Springer Series in Statistics

Zhidong Bai
Jack W. Silverstein

Spectral Analysis of Large Dimensional Random Matrices

Second Edition

Springer

Springer Series in Statistics

Advisors:
P. Bickel, P. Diggle, S. Fienberg, U. Gather,
I. Olkin, S. Zeger

For other titles published in this series, go to
http://www.springer.com/series/692

Zhidong Bai
Jack W. Silverstein

Spectral Analysis of Large Dimensional Random Matrices

Second Edition

Springer

Zhidong Bai
School of Mathematics and Statistics
KLAS MOE
Northeast Normal University
5268 Renmin Street
Changchun, Jilin 130024
China
baizd@nenu.edu.cn
&
Department of Statistics and Applied Probability
National University of Singapore
6 Science Drive 2
Singapore 117546
Singapore
stabaizd@nus.edu.sg

Jack W. Silverstein
Department of Mathematics
Box 8205
North Carolina State University
Raleigh, NC 27695-8205
jack@unity.ncsu.edu

ISSN 0172-7397
ISBN 978-1-4419-0660-1 e-ISBN 978-1-4419-0661-8
DOI 10.1007/978-1-4419-0661-8
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2009942423

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

This book is dedicated to:

Professor Calyampudi Radhakrishna Rao's 90th Birthday
Professor Ulf Grenander's 87th Birthday
Professor Yongquan Yin's 80th Birthday

and to

My wife, Xicun Dan, my sons
Li and Steve Gang, and grandsons
Yongji, and Yonglin

— Zhidong Bai

My children, Hila and Idan

— Jack W. Silverstein

Preface to the Second Edition

The ongoing developments being made in large dimensional data analysis continue to generate great interest in random matrix theory in both theoretical investigations and applications in many disciplines. This has doubtlessly contributed to the significant demand for this monograph, resulting in its first printing being sold out. The authors have received many requests to publish a second edition of the book.

Since the publication of the first edition in 2006, many new results have been reported in the literature. However, due to limitations in space, we cannot include all new achievements in the second edition. In accordance with the needs of statistics and signal processing, we have added a new chapter on the limiting behavior of eigenvectors of large dimensional sample covariance matrices. To illustrate the application of RMT to wireless communications and statistical finance, we have added a chapter on these areas. Certain new developments are commented on throughout the book. Some typos and errors found in the first edition have been corrected.

The authors would like to express their appreciation to Ms. Lü Hong for her help in the preparation of the second edition. They would also like to thank Professors Ying-Chang Liang, Zhaoben Fang, Baoxue Zhang, and Shurong Zheng, and Mr. Jiang Hu, for their valuable comments and suggestions. They also thank the copy editor, Mr. Hal Heinglein, for his careful reading, corrections, and helpful suggestions. The first author would like to acknowledge the support from grants NSFC 10871036, NUS R-155-000-079-112, and R-155-000-096-720.

Changchun, China, and Singapore — Zhidong Bai
Cary, North Carolina, USA — Jack W. Silverstein
March 2009

Preface to the First Edition

This monograph is an introductory book on the theory of random matrices (RMT). The theory dates back to the early development of quantum mechanics in the 1940s and 1950s. In an attempt to explain the complex organizational structure of heavy nuclei, E. Wigner, Professor of Mathematical Physics at Princeton University, argued that one should not compute energy levels from Schrödinger's equation. Instead, one should imagine the complex nuclei system as a black box described by $n \times n$ Hamiltonian matrices with elements drawn from a probability distribution with only mild constraints dictated by symmetry considerations. Under these assumptions and a mild condition imposed on the probability measure in the space of matrices, one finds the joint probability density of the n eigenvalues. Based on this consideration, Wigner established the well-known semicircular law. Since then, RMT has been developed into a big research area in mathematical physics and probability. Its rapid development can be seen from the following statistics from the Mathscinet database under keyword Random Matrix on 10 June 2005 (Table 0.1).

Table 0.1 Publication numbers on RMT in 10 year periods since 1955

1955–1964	1965–1974	1975–1984	1985–1994	1995–2004
23	138	249	635	1205

Modern developments in computer science and computing facilities motivate ever widening applications of RMT to many areas.

In statistics, classical limit theorems have been found to be seriously inadequate in aiding in the analysis of very high dimensional data.

In the biological sciences, a DNA sequence can be as long as several billion strands. In financial research, the number of different stocks can be as large as tens of thousands.

In wireless communications, the number of users can be several million.

All of these areas are challenging classical statistics. Based on these needs, the number of researchers on RMT is gradually increasing. The purpose of this monograph is to introduce the basic results and methodologies developed in RMT. We assume readers of this book are graduate students and beginning researchers who are interested in RMT. Thus, we are trying to provide the most advanced results with proofs using standard methods as detailed as we can.

After more than a half century, many different methodologies of RMT have been developed in the literature. Due to the limitation of our knowledge and length of the book, it is impossible to introduce all the procedures and results. What we shall introduce in this book are those results obtained either under moment restrictions using the moment convergence theorem or the Stieltjes transform.

In an attempt at complementing the material presented in this book, we have listed some recent publications on RMT that we have not introduced.

The authors would like to express their appreciation to Professors Chen Mufa, Lin Qun, and Shi Ningzhong, and Ms. Lü Hong for their encouragement and help in the preparation of the manuscript. They would also like to thank Professors Zhang Baoxue, Lee Sungchul, Zheng Shurong, Zhou Wang, and Hu Guorong for their valuable comments and suggestions.

Changchun, China — Zhidong Bai
Cary, North Carolina, USA — Jack W. Silverstein
June 2005

Contents

Chapter 1
Introduction

1.1 Large Dimensional Data Analysis

The aim of this book is to investigate the spectral properties of random matrices (RM) when their dimensions tend to infinity. All classical limiting theorems in statistics are under the assumption that the dimension of data is fixed. Then, it is natural to ask why the dimension needs to be considered large and whether there are any differences between the results for a fixed dimension and those for a large dimension.

In the past three or four decades, a significant and constant advancement in the world has been in the rapid development and wide application of computer science. Computing speed and storage capability have increased a thousand folds. This has enabled one to collect, store, and analyze data sets of very high dimension. These computational developments have had a strong impact on every branch of science. For example, Fisher's resampling theory had been silent for more than three decades due to the lack of efficient random number generators until Efron proposed his renowned bootstrap in the late 1970s; the minimum L_1 norm estimation had been ignored for centuries since it was proposed by Laplace until Huber revived it and further extended it to robust estimation in the early 1970s. It is difficult to imagine that these advanced areas in statistics would have received such deep development if there had been no assistance from the present-day computer.

Although modern computer technology helps us in so many respects, it also brings a new and urgent task to the statistician; that is, whether the classical limit theorems (i.e., those assuming a fixed dimension) are still valid for analyzing high dimensional data and how to remedy them if they are not.

Basically, there are two kinds of limiting results in multivariate analysis: those for a fixed dimension (classical limit theorems) and those for a large dimension (large dimensional limit theorems). The problem turns out to be which kind of result is closer to reality. As argued by Huber in [157], some statisticians might say that five samples for each parameter on average are

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_1,

enough to use asymptotic results. Now, suppose there are $p = 20$ parameters and we have a sample of size $n = 100$. We may consider the case as $p = 20$ being fixed and n tending to infinity, $p = 2\sqrt{n}$, or $p = 0.2n$. So, we have at least three different options from which to choose for an asymptotic setup. A natural question is then which setup is the best choice among the three. Huber strongly suggested studying the situation of an increasing dimension together with the sample size in linear regression analysis.

This situation occurs in many cases. In parameter estimation for a structured covariance matrix, simulation results show that parameter estimation becomes very poor when the number of parameters is more than four. Also, it is found in linear regression analysis that if the covariates are random (or have measurement errors) and the number of covariates is larger than six, the behavior of the estimates departs far away from the theoretic values unless the sample size is very large. In signal processing, when the number of signals is two or three and the number of sensors is more than 10, the traditional MUSIC (MUltiple SIgnal Classification) approach provides very poor estimation of the number of signals unless the sample size is larger than 1000. Paradoxically, if we use only half of the data set—namely, we use the data set collected by only five sensors—the signal number estimation is almost 100% correct if the sample size is larger than 200. Why would this paradox happen? Now, if the number of sensors (the dimension of data) is p, then one has to estimate p^2 parameters ($\frac{1}{2}p(p+1)$ real parts and $\frac{1}{2}p(p-1)$ imaginary parts of the covariance matrix). Therefore, when p increases, the number of parameters to be estimated increases proportional to p^2 while the number ($2np$) of observations increases proportional to p. This is the underlying reason for this paradox. This suggests that one has to revise the traditional MUSIC method if the sensor number is large.

An interesting problem was discussed by Bai and Saranadasa [27], who theoretically proved that when testing the difference of means of two high dimensional populations, Dempster's [91] nonexact test is more powerful than Hotelling's T^2 test even when the T^2 statistic is well defined.

It is well known that statistical efficiency will be significantly reduced when the dimension of data or number of parameters becomes large. Thus, several techniques for dimension reduction have been developed in multivariate statistical analysis. As an example, let us consider a problem in principal component analysis. If the data dimension is 10, one may select three principal components so that more than 80% of the information is reserved in the principal components. However, if the data dimension is 1000 and 300 principal components are selected, one would still have to face a high dimensional problem. If one only chooses three principal components, he would have lost 90% or even more of the information carried in the original data set. Now, let us consider another example.

Example 1.1. Let X_{ij} be iid standard normal variables. Write

$$S_n = \left(\frac{1}{n}\sum_{k=1}^{n} X_{ik}X_{jk}\right)_{i,j=1}^{p},$$

which can be considered as a sample covariance matrix with n samples of a p-dimensional mean-zero random vector with population matrix I. An important statistic in multivariate analysis is

$$T_n = \log(\det S_n) = \sum_{j=1}^{p} \log(\lambda_{n,j}),$$

where $\lambda_{n,j}$, $j = 1, \cdots, p$, are the eigenvalues of S_n. When p is fixed, $\lambda_{n,j} \to 1$ almost surely as $n \to \infty$ and thus $T_n \xrightarrow{\text{a.s.}} 0$.

Further, by taking a Taylor expansion on $\log(1 + x)$, one can show that

$$\sqrt{n/p}\,T_n \xrightarrow{\mathscr{D}} N(0, 2),$$

for any fixed p. This suggests the possibility that T_n is asymptotically normal, provided that $p = O(n)$. However, this is not the case. Let us see what happens when $p/n \to y \in (0, 1)$ as $n \to \infty$. Using results on the limiting spectral distribution of $\{S_n\}$ (see Chapter 3), we will show that with probability 1

$$\frac{1}{p}T_n \to \int_{a(y)}^{b(y)} \frac{\log x}{2\pi xy}\sqrt{(b(y) - x)(x - a(y))}dx = \frac{y-1}{y}\log(1-y) - 1 \equiv d(y) < 0 \tag{1.1.1}$$

where $a(y) = (1 - \sqrt{y})^2$, $b(y) = (1 + \sqrt{y})^2$. This shows that almost surely

$$\sqrt{n/p}\,T_n \sim d(y)\sqrt{np} \to -\infty.$$

Thus, any test that assumes asymptotic normality of T_n will result in a serious error.

These examples show that the classical limit theorems are no longer suitable for dealing with high dimensional data analysis. Statisticians must seek out special limiting theorems to deal with large dimensional statistical problems. Thus, the theory of random matrices (RMT) might be one possible method for dealing with large dimensional data analysis and hence has received more attention among statisticians in recent years. For the same reason, the importance of RMT has found applications in many research areas, such as signal processing, network security, image processing, genetic statistics, stock market analysis, and other finance or economic problems.

1.2 Random Matrix Theory

RMT traces back to the development of quantum mechanics (QM) in the 1940s and early 1950s. In QM, the energy levels of a system are described by eigenvalues of a Hermitian operator $\mathbf{A}$ on a Hilbert space, called the Hamiltonian. To avoid working with an infinite dimensional operator, it is common to approximate the system by discretization, amounting to a truncation, keeping only the part of the Hilbert space that is important to the problem under consideration. Hence, the limiting behavior of large dimensional random matrices has attracted special interest among those working in QM, and many laws were discovered during that time. For a more detailed review on applications of RMT in QM and other related areas, the reader is referred to the book *Random Matrices* by Mehta [212].

Since the late 1950s, research on the limiting spectral analysis of large dimensional random matrices has attracted considerable interest among mathematicians, probabilists, and statisticians. One pioneering work is the semicircular law for a Gaussian (or Wigner) matrix (see Chapter 2 for the definition), due to Wigner [296, 295]. He proved that the expected spectral distribution of a large dimensional Wigner matrix tends to the so-called semicircular law. This work was generalized by Arnold [8, 7] and Grenander [136] in various aspects. Bai and Yin [37] proved that the spectral distribution of a sample covariance matrix (suitably normalized) tends to the semicircular law when the dimension is relatively smaller than the sample size. Following the work of Marčenko and Pastur [201] and Pastur [230, 229], the asymptotic theory of spectral analysis of large dimensional sample covariance matrices was developed by many researchers, including Bai, Yin, and Krishnaiah [41], Grenander and Silverstein [137], Jonsson [169], Wachter [291, 290], Yin [300], and Yin and Krishnaiah [304]. Also, Yin, Bai, and Krishnaiah [301, 302], Silverstein [260], Wachter [290], Yin [300], and Yin and Krishnaiah [304] investigated the limiting spectral distribution of the multivariate F-matrix, or more generally of products of random matrices. In the early 1980s, major contributions on the existence of the limiting spectral distribution (LSD) and their explicit forms for certain classes of random matrices were made. In recent years, research on RMT has turned toward second-order limiting theorems, such as the central limit theorem for linear spectral statistics, the limiting distributions of spectral spacings, and extreme eigenvalues.

1.2.1 Spectral Analysis of Large Dimensional Random Matrices

Suppose $\mathbf{A}$ is an $m \times m$ matrix with eigenvalues $\lambda_j, j = 1, 2, \cdots, m$. If all these eigenvalues are real (e.g., if $\mathbf{A}$ is Hermitian), we can define a one-dimensional distribution function

$$F^{\mathbf{A}}(x) = \frac{1}{m}\#\{j \le m : \lambda_j \le x\} \tag{1.2.1}$$

called the empirical spectral distribution (ESD) of the matrix $\mathbf{A}$. Here $\#E$ denotes the cardinality of the set E. If the eigenvalues λ_j's are not all real, we can define a two-dimensional empirical spectral distribution of the matrix $\mathbf{A}$:

$$F^{\mathbf{A}}(x, y) = \frac{1}{m}\#\{j \le m : \Re(\lambda_j) \le x, \Im(\lambda_j) \le y\}. \tag{1.2.2}$$

One of the main problems in RMT is to investigate the convergence of the sequence of empirical spectral distributions $\{F^{\mathbf{A}_n}\}$ for a given sequence of random matrices $\{\mathbf{A}_n\}$. The limit distribution F (possibly defective; that is, total mass is less than 1 when some eigenvalues tend to $\pm\infty$), which is usually nonrandom, is called the *limiting spectral distribution* (LSD) of the sequence $\{\mathbf{A}_n\}$.

We are especially interested in sequences of random matrices with dimension (number of columns) tending to infinity, which refers to *the theory of large dimensional random matrices.*

The importance of ESD is due to the fact that many important statistics in multivariate analysis can be expressed as functionals of the ESD of some RM. We now give a few examples.

Example 1.2. Let $\mathbf{A}$ be an $n \times n$ positive definite matrix. Then

$$\det(\mathbf{A}) = \prod_{j=1}^{n} \lambda_j = \exp\left(n \int_0^\infty \log x F^{\mathbf{A}}(dx)\right).$$

Example 1.3. Let the covariance matrix of a population have the form $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_q + \sigma^2\mathbf{I}$, where the dimension of $\boldsymbol{\Sigma}$ is p and the rank of $\boldsymbol{\Sigma}_q$ is $q(< p)$. Suppose $\mathbf{S}$ is the sample covariance matrix based on n iid samples drawn from the population. Denote the eigenvalues of $\mathbf{S}$ by $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_p$. Then the test statistic for the hypothesis H_0 : rank$(\boldsymbol{\Sigma}_q) = q$ against H_1 : rank$(\boldsymbol{\Sigma}_q) > q$ is given by

$$T = \frac{1}{p-q}\sum_{j=q+1}^{p} \sigma_j^2 - \left(\frac{1}{p-q}\sum_{j=q+1}^{p} \sigma_j\right)^2$$

$$= \frac{p}{p-q}\int_0^{\sigma_q} x^2 F^{\mathbf{S}}(dx) - \left(\frac{p}{p-q}\int_0^{\sigma_q} x F^{\mathbf{S}}(dx)\right)^2.$$

1.2.2 Limits of Extreme Eigenvalues

In applications of the asymptotic theorems of spectral analysis of large dimensional random matrices, two important problems arise after the LSD is found. The first is the bound on extreme eigenvalues; the second is the convergence rate of the ESD with respect to sample size. For the first problem, the literature is extensive. The first success was due to Geman [118], who proved that the largest eigenvalue of a sample covariance matrix converges almost surely to a limit under a growth condition on all the moments of the underlying distribution. Yin, Bai, and Krishnaiah [301] proved the same result under the existence of the fourth moment, and Bai, Silverstein, and Yin [33] proved that the existence of the fourth moment is also necessary for the existence of the limit. Bai and Yin [38] found the necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. By the symmetry between the largest and smallest eigenvalues of a Wigner matrix, the necessary and sufficient conditions for almost sure convergence of the smallest eigenvalue of a Wigner matrix were also found.

Compared to almost sure convergence of the largest eigenvalue of a sample covariance matrix, a relatively harder problem is to find the limit of the smallest eigenvalue of a large dimensional sample covariance matrix. The first attempt was made in Yin, Bai, and Krishnaiah [302], in which it was proved that the almost sure limit of the smallest eigenvalue of a Wishart matrix has a positive lower bound when the ratio of the dimension to the degrees of freedom is less than 1/2. Silverstein [262] modified the work to allow a ratio less than 1. Silverstein [263] further proved that, with probability 1, the smallest eigenvalue of a Wishart matrix tends to the lower bound of the LSD when the ratio of the dimension to the degrees of freedom is less than 1. However, Silverstein's approach strongly relies on the normality assumption on the underlying distribution and thus cannot be extended to the general case. The most current contribution was made in Bai and Yin [36], in which it is proved that, under the existence of the fourth moment of the underlying distribution, the smallest eigenvalue (when $p \leq n$) or the $p - n + 1$st smallest eigenvalue (when $p > n$) tends to $a(y) = \sigma^2(1 - \sqrt{y})^2$, where $y = \lim(p/n) \in (0, \infty)$. Compared to the case of the largest eigenvalues of a sample covariance matrix, the existence of the fourth moment seems to be necessary also for the problem of the smallest eigenvalue. However, this problem has not yet been solved.

1.2.3 Convergence Rate of the ESD

The second problem, the convergence rate of the spectral distributions of large dimensional random matrices, is of practical interest. Indeed, when the LSD is used in estimating functionals of eigenvalues of a random matrix, it is important to understand the reliability of performing the substitution. This problem had been open for decades. In finding the limits of both the LSD and the extreme eigenvalues of symmetric random matrices, a very useful and powerful method is the moment method, which does not give any information about the rate of the convergence of the ESD to the LSD. The first success was made in Bai [16, 17], in which a Berry-Esseen type inequality of the difference of two distributions was established in terms of their Stieltjes transforms. Applying this inequality, a convergence rate for the expected ESD of a large Wigner matrix was proved to be $O(n^{-1/4})$ and that for the sample covariance matrix was shown to be $O(n^{-1/4})$ if the ratio of the dimension to the degrees of freedom is far from 1 and $O(n^{-5/48})$ if the ratio is close to 1. Some further developments can be found in Bai et al. [23, 24, 25], Bai et al. [26], Götze et al. [132], and Götze and Tikhomirov [133, 134].

1.2.4 Circular Law

The most perplexing problem is the so-called circular law, which conjectures that the spectral distribution of a nonsymmetric random matrix, after suitable normalization, tends to the uniform distribution over the unit disk in the complex plane. The difficulty exists in that two of the most important tools used for symmetric matrices do not apply for nonsymmetric matrices. Furthermore, certain truncation and centralization techniques cannot be used. The first known result was given in Mehta [212] (1967 edition) and in an unpublished paper of Silverstein (1984) that was reported in Hwang [159]. They considered the case where the entries of the matrix are iid standard complex normal. Their method uses the explicit expression of the joint density of the complex eigenvalues of the random matrix that was found by Ginibre [120]. The first attempt to prove this conjecture under some general conditions was made in Girko [123, 124]. However, his proofs contain serious mathematical gaps and have been considered questionable in the literature. Recently, Edelman [98] found the conditional joint distribution of complex eigenvalues of a random matrix whose entries are real normal $N(0, 1)$ when the number of its real eigenvalues is given and proved that the expected spectral distribution of the real Gaussian matrix tends to the circular law. Under the existence of the $4 + \varepsilon$ moment and the existence of a density, Bai [14] proved the strong version of the circular law. Recent work has eliminated the density requirement and weakened the moment condition. Further details are given in Chapter 11. Some consequent achievements can be found in Pan and Zhou [227] and Tao and Vu [273].

1.2.5 CLT of Linear Spectral Statistics

As mentioned above, functionals of the ESD of RMs are important in multivariate inference. Indeed, a parameter θ of the population can sometimes be expressed as

$$\theta = \int f(x)dF(x).$$

To make statistical inference on θ, one may use the integral

$$\hat{\theta} = \int f(x)dF_n(x),$$

which we call *linear spectral statistics* (LSS), as an estimator of θ, where $F_n(x)$ is the ESD of the RM computed from the data set. Further, one may want to know the limiting distribution of $\hat{\theta}$ through suitable normalization. In Bai and Silverstein [30], the normalization was found to be n by showing the limiting distribution of the linear functional

$$X_n(f) = n \int f(t)d(F_n(t) - F(t))$$

to be Gaussian under certain assumptions.

The first work in this direction was done by Jonsson [169], in which $f(t) = t^r$ and F_n is the ESD of a normalized standard Wishart matrix. Further work was done by Johansson [165], Bai and Silverstein [30], Bai and Yao [35], Sinai and Soshnikov [269], Anderson and Zeitouni [2], and Chatterjee [77], among others.

It would seem natural to pursue the properties of linear functionals by way of proving results on the process $G_n(t) = \alpha_n(F_n(t) - F(t))$ when viewed as a random element in $D[0, \infty)$, the metric space of functions with discontinuities of the first kind, along with the Skorohod metric. Unfortunately, this is impossible. The work done in Bai and Silverstein [30] shows that $G_n(t)$ cannot converge weakly to any nontrivial process for any choice of α_n. This fact appears to occur in other random matrix ensembles. When F_n is the empirical distribution of the angles of eigenvalues of an $n \times n$ Haar matrix, Diaconis and Evans [94] proved that all finite dimensional distributions of $G_n(t)$ converge in distribution to independent Gaussian variables when $\alpha_n = n/\sqrt{\log n}$. This shows that with $\alpha_n = n/\sqrt{\log n}$, the process G_n cannot be tight in $D[0, \infty)$.

The result of Bai and Silverstein [30] has been applied in several areas, especially in wireless communications, where sample covariance matrices are used to model transmission between groups of antennas. See, for example, Tulino and Verdu [283] and Kamath and Hughes [170].

1.2.6 Limiting Distributions of Extreme Eigenvalues and Spacings

The first work on the limiting distributions of extreme eigenvalues was done by Tracy and Widom [278], who found the expression for the largest eigenvalue of a Gaussian matrix when suitably normalized. Further, Johnstone [168] found the limiting distribution of the largest eigenvalue of the large Wishart matrix. In El Karoui [101], the Tracy-Widom law of the largest eigenvalue is established for the complex Wishart matrix when the population covariance matrix differs from the identity.

When the majority of the population eigenvalues are 1 and some are larger than 1, Johnstone proposed the *spiked eigenvalues model* in [168]. Then, Baik et al. [43] and Baik and Silverstein [44] investigated the strong limit of spiked eigenvalues. Bai and Yao [34] investigated the CLT of spiked eigenvalues. A special case of the CLT when the underlying distribution is complex Gaussian was considered in Baik et al. [43], and the real Gaussian case was considered in Paul [231].

The work on spectrum spacing has a long history that dates back to Mehta [213]. Most of the work in these two directions assumes the Gaussian (or generalized) distributions.

1.3 Methodologies

The eigenvalues of a matrix can be regarded as continuous functions of entries of the matrix. But these functions have no closed form when the dimension of the matrix is larger than 4. So special methods are needed to understand them. There are three important methods employed in this area: the moment method, Stieltjes transform, and orthogonal polynomial decomposition of the exact density of eigenvalues. Of course, the third method needs the assumption of the existence and special forms of the densities of the underlying distributions in the RM.

1.3.1 Moment Method

In the following, $\{F_n\}$ will denote a sequence of distribution functions, and the k-th moment of the distribution F_n is denoted by

$$\beta_{n,k} = \beta_k(F_n) := \int x^k dF_n(x). \tag{1.3.1}$$

The moment method is based on the moment convergence theorem (MCT); see Lemmas B.1, B.2, and B.3.

Let $\mathbf{A}$ be an $n \times n$ Hermitian matrix, and denote its eigenvalues by $\lambda_1 \leq \cdots \leq \lambda_n$. The ESD, $F^{\mathbf{A}}$, of $\mathbf{A}$ is defined as in (1.2.1) with m replaced by n. Then, the k-th moment of $F^{\mathbf{A}}$ can be written as

$$\beta_{n,k}(\mathbf{A}) = \int_{-\infty}^{\infty} x^k F^{\mathbf{A}}(dx) = \frac{1}{n}\mathrm{tr}(\mathbf{A}^k). \tag{1.3.2}$$

This expression plays a fundamental role in RMT. By MCT, the problem of showing that the ESD of a sequence of random matrices $\{\mathbf{A}_n\}$ (strongly or weakly or in another sense) tends to a limit reduces to showing that, for each fixed k, the sequence $\{\frac{1}{n}\mathrm{tr}(\mathbf{A}^k)\}$ tends to a limit β_k in the corresponding sense and then verifying the Carleman condition (B.1.4),

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

Note that in most cases the LSD has finite support, and hence the characteristic function of the LSD is analytic and the necessary condition for the MCT holds automatically. Most results in finding the LSD or proving the existence of the LSD were obtained by estimating the mean, variance, or higher moments of $\frac{1}{n}\mathrm{tr}(\mathbf{A}^k)$.

1.3.2 Stieltjes Transform

The definition and simple properties of the Stieltjes transform can be found in Appendix B, Section B.2. Here, we just illustrate how it can be used in RMT. Let $\mathbf{A}$ be an $n \times n$ Hermitian matrix and F_n be its ESD. Then, the Stieltjes transform of F_n is given by

$$s_n(z) = \int \frac{1}{x - z} dF_n(x) = \frac{1}{n}\mathrm{tr}(\mathbf{A} - z\mathbf{I})^{-1}.$$

Using the inverse matrix formula (see Theorem A.4), we get

$$s_n(z) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{a_{kk} - z - \boldsymbol{\alpha}_k^*(\mathbf{A}_k - z\mathbf{I})^{-1}\boldsymbol{\alpha}_k}$$

where $\mathbf{A}_k$ is the $(n-1) \times (n-1)$ matrix obtained from $\mathbf{A}$ with the k-th row and column removed and $\boldsymbol{\alpha}_k$ is the k-th column vector of $\mathbf{A}$ with the k-th element removed.

If the denominator $a_{kk} - z - \boldsymbol{\alpha}_k^*(\mathbf{A}_k - z\mathbf{I})^{-1}\boldsymbol{\alpha}_k$ can be proven to be equal to $g(z, s_n(z)) + o(1)$ for some function g, then the LSD F exists and its Stieltjes transform of F is the solution to the equation

$$s = 1/g(z, s).$$

Its applications will be discussed in more detail later.

1.3.3 Orthogonal Polynomial Decomposition

Assume that the matrix $\mathbf{A}$ has a density $p_n(\mathbf{A}) = H(\lambda_1, \cdots, \lambda_n)$. It is known that the joint density function of the eigenvalues will be of the form

$$p_n(\lambda_1, \cdots, \lambda_n) = cJ(\lambda_1, \cdots, \lambda_n)H(\lambda_1, \cdots, \lambda_n),$$

where J comes from the integral of the Jacobian of the transform from the matrix space to its eigenvalue-eigenvector space. Generally, it is assumed that H has the form $H(\lambda_1, \cdots, \lambda_n) = \prod_{k=1}^n g(\lambda_k)$ and J has the form $\prod_{i<j}(\lambda_i - \lambda_j)^\beta \prod_{k=1}^n h_n(\lambda_k)$. For example, $\beta = 1$ and $h_n = 1$ for a real Gaussian matrix, $\beta = 2$, $h_n = 1$ for a complex Gaussian matrix, $\beta = 4$, $h_n = 1$ for a quaternion Gaussian matrix, and $\beta = 1$ and $h_n(x) = x^{n-p}$ for a real Wishart matrix with $n \geq p$.

Examples considered in the literature are the following

(1) Real Gaussian matrix (symmetric; i.e., $\mathbf{A}' = \mathbf{A}$):

$$p_n(\mathbf{A}) = c \exp\left(-\frac{1}{4\sigma^2}\mathrm{tr}(\mathbf{A}^2)\right).$$

In this case, the diagonal entries of $\mathbf{A}$ are iid real $N(0, 2\sigma^2)$ and entries above diagonal are iid real $N(0, \sigma^2)$.

(2) Complex Gaussian matrix (Hermitian; i.e., $\mathbf{A}^* = \mathbf{A}$):

$$p_n(\mathbf{A}) = c \exp\left(-\frac{1}{2\sigma^2}\mathrm{tr}(\mathbf{A}^2)\right).$$

In this case, the diagonal entries of $\mathbf{A}$ are iid real $N(0, \sigma^2)$ and entries above diagonal are iid complex $N(0, \sigma^2)$ (whose real and imaginary parts are iid $N(0, \sigma^2/2)$).

(3) Real Wishart matrix of order $p \times n$:

$$p_n(\mathbf{A}) = c \exp\left(-\frac{1}{2\sigma^2}\mathrm{tr}(\mathbf{A}'\mathbf{A})\right).$$

In this case, the entries of $\mathbf{A}$ are iid real $N(0, \sigma^2)$.

(4) Complex Wishart matrix of order $p \times n$:

$$p_n(\mathbf{A}) = c\exp\left(-\frac{1}{\sigma^2}\mathrm{tr}(\mathbf{A}^*\mathbf{A})\right).$$

In this case, the entries of $\mathbf{A}$ are iid complex $N(0, \sigma^2)$.

For generalized densities, there are the following.

(1) Symmetric matrix:
$$p_n(\mathbf{A}) = c\exp(-\mathrm{tr}V(\mathbf{A})).$$

(2) Hermitian matrix:
$$p_n(\mathbf{A}) = c\exp(-\mathrm{tr}V(\mathbf{A})).$$

In the two cases above, V is assumed to be a polynomial of even degree with a positive leading coefficient.

(3) Real covariance matrix of dimension p and degrees of freedom n:
$$p_n(\mathbf{A}) = c\exp(-\mathrm{tr}V(\mathbf{A}'\mathbf{A})).$$

(4) Complex covariance matrix of dimension p and degrees of freedom n:
$$p_n(\mathbf{A}) = c\exp(-\mathrm{tr}V(\mathbf{A}^*\mathbf{A})).$$

In the two cases above, V is assumed to be a polynomial with a positive leading coefficient.

Note that the factor $\prod_{i<j}(\lambda_i - \lambda_j)$ is the determinant of the Vandermonde matrix generated by $\lambda_1, \cdots, \lambda_n$. Therefore, we may rewrite the density of the eigenvalues of the matrices as

$$\begin{aligned}
&p_n(\lambda_1, \cdots, \lambda_n) \\
&= c\prod_{k=1}^n h_n(\lambda_k)g(\lambda_k)\det\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & \cdots & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{pmatrix}^\beta \\
&= c\prod_{k=1}^n h_n(\lambda_k)g(\lambda_k)\det\begin{pmatrix} 1 & 1 & \cdots & 1 \\ m_1(\lambda_1) & m_1(\lambda_2) & \cdots & m_1(\lambda_n) \\ \vdots & \vdots & \cdots & \vdots \\ m_{n-1}(\lambda_1) & m_{n-1}(\lambda_2) & \cdots & m_{n-1}(\lambda_n) \end{pmatrix}^\beta,
\end{aligned}$$

where m_k is any polynomial of degree k and having leading coefficient 1. For ease of finding the marginal densities of several eigenvalues, one may choose the m functions as orthogonal polynomials with respective $[g(x)h_n(x)]^{2/\beta}$. Then, through mathematical analysis, one can draw various conclusions from the expression above.

Note that the moment method and Stieltjes transform method can be done under moment assumptions. This book will primarily concentrate on results without assuming density conditions. Readers who are interested in the method of orthogonal polynomials are referred to Deift [88].

1.3.4 Free Probability

Free probability is a mathematical theory that studies noncommutative random variables. The "freeness" property is the analogue of the classical notion of independence, and it is connected with free products. This theory was initiated by Dan Voiculescu around 1986 in order to attack the free group factors isomorphism problem, an important unsolved problem in the theory of operator algebras. Typically the random variables lies in a unital algebra A such as a C^* algebra or a von Neumann algebra. The algebra comes equipped with a noncommutative expectation, a linear functional $\varphi : A \to C$ such that $\varphi(1) = 1$. Unital subalgebras $A_1, \cdots, A_n$ are then said to be free if the expectation of the product $a_1 \cdots a_n$ is zero whenever each a_j has zero expectation, lies in an A_k, and no adjacent a_j's come from the same subalgebra A_k. Random variables are free if they generate free unital subalgebras.

An interesting aspect and active research direction of free probability lies in its applications to RMT. The functional φ stands for the normalized expected trace of a random matrix. For any $n \times n$ Hermitian random matrix $\mathbf{A}_n$ and a given integer k, $\varphi(\mathbf{A}_n^k) = \frac{1}{n}\mathrm{tr}(\mathrm{E}\mathbf{A}_n^k)$. If $\lim_n \varphi(\mathbf{A}_n^k) = \alpha_k$, for all k, then instead of referring to the collection of numbers α_k, it is better to use some random variable A (if it exists) to characterize the α_k's as moments of A. By setting $\varphi(A^k) = \alpha_k$, one may say that $\mathbf{A}_n \to A$ in distribution. A general definition is given as follows.

Definition 1.4. Consider $n \times n$ random matrices $A_n^{(1)}, \cdots, A_n^{(m)}$ and variables $A_1, \cdots, A_m$. We say that

$$(A_n^{(1)}, \cdots, A_n^{(m)}) \to (A_1, \cdots, A_m) \text{ in distribution}$$

if

$$\lim_{n \to \infty} \varphi(A_n^{(i_1)} \cdots A_n^{(i_k)}) = \varphi(A_{i_1} \cdots A_{i_k})$$

for all choices of k, $1 \le i_1, \cdots, i_k \le m$.

When $m = 1$, the definition of convergence in distribution is to say that if the normalized expected trace of $\mathbf{A}_n^k$ tends to the k-th moment of A, then we define $\mathbf{A}_n$ tending to A. For example, let $\mathbf{A}_n$ be the normalized Wigner matrix (see Chapter 2). Then A is the semicircular law. Now, suppose we have two independent sequences of normalized Wigner matrices, $\{\mathbf{A}_n\}$ and $\{\mathbf{B}_n\}$. How do we characterize their limits? If individually, then $\mathbf{A}_n \to s_a$ and $\mathbf{B}_n \to s_b$, and both s_a and s_b are semicircular laws. The problem is how to consider the joint limit of the sequences of pairs $(\mathbf{A}_n, \mathbf{B}_n)$. Or equivalently, what is the relationship of s_a and s_b? According to free probability, we have the following definition.

Definition 1.5. The matrices $\mathbf{A}_1, \cdots, \mathbf{A}_m$ are called free if

$$\varphi([p_1(\mathbf{A}_{i_1}) \cdots p_k(\mathbf{A}_{i_k})]) = 0$$

whenever

- $p_1, \cdots, p_k$ are polynomials in one variable,
- $i_1 \neq i_2 \neq i_3 \neq \cdots \neq i_k$ (only neighboring elements are required to be distinct),
- $\varphi(p_j(\mathbf{A}_{i_j})) = 0$ for all $j = 1, \cdots, k$.

Note that the definition of freeness can be considered as a way of organizing the information about all joint moments of free variables in a systematic and conceptual way. Indeed, the definition above allows one to calculate mixed moments of free variables in terms of moments of the single variables. For example, if a, b are free, then the definition of freeness requires that $\varphi[(a - \varphi(a)1)(b - \varphi(b)1)] = 0$, which implies that $\varphi(ab) = \varphi(a)\varphi(b)$. In the same way, $\varphi[(a - \varphi(a)1)(b - \varphi(b)1)(a - \varphi(a)1)(b - \varphi(b)1)] = 0$ leads finally to $\varphi(abab) = \varphi(aa)\varphi(b)\varphi(b) + \varphi(a)\varphi(a)\varphi(bb) - \varphi(a)\varphi(b)\varphi(a)\varphi(b)$. Analogously, all mixed moments can (at least in principle) be calculated by reducing them to alternating products of centered variables as in the definition of freeness. Thus the statements s_a, s_b are free, and each of them being semicircular determines all joint moments in s_a and s_b. This shows that s_a and s_b are not ordinary random variables but take values on some noncommutative algebra.

To apply the theory of free probability to RMT, we need to extend the definition of free to asymptotic freeness; that is, replacing the state functional φ by ϕ, where

$$\phi(\mathbf{A}) = \lim_{n\to\infty} \frac{1}{n} \mathrm{trE}(\mathbf{A}_n).$$

Since normalized traces of powers of a Hermitian matrix are the moments of the ESD of the matrix, free probability reveals important information on their LSD. It is shown that freeness of random matrices corresponds to independence and to distributions being invariant under orthogonal transformations. Formulas have been derived that express the LSD of sums and products of free random matrices in terms of their individual LSDs.

For an excellent introduction to free probability, see Biane [52] and Nica and Speicher [221].

Chapter 2
Wigner Matrices and Semicircular Law

A Wigner matrix is a symmetric (or Hermitian in the complex case) random matrix. Wigner matrices play an important role in nuclear physics and mathematical physics. The reader is referred to Mehta [212] for applications of Wigner matrices to these areas. Here we mention that they also have a strong statistical meaning. Consider the limit of a normalized Wishart matrix. Suppose that $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are iid samples drawn from a p-dimensional multivariate normal population $N(\boldsymbol{\mu}, \mathbf{I}_p)$. Then, the sample covariance matrix is defined as

$$\mathbf{S}_n = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})',$$

where $\overline{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$. When n tends to infinity, $\mathbf{S}_n \to \mathbf{I}_p$ and $\sqrt{n}\,(\mathbf{S}_n - \mathbf{I}_p) \to \sqrt{p}\mathbf{W}_p$. It can be seen that the entries above the main diagonal of $\sqrt{p}\mathbf{W}_p$ are iid $N(0,1)$ and the entries on the diagonal are iid $N(0,2)$. This matrix is called the (standard) Gaussian matrix or Wigner matrix.

A generalized definition of Wigner matrix only requires the matrix to be a Hermitian random matrix whose entries on or above the diagonal are independent. The study of spectral analysis of the large dimensional Wigner matrix dates back to Wigner's [295] famous **semicircular law**. He proved that the expected ESD of an $n \times n$ standard Gaussian matrix, normalized by $1/\sqrt{n}$, tends to the semicircular law F whose density is given by

$$F'(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4 - x^2}, & \text{if } |x| \le 2, \\ 0, & \text{otherwise.} \end{cases} \tag{2.0.1}$$

This work has been extended in various aspects. Grenander [136] proved that $\|F^{\mathbf{W}_n} - F\| \to 0$ in probability. Further, this result was improved as in the sense of "almost sure" by Arnold [8, 7]. Later on, this result was further generalized, and it will be introduced in the following sections.

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_2,

2.1 Semicircular Law by the Moment Method

In order to apply the moment method (see Appendix B, Section B.1) to prove the convergence of the ESD of Wigner matrices to the semicircular distribution, we calculate the moments of the semicircular distribution and show that they satisfy the Carleman condition. In the remainder of this section, we will show the convergence of the ESD of the Wigner matrix by the moment method.

2.1.1 Moments of the Semicircular Law

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

Lemma 2.1. *For $k = 0, 1, 2, \cdots,$ we have*

$$\beta_{2k} = \frac{1}{k+1}\binom{2k}{k},$$
$$\beta_{2k+1} = 0.$$

Proof. Since the semicircular distribution is symmetric about 0, thus we have $\beta_{2k+1} = 0$. Also, we have

$$\begin{aligned}
\beta_{2k} &= \frac{1}{2\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}dx \\
&= \frac{1}{\pi}\int_{0}^{2} x^{2k}\sqrt{4-x^2}dx \\
&= \frac{2^{2k+1}}{\pi}\int_{0}^{1} y^{k-1/2}(1-y)^{1/2}dy \quad \text{(by setting } \ x = 2\sqrt{y}) \\
&= \frac{2^{2k+1}}{\pi}\frac{\Gamma(k+1/2)\Gamma(3/2)}{\Gamma(k+2)} = \frac{1}{k+1}\binom{2k}{k}.
\end{aligned}$$

2.1.2 Some Lemmas in Combinatorics

In order to calculate the limits of moments of the ESD of a Wigner matrix, we need some information from combinatorics. This is because the mean and variance of each empirical moment will be expressed as a sum of expectations of products of matrix entries, and we need to be able to systematically count the number of significant terms. To this end, we introduce some concepts from graph theory and establish some lemmas.

A graph is a triple (E, V, F), where E is the set of edges, V is the set of vertices, and F is a function, $F : E \mapsto V \times V$. If $F(e) = (v_1, v_2)$, the vertices v_1, v_2 are called the ends of the edge e, v_1 is the initial of e, and v_2 is the terminal of e. If $v_1 = v_2$, edge e is a loop. If two edges have the same set of ends, they are said to be coincident.

Let $\mathbf{i} = (i_1, \cdots, i_k)$ be a vector valued on $\{1, \cdots, n\}^k$. With the vector $\mathbf{i}$, we define a Γ-graph as follows. *Draw a horizontal line and plot the numbers* $i_1, \cdots, i_k$ *on it. Consider the distinct numbers as vertices, and draw* k *edges* e_j *from* i_j *to* i_{j+1}, $j = 1, \cdots, k$, *where* $i_{k+1} = i_1$ *by convention. Denote the number of distinct* i_j*'s by* t. Such a graph is called a $\Gamma(k, t)$-graph. An example of $\Gamma(6, 4)$ is shown in Fig. 2.1.

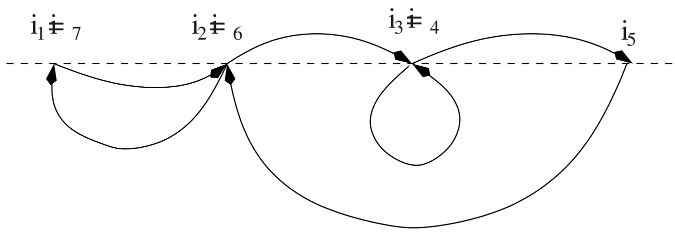

Fig. 2.1 A Γ-graph

By definition, a $\Gamma(k, t)$-graph starts from vertex i_1, and the k edges consecutively connect one after another and finally return to vertex i_1. That is, a $\Gamma(k, t)$-graph forms a cycle.

Two $\Gamma(k, t)$-graphs are said to be isomorphic if one can be converted to the other by a permutation of $(1, \cdots, n)$. By this definition, all Γ-graphs are classified into isomorphism classes.

We shall call the $\Gamma(k, t)$-graph canonical if it has the following properties:

1. Its vertex set is $V = \{1, \cdots, t\}$.
2. Its edge set is $E = \{e_1, \cdots, e_k\}$.
3. There is a function g from $\{1, 2, \cdots, k\}$ onto $\{1, 2, \cdots, t\}$ satisfying $g(1) = 1$ and $g(i) \le \max\{g(1), \cdots, g(i-1)\} + 1$ for $1 < i \le k$.
4. $F(e_i) = (g(i), g(i+1))$, for $i = 1, \cdots, k$, with convention $g(k+1) = g(1) = 1$.

It is easy to see that each isomorphism class contains one and only one canonical Γ-graph that is associated with a function g, and a general graph in this class can be defined by $F(e_j) = (i_{g(j)}, i_{g(j+1)})$. Therefore, we have the following lemma.

Lemma 2.2. *Each isomorphism class contains* $n(n-1)\cdots(n-t+1)$ $\Gamma(k, t)$ *graphs.*

The canonical $\Gamma(k,t)$-graphs can be classified into three categories.

Category 1 (denoted by $\Gamma_1(k)$): A canonical graph $\Gamma(k,t)$ is said to belong to category 1 if each edge is coincident with exactly one other edge of opposite direction and the graph of noncoincident edges forms a tree (i.e., a connected graph without cycles). It is obvious that there is no $\Gamma_1(k)$ if k is odd.

Category 2 ($\Gamma_2(k,t)$) consists of all those canonical $\Gamma(k,t)$-graphs that have at least one single edge; i.e., an edge not coincident with any other edges.

Category 3 ($\Gamma_3(k,t)$) consists of all other canonical $\Gamma(k,t)$-graphs. If we classify the k edges into coincidence classes, a $\Gamma_3(k,t)$-graph contains either a coincidence class of at least three edges or a cycle of noncoincident edges. In both cases, $t \le (k+1)/2$. Then, in fact we have proved the following lemma.

Lemma 2.3. *In a $\Gamma_3(k,t)$-graph, $t \le (k+1)/2$.*

Now, we begin to count the number of $\Gamma_1(k)$-graphs for $k = 2m$. We have the following lemma.

Lemma 2.4. *The number of $\Gamma_1(2m)$-graphs is $\frac{1}{m+1}\binom{2m}{m}$.*

Proof. Suppose G is a graph of $\Gamma_1(2m)$. We define a function $H : E \to \{-1,1\}$; $H(e) = +1$ if e is single up to itself (called an innovation) and $= -1$ otherwise (called a Type 3 (T_3) edge, the edge that coincides with an innovation that is single up to it). Corresponding to the graph G, we call the sequence $(H(e_1),\cdots,H(e_k)) = (a_1 = 1, a_2, \cdots, a_{2m-1}, a_{2m} = -1)$ the characteristic sequence of the graph G. By definition, all partial sums of the characteristic sequence are nonnegative; i.e., for all $1 \le \ell \le 2m$,

$$a_1 + a_2 + \cdots + a_\ell \ge 0. \tag{2.1.1}$$

We show that there is a one-to-one correspondence between $\Gamma_1(2m)$-graphs and the characteristic sequences. That is, we need to show that any sequence of ± 1 satisfying (2.1.1) corresponds to a $\Gamma_1(2m)$-graph. Suppose $(a_1,\cdots,a_{2m})$ is a given sequence satisfying (2.1.1). We construct a $\Gamma_1(2m)$-graph with the given sequence as its characteristic sequence.

By (2.1.1), $a_1 = 1$ and $F(e_1) = (1,2)$; i.e., $g(1) = 1$, $g(2) = 2$. Suppose $g(1), g(2), \cdots, g(s)$ $(2 \le s < 2m)$ have been defined with the following properties:

(i) For each $i \le s$, we have $g(i) \le \max\{g(1),\cdots,g(i-1)\} + 1$.

(ii) If we define $(g(i), g(i+1))$, $i = 1,\cdots,s-1$, as edges, then from $g(1) = 1$ to $g(s)$ there is a path of single innovations if $g(s) \ne 1$. All other edges not on the path must coincide with another edge of opposite direction. If $g(s) = 1$, then each edge coincides with another edge of opposite direction.

(iii) $H(g(i), g(i+1)) = a_i$ for all $i < s$.

Now, we define $g(s+1)$ in the following way:

Case 1. If $a_s = 1$, define $g(s+1) = \max\{g(1), \cdots, g(s)\} + 1$. Obviously, the edge $(g(s), g(s+1))$ is a single innovation that, combining the original path of single innovations, forms the new path of single innovations from $g(1) = 1$ to $g(s+1)$ if $g(s) \neq 1$. If $g(s) = 1$, then $g(s+1) \neq 1$ and the edge $(g(s), g(s+1))$ forms the new path of single innovations. Also, all other edges coincide with an edge of opposite directions. That is, conditions (i)–(iii) are satisfied.

Case 2. If $a_s = -1$, then $g(s) \neq 1$ for otherwise condition (2.1.1) will be violated. Hence, there is an $i < s$ such that $(g(i), g(s))$ is a single innovation (the last edge of a path of single innovations). Then, define $g(s+1) = g(i)$. If $g(i) = 1$, then the new graph has no single edges. If $g(i) \neq 1$, the original path of single innovations has at least two single innovations. Then, the new path of single innovations is obtained by cutting the last edge from the original path of single innovations. Also, conditions (i)–(iii) are satisfied.

By induction, the functions $g(1), \cdots, g(2m)$ are well defined, and hence a $\Gamma_1(2m)$ with characteristic sequence $(a_1, \cdots, a_{2m})$ is defined.

Therefore, to count the number of $\Gamma_1(2m)$-graphs is equivalent to counting the number of characteristic sequences of isomorphism classes.

Arbitrarily arrange m ones and m minus ones. The total number of possibilities is obviously $\binom{2m}{m}$. We shall use the symmetrization principle to count the number of noncharacteristic sequences. Write the sequence of ± 1s as $(a_1, \cdots, a_{2m})$ and $S_0 = 0$ and $S_i = S_{i-1} + a_i$, for $i = 1, 2, \cdots, 2m$. Plot the graph of $(i, S(i))$ on the plane. The graph should start from $(0, 0)$ and return to $(2m, 0)$. If for all i, $S_i \geq 0$ (that is, the figure is totally above or on the horizontal axis), then $(a_1, \cdots, a_{2m})$ is a characteristic sequence. Otherwise, if $(a_1, \cdots, a_{2m})$ is not a characteristic sequence, then there must be an $i \geq 1$ such that $S_i = -1$. Then we turn over the rear part after i along the line $S = -1$ and we get a new graph $(0, 0)$ to $(2m, -2)$, as shown in Fig. 2.2.

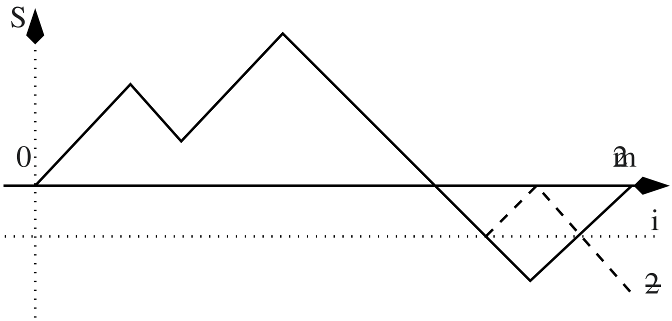

Fig. 2.2 Symmetrization principle

This is equivalent to defining $b_j = a_j$ for $j \leq i$ and $b_j = -a_j$ for $j > i$. Then, the sequence $(b_1, \cdots, b_{2m})$ contains $m - 1$ ones and $m + 1$ minus ones.

Conversely, for any sequence of $m-1$ ones and $m+1$ minus ones, there must be a smallest integer $i < 2m$ such that $b_1 + \cdots + b_i = -1$. Then the sequence $(b_1, \cdots, b_k, -b_{k+1}, \cdots, -b_{2m})$ contains m ones and m minus ones which is a noncharacteristic sequence. The number of b-sequences is $\binom{2m}{m-1}$. Thus, the number of characteristic sequences is

$$\binom{2m}{m} - \binom{2m}{m-1} = \frac{1}{m+1}\binom{2m}{m}.$$

The proof of the lemma is complete.

2.1.3 Semicircular Law for the iid Case

In this subsection, we will show the semicircular law for the iid case; that is, we shall prove the following theorem. For brevity of notation, we shall use $\mathbf{X}_n$ for an $n \times n$ Wigner matrix and save the notation $\mathbf{W}_n$ for the normalized Wigner matrix, i.e., $\frac{1}{\sqrt{n}}\mathbf{X}_n$.

Theorem 2.5. *Suppose that $\mathbf{X}_n$ is an $n \times n$ Hermitian matrix whose diagonal entries are iid real random variables and those above the diagonal are iid complex random variables with variance $\sigma^2 = 1$. Then, with probability 1, the ESD of $\mathbf{W}_n = \frac{1}{\sqrt{n}}\mathbf{X}_n$ tends to the semicircular law.*

Before applying the MCT to the proof of Theorem 2.5, we first remove the diagonal entries of $\mathbf{X}_n$, truncate the off-diagonal entries of the matrix, and renormalize them, without changing the LSD. We will proceed with the proof by taking the following steps.

Step 1. Removing the Diagonal Elements

Let $\widetilde{\mathbf{W}}_n$ be the matrix obtained from $\mathbf{W}_n$ by replacing the diagonal elements with zero. We shall show that the two matrices are asymptotically equivalent; i.e., their LSDs are the same if one of them exists.

Let $N_n = \#\{|x_{ii}| \geq \sqrt[4]{n}\}$. Replace the diagonal elements of $\mathbf{W}_n$ by $\frac{1}{\sqrt{n}} x_{ii} I(|x_{ii}| < \sqrt[4]{n})$, and denote the resulting matrix by $\widehat{\mathbf{W}}_n$. Then, by Corollary A.41, we have

$$L^3(F^{\widehat{\mathbf{W}}_n}, F^{\widetilde{\mathbf{W}}_n}) \leq \frac{1}{n}\mathrm{tr}[(\widetilde{\mathbf{W}}_n - \widehat{\mathbf{W}}_n)^2] \leq \frac{1}{n^2}\sum_{i=1}^{n} |x_{ii}|^2 I(|x_{ii}| < \sqrt[4]{n}) \leq \frac{1}{\sqrt{n}}.$$

On the other hand, by Theorem A.43, we have

$$\left\| F^{\mathbf{W}_n} - F^{\widetilde{\mathbf{W}}_n} \right\| \leq \frac{N_n}{n}.$$

Therefore, to complete the proof of our assertion, it suffices to show that $N_n/n \to 0$ almost surely. Write $p_n = \mathrm{P}(|x_{11}| \geq \sqrt[4]{n}) \to 0$. By Bernstein's inequality,[1] we have, for any $\varepsilon > 0$,

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

for some positive constant $b > 0$. This completes the proof of our assertion.

In the following subsections, we shall assume that the diagonal elements of $\mathbf{W}_n$ are all zero.

Step 2. Truncation

For any fixed positive constant C, truncate the variables at C and write $x_{ij(C)} = x_{ij}I(|x_{ij}| \leq C)$. Define a truncated Wigner matrix $\mathbf{W}_{n(C)}$ whose diagonal elements are zero and off-diagonal elements are $\frac{1}{\sqrt{n}}x_{ij(C)}$. Then, we have the following truncation lemma.

Lemma 2.6. *Suppose that the assumptions of Theorem 2.5 are true. Truncate the off-diagonal elements of $\mathbf{X}_n$ at C, and denote the resulting matrix by $\mathbf{X}_{n(C)}$. Write $\mathbf{W}_{n(C)} = \frac{1}{\sqrt{n}}\mathbf{X}_{n(C)}$. Then, for any fixed constant C,*

$$\limsup_n L^3(F^{\mathbf{W}_n}, F^{\mathbf{W}_{n(C)}}) \leq \mathrm{E}\big(|x_{11}|^2 I(|x_{11}| > C)\big), \quad \text{a.s.} \tag{2.1.2}$$

Proof. By Corollary A.41 and the law of large numbers, we have

$$L^3(F^{\mathbf{W}_n}, F^{\mathbf{W}_{n(C)}}) \leq \frac{2}{n^2}\left(\sum_{1\leq i<j\leq n}|x_{ij}|^2 I(|x_{11}| > C)\right)$$
$$\to \mathrm{E}\big(|x_{11}|^2 I(|x_{11}| > C)\big).$$

This completes the proof of the lemma.

Note that the right-hand side of (2.1.2) can be made arbitrarily small by making C large. Therefore, in the proof of Theorem 2.5, we can assume that the entries of the matrix $\mathbf{X}_n$ are uniformly bounded.

Step 3. Centralization

Applying Theorem A.43, we have

$$\left\|F^{\mathbf{W}_{n(C)}} - F^{\mathbf{W}_{n(C)} - a\mathbf{1}\mathbf{1}'}\right\| \leq \frac{1}{n}, \tag{2.1.3}$$

[1] Bernstein's inequality states that if $X_1, \cdots, X_n$ are independent random variables with mean zero and uniformly bounded by b, then, for any $\varepsilon > 0$,
$P(|S_n| \geq \varepsilon) \leq 2\exp(-\varepsilon^2/[2(B_n^2 + b\varepsilon)])$, where $S_n = X_1 + \cdots + X_n$ and $B_n^2 = \mathrm{E}S_n^2$.

where $a = \frac{1}{\sqrt{n}}\Re(\mathrm{E}(x_{12(C)}))$. Furthermore, by Corollary A.41, we have

$$L(F^{\mathbf{W}_{n(C)}-\Re(\mathrm{E}(\mathbf{W}_{n(C)}))}, F^{\mathbf{W}_{n(C)}-a\mathbf{1}\mathbf{1}'}) \le \frac{|\Re(\mathrm{E}(x_{12(C)}))|^2}{n} \to 0. \tag{2.1.4}$$

This shows that we can assume that the real parts of the mean values of the off-diagonal elements are 0. In the following, we proceed to remove the imaginary part of the mean values of the off-diagonal elements.

Before we treat the imaginary part, we introduce a lemma about eigenvalues of a skew-symmetric matrix.

Lemma 2.7. *Let $\mathbf{A}_n$ be an $n \times n$ skew-symmetric matrix whose elements above the diagonal are 1 and those below the diagonal are -1. Then, the eigenvalues of $\mathbf{A}_n$ are $\lambda_k = i\cot(\pi(2k-1)/2n)$, $k = 1, 2, \cdots, n$. The eigenvector associated with λ_k is $\mathbf{u}_k = \frac{1}{\sqrt{n}}(1, \rho_k, \cdots, \rho_k^{n-1})'$, where $\rho_k = (\lambda_k - 1)/(\lambda_k + 1) = \exp(-i\pi(2k-1)/n)$.*

Proof. We first compute the characteristic polynomial of $\mathbf{A}_n$.

$$D_n = |\lambda\mathbf{I} - \mathbf{A}_n| = \begin{vmatrix} \lambda & -1 & -1 & \cdots & -1 \\ 1 & \lambda & -1 & \cdots & -1 \\ 1 & 1 & \lambda & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & \lambda \end{vmatrix}$$
$$= \begin{vmatrix} \lambda-1 & -(1+\lambda) & 0 & \cdots & 0 \\ 0 & \lambda-1 & -(1+\lambda) & \cdots & 0 \\ 0 & 0 & \lambda-1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & \lambda \end{vmatrix}.$$

Expanding the above along the first row, we get the following recursive formula

$$D_n = (\lambda - 1)D_{n-1} + (1 + \lambda)^{n-1},$$

with the initial value $D_1 = \lambda$. The solution is

$$\begin{aligned} D_n &= \lambda(\lambda-1)^{n-1} + (\lambda+1)(\lambda-1)^{n-2} + \cdots + (\lambda+1)^{n-1} \\ &= \frac{1}{2}\left((\lambda-1)^n + (\lambda+1)^n\right). \end{aligned}$$

Setting $D_n = 0$, we get

$$\frac{\lambda+1}{\lambda-1} = e^{i\pi(2k-1)/n}, \quad k = 1, 2, \cdots, n, \tag{2.1.5}$$

which implies that $\lambda = i\cot(\pi(2k-1)/2n)$.

Comparing the two sides of the equation $\mathbf{A}_n\mathbf{u}_k = \lambda_k\mathbf{u}_k$, we obtain

$$-u_{k,1} - \cdots - u_{k,\ell-1} + u_{k,\ell+1} + \cdots + u_{k,n} = \lambda_k u_{k,\ell}$$

for $\ell = 1, 2, \cdots, n$. Thus, subtracting the equations for $\ell + 1$ from that for ℓ, we get

$$u_{k,\ell} + u_{k,\ell+1} = \lambda_k(u_{k,\ell} - u_{k,\ell+1}),$$

which implies that

$$\frac{u_{k,\ell+1}}{u_{k,\ell}} = \frac{\lambda_k - 1}{\lambda_k + 1} = e^{-i\pi(2k-1)/n} := \rho_k.$$

Therefore, one can choose $u_{k,\ell} = \rho_k^{\ell-1}/\sqrt{n}$.

The proof of the lemma is complete.

Write $b = \mathrm{E}\Im(x_{12(C)})$. Then, $\mathrm{E}\Im(\mathbf{W}_{n(C)}) = ib\mathbf{A}_n$. By Lemma 2.7, the eigenvalues of the matrix $i\Im(\mathrm{E}(\mathbf{W}_{n(C)})) = ib\mathbf{A}_n$ are $ib\lambda_k = -n^{-1/2}b\cot(\pi(2k-1)/2n)$, $k = 1, \cdots, n$. If the spectral decomposition of $\mathbf{A}_n$ is $\mathbf{U}_n\mathbf{D}_n\mathbf{U}_n^*$, then we rewrite $i\Im(\mathrm{E}(\mathbf{W}_{n(C)})) = \mathbf{B}_1 + \mathbf{B}_2$, where $\mathbf{B}_j = -\frac{1}{\sqrt{n}}b\mathbf{U}_n\mathbf{D}_{nj}\mathbf{U}_n^*$, $j = 1, 2$, where $\mathbf{U}_n$ is a unitary matrix, $\mathbf{D}_n = \mathrm{diag}[\lambda_1, \cdots, \lambda_n]$, and

$$\mathbf{D}_{n1} = \mathbf{D}_n - \mathbf{D}_{n2} = \mathrm{diag}[0, \cdots, 0, \lambda_{[n^{3/4}]}, \lambda_{[n^{3/4}]+1}, \cdots, \lambda_{n-[n^{3/4}]}, 0, \cdots, 0].$$

For any $n \times n$ Hermitian matrix $\mathbf{C}$, by Corollary A.41, we have

$$\begin{aligned} L^3(F^{\mathbf{C}}, F^{\mathbf{C}-\mathbf{B}_1}) &\le \frac{1}{n^2} \sum_{n^{3/4} \le k \le n - n^{3/4}} \cot^2(\pi(2k-1)/2n) \\ &< \frac{2}{n \sin^2(n^{-1/4}\pi)} \to 0 \end{aligned} \tag{2.1.6}$$

and, by Theorem A.43,

$$\|F^{\mathbf{C}} - F^{\mathbf{C}-\mathbf{B}_2}\| \le \frac{2n^{3/4}}{n} \to 0. \tag{2.1.7}$$

Summing up estimations (2.1.3)–(2.1.7), we established the following centralization lemma.

Lemma 2.8. *Under the conditions assumed in Lemma 2.6, we have*

$$L(F^{\mathbf{W}_{n(C)}}, F^{\mathbf{W}_{n(C)} - \mathrm{E}(\mathbf{W}_{n(C)})}) = o(1). \tag{2.1.8}$$

Step 4. Rescaling

Write $\sigma^2(C) = \mathrm{Var}(x_{11(C)})$, and define $\widetilde{\mathbf{W}}_n = \sigma^{-1}(C)\big(\mathbf{W}_{n(C)} - \mathrm{E}(\mathbf{W}_{n(C)})\big)$. Note that the off-diagonal entries of $\sqrt{n}\widetilde{\mathbf{W}}_n$ are $\widehat{x}_{kj} = \sigma^{-1}(C)\big(x_{kj(C)} - \mathrm{E}(x_{kj(C)})\big)$.

Applying Corollary A.41, we obtain

$$L^3(F^{\widetilde{\mathbf{W}}_n}, F^{\mathbf{W}_{n(C)}-\mathrm{E}(\mathbf{W}_{n(C)})}) \leq \frac{2(\sigma(C)-1)^2}{n^2\sigma^2(C)} \sum_{1\leq i<j\leq n} |x_{kj(C)} - \mathrm{E}(x_{kj(C)})|^2$$
$$\to (\sigma(C)-1)^2, \quad \text{a.s.} \tag{2.1.9}$$

Note that $(\sigma(C)-1)^2$ can be made arbitrarily small if C is large. Combining (2.1.9) with Lemmas 2.6 and 2.8, to prove the semicircular law, we may assume that the entries of $\mathbf{X}$ are bounded by C, having mean zero and variance 1. Also, we may assume the diagonal elements are zero.

Step 5. Proof of the Semicircular Law

We will prove Theorem 2.5 by the moment method. For simplicity, we still use $\mathbf{W}_n$ and x_{ij} to denote the Wigner matrix and basic variables after truncation, centralization, and rescaling.

The semicircular distribution satisfies the Riesz condition. Therefore it is enough to show that the moments of the spectral distribution converge to the corresponding moments of the semicircular distribution almost surely. The k-th moment of the ESD of $\mathbf{W}_n$ is

$$\begin{aligned}\beta_k(\mathbf{W}_n) &= \beta_k(F^{\mathbf{W}_n}) = \int x^k \, dF^{\mathbf{W}_n}(x) \\ &= \frac{1}{n}\sum_{i=1}^n \lambda_i^k = \frac{1}{n}\mathrm{tr}(\mathbf{W}_n^k) = \frac{1}{n^{1+\frac{k}{2}}}\mathrm{tr}(\mathbf{X}_n^k) \\ &= \frac{1}{n^{1+\frac{k}{2}}}\sum_{\mathbf{i}} X(\mathbf{i}),\end{aligned} \tag{2.1.10}$$

where λ_i's are the eigenvalues of the matrix $\mathbf{W}_n$, $X(\mathbf{i}) = x_{i_1i_2}x_{i_2i_3}\cdots x_{i_ki_1}$, $\mathbf{i} = (i_1, \cdots, i_k)$, and the summation $\sum_{\mathbf{i}}$ runs over all possibilities that $\mathbf{i} \in \{1, \cdots, n\}^k$.

By applying the moment convergence theorem, we complete the proof of the semicircular law for the iid case by showing the following:

(1) $E[\beta_k(\mathbf{W}_n)]$ converges to the k-th moment β_k of the semicircular distribution, which are $\beta_{2m-1} = 0$ and $\beta_{2m} = (2m)!/m!(m+1)!$ given in Lemma 2.1.
(2) For each fixed k, $\sum_n \mathrm{Var}[\beta_k(\mathbf{W}_n)] < \infty$.

The Proof of (1); i.e., $\mathrm{E}[\beta_k(\mathbf{W}_n)] \to \beta_k$.

We have
$$\mathrm{E}[\beta_k(\mathbf{W}_n)] = \frac{1}{n^{1+k/2}}\sum \mathrm{E}X(\mathbf{i}).$$

For each vector $\mathbf{i}$, construct a graph $G(\mathbf{i})$ as in Subsection 2.1.2. To specify the graph, we rewrite $X(\mathbf{i}) = X(G(\mathbf{i}))$. The summation is taken over all sequences $\mathbf{i} = (i_1, i_2, \cdots, i_k) \in \{1, 2, \cdots, n\}^k$.

Note that isomorphic graphs correspond to equal terms. Thus, we first group the terms according to isomorphism classes and then split $\mathrm{E}[\beta_k(\mathbf{W}_n)]$ into three sums according to categories. Then

$$\mathrm{E}[\beta_k(\mathbf{W}_n)] = S_1 + S_2 + S_3,$$

where

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

in which the summation $\sum_{\Gamma(k,t)\in C_j}$ is taken on all canonical $\Gamma(k,t)$-graphs in category j and the summation $\sum_{G(\mathbf{i})\in\Gamma(k,t)}$ is taken on all isomorphic graphs for a given canonical graph.

By definition of the categories and by the assumptions on the entries of the random matrices, we have

$$S_2 = 0.$$

Since the random variables are bounded by C, the number of isomorphic graphs is less than n^t by Lemma 2.2, and $t \le (k+1)/2$ by Lemma 2.3, we conclude that

$$|S_3| \le n^{-1-k/2} O(n^t) = o(1).$$

If $k = 2m-1$, then $S_1 = 0$ since there are no terms in S_1. We consider the case where $k = 2m$. Since each edge coincides with an edge of opposite direction, each term in S_1 is $(\mathrm{E}|x_{12}|^2)^m = 1$. So, by Lemma 2.4,

$$\begin{aligned} S_1 &= n^{-1-m} \sum_{\Gamma(2m,t)\in C_1} n(n-1)\cdots(n-m) \\ &= \beta_{2m}\left(1-\frac{1}{m}\right)\cdots\left(1-\frac{m}{n}\right) \to \beta_{2m}. \end{aligned}$$

Assertion (1) is then proved.

The proof of (2). We only need to show that $\mathrm{Var}(\beta_k(\mathbf{W}_n))$ is summable for all fixed k. We have

$$\begin{aligned} &\mathrm{Var}(\beta_k(\mathbf{W}_n)) = E[|\beta_k(\mathbf{W}_n)|^2] - |E[\beta_k(\mathbf{W}_n)]|^2 \\ &= \frac{1}{n^{2+k}} \sum{}^{*} \{\mathrm{E}[X(\mathbf{i})X(\mathbf{j})] - \mathrm{E}[X(\mathbf{i})]\mathrm{E}[X(\mathbf{j})]\}, \end{aligned} \tag{2.1.11}$$

where $\mathbf{i} = (i_1, \cdots, i_k)$, $\mathbf{j} = (j_1, \cdots, j_k)$, and $\sum^*$ is taken over all possibilities for $\mathbf{i}$, $\mathbf{j} \in \{1, \cdots, n\}^k$. Here, the reader should notice that $\beta_k(\mathbf{W}_n)$ is real and hence the second equality in the above is meaningful, although the variables $X(\mathbf{i})$ and $X(\mathbf{j})$ are complex.

Using $\mathbf{i}$ and $\mathbf{j}$, one can construct two graphs $G(\mathbf{i})$ and $G(\mathbf{j})$, as in the proof of (1). If there are no coincident edges between $G(\mathbf{i})$ and $G(\mathbf{j})$, then $X(\mathbf{i})$ is independent of $X(\mathbf{j})$, and thus the corresponding term in the sum is 0. If the combined graph $G = G(\mathbf{i}) \cup G(\mathbf{j})$ has a single edge, then $\mathrm{E}[X(\mathbf{i})X(\mathbf{j})] = \mathrm{E}[X(\mathbf{i})]\mathrm{E}[X(\mathbf{j})] = 0$, and hence the corresponding term in (2.1.11) is also 0.

Now, suppose that G contains no single edges and the graph of noncoincident edges has a cycle. Then the noncoincident vertices of G are not more than k. If G contains no single edges and the graph of noncoincident edges has no cycles, then there is at least one edge with coincidence multiplicity greater than or equal to 4, and thus the number of noncoincident vertices is not larger than k. Also, each term in (2.1.11) is not larger than $2C^{2k}n^{-2-k}$. Consequently, we can conclude that

$$\mathrm{Var}(\beta_k(\mathbf{W}_n)) \leq K_k C^{2k} n^{-2}, \tag{2.1.12}$$

where K_k is a constant that depends on k only. This completes the proof of assertion (2).

The proof of Theorem 2.5 is then complete.

2.2 Generalizations to the Non-iid Case

Sometimes, it is of practical interest to consider the case where, for each n, the entries above or on the diagonal of $\mathbf{W}_n$ are independent complex random variables with mean zero and variance σ^2 (for simplicity we assume $\sigma = 1$ in the following), but may depend on n. For this case, we present the following theorem.

Theorem 2.9. *Suppose that $\mathbf{W}_n = \frac{1}{\sqrt{n}}\mathbf{X}_n$ is a Wigner matrix and the entries above or on the diagonal of $\mathbf{X}_n$ are independent but may be dependent on n and may not necessarily be identically distributed. Assume that all the entries of $\mathbf{X}_n$ are of mean zero and variance 1 and satisfy the condition that, for any constant $\eta > 0$,*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{jk} \mathrm{E}|x_{jk}^{(n)}|^2 I(|x_{jk}^{(n)}| \geq \eta\sqrt{n}) = 0. \tag{2.2.1}$$

Then, the ESD of $\mathbf{W}_n$ converges to the semicircular law almost surely.

Remark 2.10. In Girko's book [121], it is stated that condition (2.2.1) is necessary and sufficient for the conclusion of Theorem 2.9.

2.2.1 Proof of Theorem 2.9

Again, we need to truncate, remove diagonal entries, and renormalize before we use the MCT. Because the entries are not iid, we cannot truncate the entries at constant positions. Instead, we shall truncate them at $\eta_n\sqrt{n}$ for some sequence $\eta_n \downarrow 0$.

Step 1. Truncation

Note that Corollary A.41 may not be applicable in proving the almost sure asymptotic equivalence between the ESD of the original matrix and that of the truncated one, as was done in the last section. In this case, we shall use the rank inequality (see Theorem A.43) to truncate the variables.

Note that condition (2.2.1) is equivalent to: for any $\eta > 0$,

$$\lim_{n\to\infty} \frac{1}{\eta^2 n^2} \sum_{jk} \mathrm{E}|x_{jk}^{(n)}|^2 I(|x_{jk}^{(n)}| \geq \eta\sqrt{n}) = 0. \tag{2.2.2}$$

Thus, one can select a sequence $\eta_n \downarrow 0$ such that (2.2.2) remains true when η is replaced by η_n. Define $\widetilde{\mathbf{W}}_n = \frac{1}{\sqrt{n}} n(x_{ij}^{(n)} I(|x_{ij}^{(n)}| \leq \eta_n\sqrt{n})$. By using Theorem A.43, one has

$$\begin{aligned}\|F^{\mathbf{W}_n} - F^{\widetilde{\mathbf{W}}_n}\| &\leq \frac{1}{n}\mathrm{rank}(\mathbf{W}_n - \mathbf{W}_{n(\eta_n\sqrt{n})}) \\ &\leq \frac{2}{n} \sum_{1\leq i\leq j\leq n} I(|x_{ij}^{(n)}| \geq \eta_n\sqrt{n}). \end{aligned}\tag{2.2.3}$$

By condition (2.2.2), we have

$$\begin{aligned}&\mathrm{E}\left(\frac{1}{n}\sum_{1\leq i\leq j\leq n} I(|x_{ij}^{(n)}| \geq \eta_n\sqrt{n})\right) \\ &\leq \frac{2}{\eta_n^2 n^2}\sum_{jk} \mathrm{E}|x_{ij}^{(n)}|^2 I(|x_{ij}^{(n)}| \geq \eta_n\sqrt{n}) = o(1),\end{aligned}$$

and

$$\begin{aligned}&\mathrm{Var}\left(\frac{1}{n}\sum_{1\leq i\leq j\leq n} I(|x_{ij}^{(n)}| \geq \eta_n\sqrt{n})\right) \\ &\leq \frac{4}{\eta_n^2 n^3}\sum_{jk} \mathrm{E}|x_{ij}^{(n)}|^2 I(|x_{ij}^{(n)}| \geq \eta_n\sqrt{n}) = o(1/n).\end{aligned}$$

Then, applying Bernstein's inequality, for all small $\varepsilon > 0$ and large n, we have

$$\mathrm{P}\left(\frac{1}{n}\sum_{1\leq i\leq j\leq n} I(|x_{ij}^{(n)}| \geq \eta_n\sqrt{n}) \geq \varepsilon\right) \leq 2e^{-\varepsilon n}, \tag{2.2.4}$$

which is summable. Thus, by (2.2.3) and (2.2.4), to prove that with probability one $F^{\mathbf{W}_n}$ converges to the semicircular law, it suffices to show that with probability one $F^{\widetilde{\mathbf{W}}_n}$ converges to the semicircular law.

Step 2. Removing diagonal elements

Let $\widehat{\mathbf{W}}_n$ be the matrix $\widetilde{\mathbf{W}}_n$ with diagonal elements replaced by 0. Then, by Corollary A.41, we have

$$L^3\left(F^{\widetilde{\mathbf{W}}_n}, F^{\widehat{\mathbf{W}}_n}\right) \le \frac{1}{n^2}\sum_{k=1}^{n}|x_{kk}^{(n)}|^2 I(|x_{kk}^{(n)}| \le \eta_n\sqrt{n}) \le \eta_n^2 \to 0.$$

Step 3. Centralization

By Corollary A.41, it follows that

$$\begin{aligned} &L^3\left(F^{\widehat{\mathbf{W}}_n}, F^{\widehat{\mathbf{W}}_n - \mathrm{E}\widehat{\mathbf{W}}_n}\right) \\ &\le \frac{1}{n^2}\sum_{i\neq j}|\mathrm{E}(x_{ij}^{(n)} I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n}))|^2 \\ &\le \frac{1}{n^3\eta_n^2}\sum_{ij}\mathrm{E}|x_{jk}^{(n)}|^2 I(|x_{jk}^{(n)}| \ge \eta_n\sqrt{n}) \to 0. \end{aligned} \tag{2.2.5}$$

Step 4. Rescaling

Write $\widetilde{\widetilde{\mathbf{W}}}_n = \frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n$, where

$$\widetilde{\mathbf{X}}_n = \left(\frac{x_{ij}^{(n)} I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n}) - \mathrm{E}(x_{ij}^{(n)} I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n}))}{\sigma_{ij}}(1-\delta_{ij})\right),$$

$\sigma_{ij}^2 = \mathrm{E}|x_{ij}^{(n)} I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n}) - \mathrm{E}(x_{ij}^{(n)} I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n}))|^2$ and δ_{ij} is Kronecker's delta.

By Corollary A.41, it follows that

$$\begin{aligned} &L^3\left(F^{\widetilde{\widetilde{\mathbf{W}}}_n}, F^{\widehat{\mathbf{W}}_n - \mathrm{E}\widehat{\mathbf{W}}_n}\right) \\ &\le \frac{1}{n^2}\sum_{i\neq j}(1-\delta_{ij}^{-1})^2|x_{ij}^{(n)} I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n}) - \mathrm{E}(x_{ij}^{(n)} I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n}))|^2. \end{aligned}$$

Note that

$$\begin{aligned} &\mathrm{E}\left(\frac{1}{n^2}\sum_{i\neq j}(1-\delta_{ij}^{-1})^2|x_{ij}^{(n)} I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n}) - \mathrm{E}(x_{ij}^{(n)} I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n}))|^2\right) \\ &\le \frac{1}{n^2\eta_n^2}\sum_{ij}(1-\sigma_{ij})^2 \end{aligned}$$

$$\le \frac{1}{n^2\eta_n^2}\sum_{ij}(1-\sigma_{ij}^2)$$
$$\le \frac{1}{n^2\eta_n^2}\sum_{ij}[\mathrm{E}|x_{jk}^{(n)}|^2 I(|x_{jk}^{(n)}| \ge \eta_n\sqrt{n}) + \mathrm{E}^2|x_{jk}^{(n)}|I(|x_{jk}^{(n)}| \ge \eta_n\sqrt{n})] \to 0.$$

Also, we have[2]

$$\mathrm{E}\left|\frac{1}{n^2}\sum_{i\ne j}(1-\delta_{ij}^{-1})^2\left|x_{ij}^{(n)}I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n}) - \mathrm{E}(x_{ij}^{(n)}I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n}))\right|^2\right|^4$$
$$\le \frac{C}{n^8}\left[\sum_{i\ne j}\mathrm{E}|x_{ij}^{(n)}|^8 I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n}) + \left(\sum_{i\ne j}\mathrm{E}|x_{ij}^{(n)}|^4 I(|x_{ij}^{(n)}| \le \eta_n\sqrt{n})\right)^2\right]$$
$$\le Cn^{-2}[n^{-1}\eta_n^6 + \eta_n^4],$$

which is summable. From the two estimates above, we conclude that

$$L\left(F^{\widetilde{\mathbf{W}}_n}, F^{\widehat{\mathbf{W}}_n - \mathrm{E}\widehat{\mathbf{W}}_n}\right) \to 0, \text{ a.s.}$$

Step 5. Proof by MCT

Up to here, we have proved that we may truncate, centralize, and rescale the entries of the Wigner matrix at $\eta_n\sqrt{n}$ and remove the diagonal elements without changing the LSD. These four steps are almost the same as those we followed for the iid case.

Now, we assume that the variables are truncated at $\eta_n\sqrt{n}$ and then centralized and rescaled.

Again for simplicity, the truncated and centralized variables are still denoted by x_{ij}, We assume:

(i) The variables $\{x_{ij},\ 1 \le i < j \le n\}$ are independent and $x_{ii} = 0$.
(ii) $\mathrm{E}(x_{ij}) = 0$ and $\mathrm{Var}(x_{ij}) = 1$.
(iii) $|x_{ij}| \le \eta_n\sqrt{n}$.

Similar to what we did in the last section, in order to prove Theorem 2.9, we need to show that:

(1) $E[\beta_k(\mathbf{W}_n)]$ converges to the k-th moment β_k of the semicircular distribution.
(2) For each fixed k, $\sum_n \mathrm{E}|\beta_k(\mathbf{W}_n) - \mathrm{E}(\beta_k(\mathbf{W}_n))|^4 < \infty$.

The proof of (1)
Let $\mathbf{i} = (i_1, \cdots, i_k) \in \{1, \cdots, n\}^k$. As in the iid case, we write

[2] Here we use the elementary inequality $\mathrm{E}|\sum X_i|^{2k} \le C_k(\sum \mathrm{E}|X_i|^{2k} + (\sum \mathrm{E}|X_i|^2)^k)$ for some constant C_k if the X_i's are independent with zero means.

$$\mathrm{E}[\beta_k(\mathbf{W}_n)] = n^{-1-k/2} \sum_{\mathbf{i}} \mathrm{E}X(G(\mathbf{i})),$$

where $X(G(\mathbf{i})) = x_{i_1,i_2} x_{i_2,i_3} \cdots, x_{i_k,i_1}$, and $G(\mathbf{i})$ is the graph defined by $\mathbf{i}$.

By the same method for the iid case, we split $\mathrm{E}[\beta_k(\mathbf{W}_n)]$ into three sums according to the categories of graphs. We know that the terms in S_2 are all zero, that is, $S_2 = 0$.

We now show that $S_3 \to 0$. Split S_3 as $S_{31} + S_{32}$, where S_{31} consists of the terms corresponding to a $\Gamma_3(k,t)$-graph that contains at least one noncoincident edge with multiplicity greater than 2 and S_{32} is the sum of the remaining terms in S_3.

To estimate S_{31}, assume that the $\Gamma_3(k,t)$-graph contains ℓ noncoincident edges with multiplicities $\nu_1, \cdots, \nu_\ell$ among which at least one is greater than or equal to 3. Note that the multiplicities are subject to $\nu_1 + \cdots + \nu_\ell = k$. Also, each term in S_{31} is bounded by

$$n^{-1-k/2} \prod_{i=1}^{\ell} E|x_{a_i,b_i}|^{\nu_i} \le n^{-1-k/2} (\eta_n \sqrt{n})^{\sum(\nu_i - 2)} = n^{-1-\ell} \eta_n^{k-2\ell}.$$

Since the graph is connected and the number of its noncoincident edges is ℓ, the number of noncoincident vertices is not more than $\ell + 1$, which implies that the number of terms in S_{31} is not more than $n^{1+\ell}$. Therefore,

$$|S_{31}| \le C_k \eta_n^{k-2\ell} \to 0$$

since $k - 2\ell \ge 1$.

To estimate S_{32}, we note that the $\Gamma_3(k,t)$-graph contains exactly $k/2$ noncoincident edges, each with multiplicity 2 (thus k must be even). Then each term of S_{32} is bounded by $n^{-1-k/2}$. Since the graph is not in category 1, the graph of noncoincident edges must contain a cycle, and hence the number of noncoincident vertices is not more than $k/2$ and therefore

$$|S_{32}| \le Cn^{-1} \to 0.$$

Then, the evaluation of S_1 is exactly the same as in the iid case and hence is omitted. Hence, we complete the proof of $\mathrm{E}\beta_k(\mathbf{W}_n) \to \beta_k$.

The proof of (2)

Unlike in the proof of (2.1.11), the almost sure convergence cannot follow by estimating the second moment of $\beta_k(\mathbf{W}_n)$. We need to estimate its fourth moment as

$$\begin{aligned} &\mathrm{E}(\beta_k(\mathbf{W}_n) - \mathrm{E}(\beta_k(\mathbf{W}_n)))^4 \\ &= n^{-4-2k} \sum_{\mathbf{i}_j, j=1,2,3,4} \mathrm{E} \prod_{j=1}^{4} \left(X[\mathbf{i}_j] - \mathrm{E}(X[\mathbf{i}_j]) \right), \end{aligned} \tag{2.2.6}$$

where $\mathbf{i}_j$ is a vector of k integers not larger than n, $j = 1, 2, 3, 4$. As in the last section, for each $\mathbf{i}_j$, we construct a graph $G_j = G(\mathbf{i}_j)$.

Obviously, if, for some j, $G(\mathbf{i}_j)$ does not have any edges coincident with edges of the other three graphs, then the term in (2.2.6) equals 0 by independence. Also, if $G = \bigcup_{j=1}^4 G_j$ has a single edge, the term in (2.2.6) equals 0 by centralization.

Now, let us estimate the nonzero terms in (2.2.6). Assume that G has ℓ noncoincident edges with multiplicities $\nu_1, \cdots, \nu_\ell$, subject to the constraint $\nu_1 + \cdots + \nu_\ell = 4k$. Then, the term corresponding to G is bounded by

$$n^{-4-2k} \prod_{j=1}^{\ell} (\eta_n \sqrt{n})^{\nu_j - 2} = \eta_n^{4k-2\ell} n^{-4-\ell}.$$

Since the graph of noncoincident edges of G can have at most two pieces of connected subgraphs, the number of noncoincident vertices of G is not greater than $\ell + 2$. If $\ell = 2k$, then $\nu_1 = \cdots = \nu_\ell = 2$. Therefore, there is at least one noncoincident edge consisting of edges from two different subgraphs and hence there must be a cycle in the graph of noncoincident edges of G. Therefore,

$$\begin{aligned} &\mathrm{E}(\beta_k(\mathbf{W}_n) - \mathrm{E}(\beta_k(\mathbf{W}_n)))^4 \\ &\le C_k n^{-2k-4} \left[\sum_{\ell<2k} n^{\ell+2} (\eta_n^2 n)^{2k-\ell} + n^{2k+1} \right] \le C_k \eta_n n^{-2}, \end{aligned}$$

which is summable, and thus (2) is proved. Consequently, the proof of Theorem 2.9 is complete.

2.3 Semicircular Law by the Stieltjes Transform

As an illustration of the use of Stieltjes transforms, in this section we shall present a proof of Theorem 2.9 using them.

2.3.1 Stieltjes Transform of the Semicircular Law

Let $z = u + iv$ with $v > 0$ and $s(z)$ be the Stieltjes transform of the semicircular law. Then, we have

$$s(z) = \frac{1}{2\pi\sigma^2} \int_{-2\sigma}^{2\sigma} \frac{1}{x - z} \sqrt{4\sigma^2 - x^2}\, dx.$$

Letting $x = 2\sigma\cos y$, then

$$\begin{aligned} s(z) &= \frac{2}{\pi}\int_0^{\pi} \frac{1}{2\sigma\cos y - z}\sin^2 y dy \\ &= \frac{1}{\pi}\int_0^{2\pi} \frac{1}{2\sigma\frac{e^{iy}+e^{-iy}}{2} - z}\left(\frac{e^{iy}-e^{-iy}}{2i}\right)^2 dy \\ &= -\frac{1}{4i\pi}\oint_{|\zeta|=1} \frac{1}{\sigma(\zeta+\zeta^{-1}) - z}(\zeta-\zeta^{-1})^2\zeta^{-1}d\zeta \text{ (setting } \zeta = e^{iy}) \\ &= -\frac{1}{4i\pi}\oint_{|\zeta|=1} \frac{(\zeta^2-1)^2}{\zeta^2(\sigma\zeta^2+\sigma-z\zeta)}d\zeta. \end{aligned} \tag{2.3.1}$$

We will use the residue theorem to evaluate the integral. Note that the integrand has three poles, at $\zeta_0 = 0$, $\zeta_1 = \frac{z+\sqrt{z^2-4\sigma^2}}{2\sigma}$, and $\zeta_2 = \frac{z-\sqrt{z^2-4\sigma^2}}{2\sigma}$, where here, and throughout the book, the square root of a complex number is specified as the one with the positive imaginary part. By this convention, we have

$$\sqrt{z} = \mathrm{sign}(\Im z)\frac{|z|+z}{\sqrt{2(|z|+\Re z)}} \tag{2.3.2}$$

or

$$\Re(\sqrt{z}) = \frac{1}{\sqrt{2}}\mathrm{sign}(\Im z)\sqrt{|z|+\Re z} = \frac{\Im z}{\sqrt{2(|z|-\Re z)}}$$

and

$$\Im(\sqrt{z}) = \frac{1}{\sqrt{2}}\sqrt{|z|-\Re z} = \frac{|\Im z|}{\sqrt{2(|z|+\Re z)}}.$$

This shows that the real part of $\sqrt{z}$ has the same sign as the imaginary part of z. Applying this to ζ_1 and ζ_2, we find that the real part of $\sqrt{z^2-4\sigma^2}$ has the same sign as z, which implies that $|\zeta_1| > |\zeta_2|$. Since $\zeta_1\zeta_2 = 1$, we conclude that $|\zeta_2| < 1$ and thus the two poles 0 and ζ_1 of the integrand are in the disk $|z| \le 1$. By simple calculation, we find that the residues at these two poles are

$$\frac{z}{\sigma^2} \text{ and } \frac{(\zeta_2^2-1)^2}{\sigma\zeta_2^2(\zeta_2-\zeta_1)} = \sigma^{-1}(\zeta_2-\zeta_1) = -\sigma^{-2}\sqrt{z^2-4\sigma^2}.$$

Substituting these into the integral of (2.3.1), we obtain the following lemma.

Lemma 2.11. *The Stieltjes transform for the semicircular law with scale parameter σ^2 is*

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

2.3.2 Proof of Theorem 2.9

At first, we truncate the underlying variables at $\eta_n\sqrt{n}$ and remove the diagonal elements and then centralize and rescale the off-diagonal elements as done in Steps 1–4 in the last section. That is, we assume that:

(i) For $i \neq j$, $|x_{ij}| \leq \eta_n\sqrt{n}$ and $x_{ii} = 0$.
(ii) For all $i \neq j$, $\mathrm{E}x_{ij} = 0$, $\mathrm{E}|x_{ij}|^2 = \sigma^2$.
(iii) The variables $\{x_{ij}, i < j\}$ are independent.

For brevity, we assume $\sigma^2 = 1$ in what follows.

By definition, the Stieltjes transform of $F^{\mathbf{W}_n}$ is given by

$$s_n(z) = \frac{1}{n}\mathrm{tr}(\mathbf{W}_n - z\mathbf{I}_n)^{-1}. \tag{2.3.3}$$

We shall then proceed in our proof by taking the following three steps:

(i) For any fixed $z \in \mathbb{C}^+ = \{z, \Im(z) > 0\}$, $s_n(z) - \mathrm{E}s_n(z) \to 0$, a.s.
(ii) For any fixed $z \in \mathbb{C}^+$, $\mathrm{E}s_n(z) \to s(z)$, the Stieltjes transform of the semicircular law.
(iii) Outside a null set, $s_n(z) \to s(z)$ for every $z \in \mathbb{C}^+$.

Then, applying Theorem B.9, it follows that, except for this null set, $F^{\mathbf{W}_n} \to F$ weakly.

Step 1. Almost sure convergence of the random part
For the first step, we show that, for each fixed $z \in \mathbb{C}^+$,

$$s_n(z) - \mathrm{E}(s_n(z)) \to 0 \ \text{ a.s.} \tag{2.3.4}$$

We need the extended Burkholder inequality.

Lemma 2.12. *Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$. Then, for $p > 1$,*

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

Proof. Burkholder [67] proved the lemma for a real martingale difference sequence. Now, both $\{\Re X_k\}$ and $\{\Im X_k\}$ are martingale difference sequences. Thus, we have

$$\begin{aligned}
\mathrm{E}\left|\sum X_k\right|^p &\leq C_p\left[\mathrm{E}\left|\sum \Re X_k\right|^p + \mathrm{E}\left|\sum \Im X_k\right|^p\right] \\
&\leq C_p\left[K_p\mathrm{E}\left(\sum |\Re X_k|^2\right)^{p/2} + K_p\mathrm{E}\left(\sum |\Im X_k|^2\right)^{p/2}\right] \\
&\leq 2C_pK_p\mathrm{E}\left(\sum |X_k|^2\right)^{p/2},
\end{aligned}$$

where $C_p = 2^{p-1}$. This lemma is proved.

For later use, we introduce here another inequality proved in [67].

Lemma 2.13. *Let $\{X_k\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\mathcal{F}_k$, and let E_k denote conditional expectation w.r.t. $\mathcal{F}_k$. Then, for $p \geq 2$,*

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

Similar to Lemma 2.12, Burkholder proved this lemma for the real case. Using the same technique as in the proof of Lemma 2.12, one may easily extend the Burkholder inequality to the complex case.

Now, we proceed to the proof of the almost sure convergence (2.3.4). Denote by $\mathrm{E}_k(\cdot)$ conditional expectation with respect to the σ-field generated by the random variables $\{x_{ij}, i, j > k\}$, with the convention that $\mathrm{E}_n s_n(z) = \mathrm{E}s_n(z)$ and $\mathrm{E}_0 s_n(z) = s_n(z)$. Then, we have

$$s_n(z) - \mathrm{E}(s_n(z)) = \sum_{k=1}^{n} [\mathrm{E}_{k-1}(s_n(z)) - \mathrm{E}_k(s_n(z))] := \sum_{k=1}^{n} \gamma_k,$$

where, by Theorem A.5,

$$\begin{aligned}\gamma_k &= \frac{1}{n}\left(\mathrm{E}_{k-1}\mathrm{tr}(\mathbf{W}_n - z\mathbf{I})^{-1} - \mathrm{E}_k\mathrm{tr}(\mathbf{W}_n - z\mathbf{I})^{-1}\right) \\ &= \frac{1}{n}\left(\mathrm{E}_{k-1}[\mathrm{tr}(\mathbf{W}_n - z\mathbf{I})^{-1} - \mathrm{tr}(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}]\right. \\ &\qquad \left. -\mathrm{E}_k[\mathrm{tr}(\mathbf{W}_n - z\mathbf{I})^{-1} - \mathrm{tr}(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}]\right) \\ &= \frac{1}{n}\left(\mathrm{E}_{k-1}\frac{1 + \boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k}{-z - \boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k}\right. \\ &\qquad \left. -\mathrm{E}_k\frac{1 + \boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k}{-z - \boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k}\right),\end{aligned}$$

where $\mathbf{W}_k$ is the matrix obtained from $\mathbf{W}_n$ with the k-th row and column removed and $\boldsymbol{\alpha}_k$ is the k-th column of $\mathbf{W}_n$ with the k-th element removed.

Note that

$$\begin{aligned}&|1 + \boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k| \\ &\leq 1 + \boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}(\mathbf{W}_k - \bar{z}\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k \\ &= v^{-1}\Im(z + \boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k)\end{aligned}$$

which implies that

$$|\gamma_k| \leq 2/nv.$$

Noting that $\{\gamma_k\}$ forms a martingale difference sequence, applying Lemma 2.12 for $p = 4$, we have

$$\begin{aligned}
\mathrm{E}|s_n(z) - \mathrm{E}(s_n(z))|^4 &\le K_4 \mathrm{E}\left(\sum_{k=1}^{n} |\gamma_k|^2\right)^2 \\
&\le K_4 \left(\sum_{k=1}^{n} \frac{2}{n^2 v^2}\right)^2 \\
&\le \frac{4K_4}{n^2 v^4}.
\end{aligned}$$

By the Borel-Cantelli lemma, we know that, for each fixed $z \in \mathbb{C}^+$,

$$s_n(z) - \mathrm{E}(s_n(z)) \to 0, \text{ a.s.}$$

Step 2. Convergence of the expected Stieltjes transform
By Theorem A.4, we have

$$\begin{aligned}
s_n(z) &= \frac{1}{n}\mathrm{tr}(\mathbf{W}_n - z\mathbf{I}_n)^{-1} \\
&= \frac{1}{n}\sum_{k=1}^{n} \frac{1}{-z - \boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k}.
\end{aligned} \tag{2.3.5}$$

Write $\varepsilon_k = \mathrm{E}s_n(z) - \boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k$. Then we have

$$\begin{aligned}
\mathrm{E}s_n(z) &= \frac{1}{n}\sum_{k=1}^{n} \mathrm{E}\frac{1}{-z - \mathrm{E}s_n(z) + \varepsilon_k} \\
&= -\frac{1}{z + \mathrm{E}s_n(z)} + \delta_n,
\end{aligned} \tag{2.3.6}$$

where

$$\delta_n = \frac{1}{n}\sum_{k=1}^{n} \mathrm{E}\left(\frac{\varepsilon_k}{(z + \mathrm{E}s_n(z))(-z - \mathrm{E}s_n(z) + \varepsilon_k)}\right).$$

Solving equation (2.3.6), we obtain two solutions:

$$\frac{1}{2}(-z + \delta_n \pm \sqrt{(z + \delta_n)^2 - 4}).$$

We show that

$$\mathrm{E}s_n(z) = \frac{1}{2}(-z + \delta_n + \sqrt{(z + \delta_n)^2 - 4}). \tag{2.3.7}$$

When fixing $\Re z$ and letting $\Im z = v \to \infty$, we have $\mathrm{E}s_n(z) \to 0$, which implies that $\delta_n \to 0$. Consequently,

$$\Im\Big(\frac{1}{2}(-z+\delta_n-\sqrt{(z+\delta_n)^2-4})\Big)\le -\frac{v-|\delta_n|}{2}\to-\infty,$$

which cannot be $\mathrm{E}s_n(z)$ since it violates the property that $\Im s_n(z)\ge 0$. Thus, assertion (2.3.7) is true when v is large. Now, we claim that assertion (2.3.7) is true for all $z\in\mathbb{C}^+$.

It is easy to see that $\mathrm{E}s_n(z)$ and $\frac{1}{2}(-z+\delta_n\pm\sqrt{(z+\delta_n)^2-4})$ are continuous functions on the upper half plane $\mathbb{C}^+$. If $\mathrm{E}s_n(z)$ takes a value on the branch $\frac{1}{2}(-z+\delta_n-\sqrt{(z+\delta_n)^2-4})$ for some z, then the two branches $\frac{1}{2}(-z+\delta_n\pm\sqrt{(z+\delta_n)^2-4})$ should cross each other at some point $z_0\in\mathbb{C}^+$. At this point, we would have $\sqrt{(z_0+\delta_n)^2-4}=0$ and hence $\mathrm{E}s_n(z_0)$ has to be one of the following:

$$\frac{1}{2}(-z_0+\delta_n)=\frac{1}{2}(-2z_0\pm 2).$$

However, both of the two values above have negative imaginary parts. This contradiction leads to the truth of (2.3.7).

From (2.3.7), to prove $\mathrm{E}s_n(z)\to s(z)$, it suffices to show that

$$\delta_n\to 0. \tag{2.3.8}$$

Now, rewrite

$$\begin{aligned}\delta_n&=-\frac{1}{n}\sum_{k=1}^n\frac{\mathrm{E}(\varepsilon_k)}{(z+\mathrm{E}s_n(z))^2}+\frac{1}{n}\sum_{k=1}^n\mathrm{E}\left(\frac{\varepsilon_k^2}{(z+\mathrm{E}s_n(z))^2(-z-\mathrm{E}s_n(z)+\varepsilon_k)}\right)\\&=J_1+J_2.\end{aligned}$$

By (A.1.10) and (A.1.12), we have

$$\begin{aligned}|\mathrm{E}\varepsilon_k|&=\left|\frac{1}{n}\mathrm{E}(\mathrm{tr}(\mathbf{W}_n-z\mathbf{I})^{-1}-\mathrm{tr}(\mathbf{W}_k-z\mathbf{I}_{n-1})^{-1})\right|\\&=\left|\frac{1}{n}\cdot\mathrm{E}\frac{1+\boldsymbol{\alpha}_k^*(\mathbf{W}_k-z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k}{-z-\boldsymbol{\alpha}_k^*(\mathbf{W}_k-z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k}\right|\le\frac{1}{nv}.\end{aligned}$$

Note that

$$|z+\mathrm{E}s_n(z)|\ge\Im(z+\mathrm{E}s_n(z))=v+\mathrm{E}(\Im(s_n(z))\ge v.$$

Therefore, for any fixed $z\in\mathbb{C}^+$,

$$|J_1|\le\frac{1}{nv^3}\to 0.$$

On the other hand, we have

$$\begin{aligned}&|-z-\mathrm{E}s_n(z)+\varepsilon_k|=|-z-\boldsymbol{\alpha}_k^*(\mathbf{W}_k-z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k|\\&\ge\Im(z+\boldsymbol{\alpha}_k^*(\mathbf{W}_k-z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k)\end{aligned}$$

$$= v(1 + \boldsymbol{\alpha}_k^*((\mathbf{W}_k - z\mathbf{I}_{n-1})(\mathbf{W}_k - \bar{z}\mathbf{I}_{n-1}))^{-1}\boldsymbol{\alpha}_k) \geq v.$$

To prove $J_2 \to 0$, it is sufficient to show that

$$\max_k \mathrm{E}|\varepsilon_k|^2 \to 0.$$

Write $(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1} = (b_{ij})_{i,j\leq n-1}$. We then have

$$\begin{aligned}
\mathrm{E}|\varepsilon_k - \mathrm{E}\varepsilon_k|^2 &= \mathrm{E}|\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k - \frac{1}{n}\mathrm{Etr}((\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1})|^2 \\
&= \mathrm{E}|\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k - \frac{1}{n}\mathrm{tr}((\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1})|^2 \\
&\quad + \mathrm{E}\left|\frac{1}{n}\mathrm{tr}((\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}) - \frac{1}{n}\mathrm{Etr}((\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1})\right|^2.
\end{aligned}$$

By elementary calculations, we have

$$\begin{aligned}
&\mathrm{E}|\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k - \frac{1}{n}\mathrm{tr}((\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1})|^2 \\
&= \frac{1}{n^2}\left[\sum_{ij\neq k}[\mathrm{E}|b_{ij}|^2\mathrm{E}|x_{ik}|^2\mathrm{E}|x_{jk}|^2 + \mathrm{E}b_{ij}^2\mathrm{E}x_{ik}^2\mathrm{E}x_{jk}^2] + \sum_{i\neq k}\mathrm{E}|b_{ii}^2|(\mathrm{E}|x_{ik}^4| - 1)\right] \\
&\leq \frac{2}{n^2}\sum_{ij}\mathrm{E}|b_{ij}|^2 + \frac{\eta_n^2}{n}\sum_{i\neq k}\mathrm{E}|b_{ii}|^2 \\
&= \frac{2}{n^2}\mathrm{Etr}((\mathbf{W}_k - z\mathbf{I}_{n-1})(\mathbf{W}_k - \bar{z}\mathbf{I}_{n-1}))^{-1} + \frac{\eta_n^2}{n}\sum_{i\neq k}\mathrm{E}|b_{ii}|^2 \\
&\leq \frac{2}{nv^2} + \eta_n^2 \to 0. \qquad (2.3.9)
\end{aligned}$$

By Theorem A.5, one can prove that

$$\mathrm{E}\left|\frac{1}{n}\mathrm{tr}((\mathbf{W}_n - z\mathbf{I}_{n-1})^{-1}) - \frac{1}{n}\mathrm{Etr}((\mathbf{W}_n - z\mathbf{I}_{n-1})^{-1})\right|^2 \leq 1/n^2v^2.$$

Then, the assertion $J_2 \to 0$ follows from the estimates above and the fact that

$$\mathrm{E}|\varepsilon_n|^2 = \mathrm{E}|\varepsilon_n - \mathrm{E}\varepsilon_n|^2 + |\mathrm{E}\varepsilon_n|^2.$$

The proof of the mean convergence is complete.

Step 3. Completion of the proof of Theorem 2.9
In this step, we need Vitali's convergence theorem.

Lemma 2.14. *Let $f_1, f_2, \cdots$ be analytic in D, a connected open set of $\mathbb{C}$, satisfying $|f_n(z)| \leq M$ for every n and z in D, and $f_n(z)$ converges as $n \to \infty$ for each z in a subset of D having a limit point in D. Then there exists a* *function f analytic in D for which $f_n(z) \to f(z)$ and $f_n'(z) \to f'(z)$ for all $z \in D$. Moreover, on any set bounded by a contour interior to D, the convergence is uniform and $\{f_n'(z)\}$ is uniformly bounded.*

Proof. The conclusions on $\{f_n\}$ are from Vitali's convergence theorem (see Titchmarsh [275], p. 168). Those on $\{f_n'\}$ follow from the dominated convergence theorem (d.c.t.) and the identity

$$f_n'(z) = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{f_n(w)}{(w-z)^2} dw,$$

where $\mathcal{C}$ is a contour in D and enclosing z. The proof of the lemma is complete.

By Steps 1 and 2, for any fixed $z \in \mathbb{C}^+$, we have

$$s_n(z) \to s(z), \quad \text{a.s.},$$

where $s(z)$ is the Stieltjes transform of the standard semicircular law. That is, for each $z \in \mathbb{C}^+$, there exists a null set N_z (i.e., $P(N_z) = 0$) such that

$$s_n(z, \omega) \to s(z) \text{ for all } \ \omega \in N_z^c.$$

Now, let $\mathbb{C}_0^+ = \{z_m\}$ be a dense subset of $\mathbb{C}^+$ (e.g., all z of rational real and imaginary parts) and let $N = \cup N_{z_m}$. Then

$$s_n(z, \omega) \to s(z) \text{ for all } \ \omega \in N^c \text{ and } z \in \mathbb{C}_0^+.$$

Let $\mathbb{C}_m^+ = \{z \in \mathbb{C}^+, \Im z > 1/m, \ |z| \le m\}$. When $z \in \mathbb{C}_m^+$, we have $|s_n(z)| \le m$. Applying Lemma 2.14, we have

$$s_n(z, \omega) \to s(z) \text{ for all } \ \omega \in N^c \text{ and } z \in \mathbb{C}_m^+.$$

Since the convergence above holds for every m, we conclude that

$$s_n(z, \omega) \to s(z) \text{ for all } \ \omega \in N^c \text{ and } z \in \mathbb{C}^+.$$

Applying Theorem B.9, we conclude that

$$F^{\mathbf{W}_n} \xrightarrow{w} F, \ \text{a.s.}$$

Chapter 3
Sample Covariance Matrices and the Marčenko-Pastur Law

The sample covariance matrix is a most important random matrix in multivariate statistical inference. It is fundamental in hypothesis testing, principal component analysis, factor analysis, and discrimination analysis. Many test statistics are defined by its eigenvalues.

The definition of a sample covariance matrix is as follows. Suppose that $\{x_{jk},\ j,k = 1,2,\cdots\}$ is a double array of iid complex random variables with mean zero and variance σ^2. Write $\mathbf{x}_j = (x_{1j},\cdots,x_{pj})'$ and $\mathbf{X} = (\mathbf{x}_1,\cdots,\mathbf{x}_n)$. The sample covariance matrix is defined by

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

where $\bar{\mathbf{x}} = \frac{1}{n}\sum \mathbf{x}_j$.

However, in most cases of spectral analysis of large dimensional random matrices, the sample covariance matrix is simply defined as

$$\mathbf{S} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k\mathbf{x}_k^* = \frac{1}{n}\mathbf{X}\mathbf{X}^* \tag{3.0.1}$$

because the $\bar{\mathbf{x}}\bar{\mathbf{x}}^*$ is a rank 1 matrix and hence the removal of $\bar{\mathbf{x}}$ does not affect the LSD due to Theorem A.44.

In spectral analysis of large dimensional sample covariance matrices, it is usual to assume that the dimension p tends to infinity proportionally to the degrees of freedom n, namely $p/n \to y \in (0,\infty)$.

The first success in finding the limiting spectral distribution of the large sample covariance matrix $\mathbf{S}_n$ (named the Marčenko-Pastur (M-P) law by some authors) was due to Marčenko and Pastur [201]. Succeeding work was done in Bai and Yin [37], Grenander and Silverstein [137], Jonsson [169], Silverstein [256], Wachter [291], and Yin [300]. When the entries of $\mathbf{X}$ are not independent, Yin and Krishnaiah [303] investigated the limiting spectral distribution of $\mathbf{S}$ when the underlying distribution is isotropic. The theorem

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_3,

in the next section is a consequence of a result in Yin [300], where the real case is considered.

3.1 M-P Law for the iid Case

3.1.1 Moments of the M-P Law

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

$$p_y(x) = \begin{cases} \frac{1}{2\pi xy\sigma^2}\sqrt{(b-x)(x-a)}, & \text{if } a \le x \le b, \\ 0, & \text{otherwise,} \end{cases} \tag{3.1.1}$$

and has a point mass $1 - 1/y$ at the origin if $y > 1$, where $a = \sigma^2(1-\sqrt{y})^2$ and $b = \sigma^2(1+\sqrt{y})^2$. Here, the constant y is the dimension to sample size ratio index and σ^2 is the scale parameter. If $\sigma^2 = 1$, the M-P law is said to be the standard M-P law.

The moments $\beta_k = \beta_k(y, \sigma^2) = \int_a^b x^k p_y(x)dx$. In the following, we shall determine the explicit expression of β_k. Note that, for all $k \ge 1$,

$$\beta_k(y, \sigma^2) = \sigma^{2k}\beta_k(y, 1).$$

We need only compute β_k for the standard M-P law.

Lemma 3.1. *We have*

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

Proof. By definition,

$$\begin{aligned}
\beta_k &= \frac{1}{2\pi y}\int_a^b x^{k-1}\sqrt{(b-x)(x-a)}dx \\
&= \frac{1}{2\pi y}\int_{-2\sqrt{y}}^{2\sqrt{y}} (1+y+z)^{k-1}\sqrt{4y-z^2}dz \quad (\text{with } \ x = 1+y+z) \\
&= \frac{1}{2\pi y}\sum_{\ell=0}^{k-1}\binom{k-1}{\ell}(1+y)^{k-1-\ell}\int_{-2\sqrt{y}}^{2\sqrt{y}} z^\ell\sqrt{4y-z^2}dz \\
&= \frac{1}{2\pi y}\sum_{\ell=0}^{[(k-1)/2]}\binom{k-1}{2\ell}(1+y)^{k-1-2\ell}(4y)^{\ell+1}\int_{-1}^{1} u^{2\ell}\sqrt{1-u^2}du, \\
&\qquad\qquad (\text{by setting } z = 2\sqrt{y}u)
\end{aligned}$$

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

$$\text{(setting } u = \sqrt{w}\text{)}$$

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

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

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

$$= \sum_{\ell=0}^{[(k-1)/2]} \sum_{r=\ell}^{k-1-\ell} \frac{(k-1)!}{\ell!(\ell+1)!(r-\ell)!(k-1-r-\ell)!} y^r$$

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

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

By definition, we have $\beta_{2k} \leq b^{2k} = (1+\sqrt{y})^{4k}$. From this, it is easy to see that the Carleman condition is satisfied.

3.1.2 Some Lemmas on Graph Theory and Combinatorics

To use the moment method to show the convergence of the ESD of large dimensional sample covariance matrices to the M-P law, we need to define a class of Δ-graphs and establish some lemmas concerning some counting problems related to Δ-graphs.

Suppose that $i_1, \cdots, i_k$ are k positive integers (not necessarily distinct) not greater than p and $j_1, \cdots, j_k$ are k positive integers (not necessarily distinct) not larger than n. A Δ-graph is defined as follows. *Draw two parallel lines, referring to the I line and the J line. Plot $i_1, \cdots, i_k$ on the I line and $j_1, \cdots, j_k$ on the J line, and draw k (down) edges from i_u to j_u, $u = 1, \cdots, k$ and k (up) edges from j_u to i_{u+1}, $u = 1, \cdots, k$ (with the convention that $i_{k+1} = i_1$).* The graph is denoted by $G(\mathbf{i}, \mathbf{j})$, where $\mathbf{i} = (i_1, \cdots, i_k)$ and $\mathbf{j} = (j_1, \cdots, j_k)$. An example of a Δ-graph is shown in Fig. 3.1.

Two graphs are said to be isomorphic if one becomes the other by a suitable permutation on $(1, 2, \cdots, p)$ and a suitable permutation on $(1, 2, \cdots, n)$.

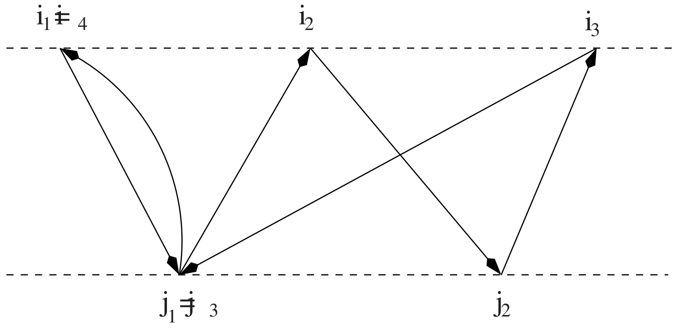

Fig. 3.1 A Δ-graph.

For each isomorphism class, there is only one graph, called *canonical*, satisfying $i_1 = j_1 = 1$, $i_u \le \max\{i_1, \cdots, i_{u-1}\} + 1$, and $j_u \le \max\{j_1, \cdots, j_{u-1}\} + 1$. A canonical Δ-graph $G(\mathbf{i}, \mathbf{j})$ is denoted by $\Delta(k, r, s)$ if G has $r + 1$ noncoincident I-vertices and s noncoincident J-vertices. A canonical $\Delta(k, r, s)$ can be directly defined in the following way:

1. Its vertex set $V = V_I + V_J$, where $V_I = \{1, \cdots, r+1\}$, called the ***I*-vertices**, and $V_J = \{1, \cdots, s\}$, called the ***J*-vertices**.
2. There are two functions, $f : \{1, \cdots, k\} \mapsto \{1, \cdots, r+1\}$ and $g : \{1, \cdots, k\} \mapsto \{1, \cdots, s\}$, satisfying

$$f(1) = 1 = g(1) = f(k + 1),$$
$$f(i) \le \max\{f(1), \cdots, f(i - 1)\} + 1,$$
$$g(j) \le \max\{g(1), \cdots, g(j - 1)\} + 1.$$

3. Its edge set $E = \{e_{1d}, e_{1u}, \cdots, e_{kd}, e_{ku}\}$, where $e_{1d}, \cdots, e_{kd}$ are called the down edges and $e_{1u}, \cdots, e_{ku}$ are called the up edges.
4. $F(e_{jd}) = (f(j), g(j))$ and $F(e_{ju}) = (g(j), f(j + 1))$ for $j = 1, \cdots, k$.

In the case where $f(j+1) = \max\{f(1), \cdots, f(j)\}+1$, the edge $e_{j,u}$ is called an up innovation, and in the case where $g(j) = \max\{g(1), \cdots, g(j - 1)\} + 1$, the edge $e_{j,d}$ is called a down innovation. Intuitively, an up innovation leads to a new I-vertex and a down innovation leads to a new J-vertex. We make the convention that the first down edge is a down innovation and the last up edge is not an innovation.

Similar to the Γ-graphs, we classify $\Delta(k, r, s)$-graphs into three categories:

Category 1 (denoted by $\Delta_1(k, r)$): Δ-graphs in which each down edge must coincide with one and only one up edge. If we glue the coincident edges, the resulting graph is a tree of k edges. In this category, $r + s = k$ and thus s is suppressed for simplicity.

Category 2 ($\Delta_2(k, r, s)$): Δ-graphs that contain at least one single edge.

Category 3 ($\Delta_3(k,r,s)$): Δ-graphs that do not belong to $\Delta_1(k,r)$ or $\Delta_2(k,r,s)$.

Similar to the arguments given in Subsection 2.1.2, the number of graphs in each isomorphism class for a given canonical $\Delta(k,r,s)$ is given by the following lemma.

Lemma 3.2. *For a given k, r, and s, the number of graphs in the isomorphism class for each canonical $\Delta(k,r,s)$-graph is*

$$p(p-1)\cdots(p-r)n(n-1)\cdots(n-s+1)=p^{r+1}n^s[1+O(n^{-1})].$$

For a Δ_3-graph, we have the following lemma.

Lemma 3.3. *The total number of noncoincident vertices of a $\Delta_3(k,r,s)$-graph is less than or equal to k.*

Proof. Let G be a graph of $\Delta_3(k,r,s)$. Note that any Δ-graph is connected. Since G is not in category 2, it does not contain single edges and hence the number of noncoincident edges is not larger than k. If the number of noncoincident edges is less than k, then the lemma is proved. If the number of noncoincident edges is exactly k, the graph of noncoincident edges must contain a cycle since it is not in category 1. In this case, the number of noncoincident vertices is also not larger than k and the lemma is proved.

A more difficult task is to count the number of $\Delta_1(k,r)$-graphs, as given in the following lemma.

Lemma 3.4. *For k and r, the number of $\Delta_1(k,r)$-graphs is*

$$\frac{1}{r+1}\binom{k}{r}\binom{k-1}{r}.$$

Proof. Define two characteristic sequences $\{u_1,\cdots,u_k\}$ and $\{d_1,\cdots,d_k\}$ of the graph G by

$$u_\ell=\begin{cases}1, & \text{if } f(\ell+1)=\max\{f(1),\cdots,f(\ell)\}+1,\\ 0, & \text{otherwise,}\end{cases}$$

and

$$d_\ell=\begin{cases}-1, & \text{if } f(\ell)\notin\{1,f(\ell+1),\cdots,f(k)\},\\ 0, & \text{otherwise.}\end{cases}$$

We can interpret the intuitive meaning of the characteristic sequences as follows: $u_\ell=1$ if and only if the ℓ-th up edge is an up innovation and $d_\ell=-1$ if and only if the ℓ-th down edge coincides with the up innovation that leads to this I-vertex. An example with $r=2$ and $s=3$ is given in Fig. 3.2.

By definition, we always have $u_k=0$, and since $f(1)=1$, we always have $d_1=0$. For a $\Delta_1(k,r)$-graph, there are exactly r up innovations and hence there are r u-variables equal to 1. Since there are r I-vertices other than 1, there are then r d-variables equal to -1.

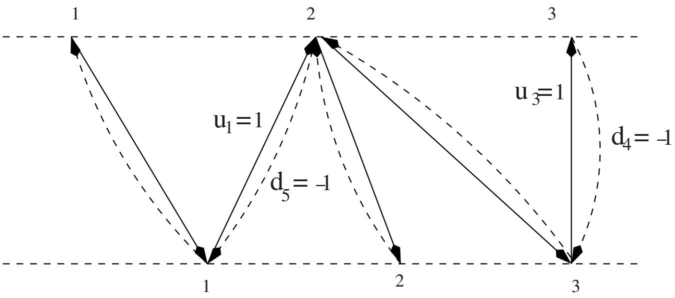

Fig. 3.2 Definition of (u, d) sequence

From its definition, one sees that $d_\ell = -1$ means that after plotting the ℓ-th down edge $(f(\ell), g(\ell))$, the future path will never revisit the I-vertex $f(\ell)$. This means that the edge $(f(\ell), g(\ell))$ must coincide with the up innovation leading to the vertex $f(\ell)$. Since there are $s = k - r$ down innovations to lead out the s J-vertices, $d_\ell = 0$ therefore implies that the edge $(f(\ell), g(\ell))$ must be a down innovation.

From the argument above, one sees that $d_\ell = -1$ must follow a $u_j = 1$ for some $j < \ell$. Therefore, the two sequences should satisfy the restriction

$$u_1 + \cdots + u_{\ell-1} + d_2 + \cdots + d_\ell \geq 0, \quad \ell = 2, \cdots, k. \tag{3.1.2}$$

From the definition of the characteristic sequences, each $\Delta_1(k, r)$-graph defines a pair of characteristic sequences. Conversely, we shall show that each pair of characteristic sequences satisfying (3.1.2) uniquely defines a $\Delta_1(k, r)$-graph. In other words, the functions f and g in the definition of the Δ-graph G are uniquely determined by the two sequences of $\{u_\ell\}$ and $\{d_\ell\}$.

At first, we notice that $u_\ell = 1$ implies that $e_{\ell,u}$ is an up innovation and thus

$$f(\ell + 1) = 1 + {}^{\#}\{j \leq \ell, u_j = 1\}.$$

Similarly, $d_\ell = 0$ implies that $e_{\ell,d}$ is a down innovation and thus

$$g(\ell) = {}^{\#}\{j \leq \ell, d_j = 0\}.$$

However, it is not easy to define the values of f and g at other points. So, we will directly plot the $\Delta_1(k, r)$-graph from the two characteristic sequences.

Since $d_1 = 0$ and hence $e_{1,d}$ is a down innovation, we draw $e_{1,d}$ from the I-vertex 1 to the J-vertex 1. If $u_1 = 0$, then $e_{1,u}$ is not an up innovation and thus the path must return the I-vertex 1 from the J-vertex 1; i.e., $f(2) = 1$. If $u_1 = 1$, $e_{1,u}$ is an up innovation leading to the new I-vertex 2; that is, $f(2) = 2$. Thus, the edge $e_{1,u}$ is from the J-vertex 1 to the I-vertex 2. This shows that the first pair of down and up edges are uniquely determined by u_1 and d_1. Suppose that the first ℓ pairs of the down and up edges are uniquely determined by the sequences $\{u_1, \cdots, u_\ell\}$ and $\{d_1, \cdots, d_\ell\}$. Also, suppose that the subgraph G_ℓ of the first ℓ pairs of down and up edges satisfies the following properties

1. G_ℓ is connected, and the undirectional noncoincident edges of G_ℓ form a tree.
2. If the end vertex $f(\ell+1)$ of $e_{\ell,u}$ is the I-vertex 1, then each down edge of G_ℓ coincides with an up edge of G_ℓ. Thus, G_ℓ does not have single innovations.
   If the end vertex $f(\ell+1)$ of $e_{\ell,u}$ is not the I-vertex 1, then from the I-vertex 1 to the I-vertex $f(\ell+1)$ there is only one path (chain without cycles) of down-up-down-up single innovations and all other down edges coincide with an up edge.

To draw the $\ell+1$-st pair of down and up edges, we consider the following four cases.

Case 1. $d_{\ell+1} = 0$ and $u_{\ell+1} = 1$. Then both edges of the $\ell+1$-st pair are innovations. Thus, adding the two innovations to G_ℓ, the resulting subgraph $G_{\ell+1}$ satisfies the two properties above with the path of down-up single innovations that consists of the original path of single innovations and the two new innovations. See Case 1 in Fig. 3.3.

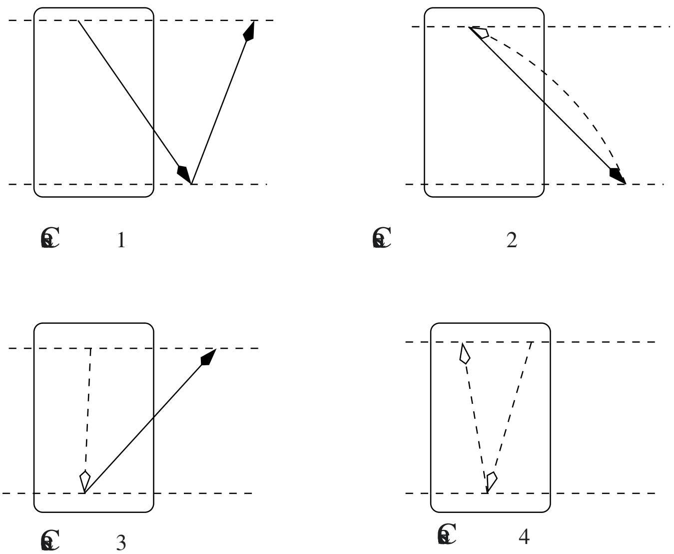

Fig. 3.3 Examples of the four cases. In the four graphs, the rectangle denotes the subgraph G_ℓ, solid arrows are new innovations, and broken arrows are new T_3 edges.

Case 2. $d_{\ell+1} = 0$ and $u_{\ell+1} = 0$. Then, $e_{\ell+1,d}$ is a down innovation and $e_{\ell+1,u}$ coincides with $e_{\ell+1,d}$. See Case 2 in Fig. 3.3. Thus, for the subgraph $G_{\ell+1}$, the two properties above can be trivially seen from the hypothesis for the subgraph G_ℓ. The single innovation chain of $G_{\ell+1}$ is exactly the same as that of G_ℓ.

Case 3. $d_{\ell+1} = -1$ and $u_{\ell+1} = 1$. In this case, by (3.1.2) we have

$$u_1 + \cdots + u_\ell + d_2 + \cdots + d_\ell \geq 1$$

which implies that the total number of I-vertices of G_ℓ other than 1 (i.e., $u_1 + \cdots + u_\ell$) is greater than the number of I-vertices of G_ℓ from which the graph ultimately leaves (i.e., $d_2+\cdots+d_\ell$). Therefore, $f(\ell+1) \neq 1$ because G_ℓ must contain single innovations by property 2. Then there must be a single up innovation leading to the vertex $f(\ell+1)$ and thus we can draw the down edge $e_{\ell+1,d}$ coincident with this up innovation. Then, the next up innovation $e_{\ell,u}$ starts from the end vertex to $g(\ell+1)$. See case 3 in Fig. 3.3. It is easy to see that the two properties above hold with the path of single innovations that is the original one with the last up innovation replaced by $e_{\ell+1,u}$.

Case 4. $d_{\ell+1} = -1$ and $u_{\ell+1} = 0$. Then, as discussed in case 3, $e_{\ell+1,d}$ can be drawn to coincide with the only up innovation ended at $f(\ell+1)$. Prior to this up innovation, there must be a single down innovation with which the up edge $e_{\ell,u}$ can be drawn to coincide. If the path of single innovations of G_ℓ has only one pair of down-up innovations, then $f(\ell+2) = 1$ and hence $G_{\ell+1}$ has no single innovations. If the path of single innovations of G_ℓ has more than two edges, then the remaining part of the path of single innovations of G_ℓ, with the last two innovations removed, forms a path of single innovations of $G_{\ell+1}$. See case 1 in Fig. 3.3. In either case, two properties for $G_{\ell+1}$ hold.

By induction, it is shown that two sequences subject to restriction (3.1.2) uniquely determine a $\Delta_1(k,r)$-graph. Therefore, counting the number of $\Delta_1(k,r)$-graphs is equivalent to counting the number of pairs of characteristic sequences.

Now, we count the number of characteristic sequences for given k and r. We have the following lemma.

Lemma 3.5. *For a given k and r $(0 \leq r \leq k-1)$, the number of $\Delta_1(k,r)$-graphs is*

$$\frac{1}{r+1}\binom{k}{r}\binom{k-1}{r}.$$

Proof. Ignoring the restriction (3.1.2), we have $\binom{k-1}{r}\binom{k-1}{r}$ ways to arrange r ones in the $k-1$ positions $u_1, \cdots, u_{k-1}$ and to arrange r minus ones in the $k-1$ positions $d_2, \cdots, d_k$. If there is an integer $2 \leq \ell \leq k$ such that

$$u_1 + \cdots + u_{\ell-1} + d_1 + \cdots + d_\ell = -1,$$

then define

$$\tilde{u}_j = \begin{cases} u_j, & \text{if } j < \ell, \\ -d_{j+1}, & \text{if } \ell \leq j < k, \end{cases}$$

and

$$\tilde{d}_j = \begin{cases} d_j, & \text{if } 1 < j \le \ell, \\ -u_{j-1}, & \text{if } \ell < j \le k. \end{cases}$$

Then we have $r-1$ u's equal to one and $r+1$ d's equal to minus one. There are $\binom{k-1}{r-1}\binom{k-1}{r+1}$ ways to arrange $r-1$ ones in the $k-1$ positions $\tilde{u}_1, \cdots, \tilde{u}_{k-1}$, and to arrange $r+1$ minus ones in the $k-1$ positions $\tilde{d}_2, \cdots, \tilde{d}_k$.

Therefore, the number of pairs of characteristic sequences with indices k and r satisfying the restriction (3.1.2) is

$$\binom{k-1}{r}^2 - \binom{k-1}{r-1}\binom{k-1}{r+1} = \frac{1}{r+1}\binom{k}{r}\binom{k-1}{r}.$$

The proof of the lemma is complete.

3.1.3 M-P Law for the iid Case

In this section, we consider the LSD of the sample covariance matrix for the case where the underlying variables are iid.

Theorem 3.6. *Suppose that $\{x_{ij}\}$ are iid real random variables with mean zero and variance σ^2. Also assume that $p/n \to y \in (0, \infty)$. Then, with probability one, $F^{\mathbf{S}}$ tends to the M-P law, which is defined in (3.1.1).*

Yin [300] considered existence of the LSD of the sequence of random matrices $\mathbf{S}_n\mathbf{T}_n$, where $\mathbf{T}_n$ is a positive definite random matrix and is independent of $\mathbf{S}_n$. When $\mathbf{T}_n = \mathbf{I}_p$, Yin's result reduces to Theorem 3.6.

In this section, we shall give a proof of the following extension to the complex random sample covariance matrix.

Theorem 3.7. *Suppose that $\{x_{ij}\}$ are iid complex random variables with variance σ^2. Also assume that $p/n \to y \in (0, \infty)$. Then, with probability one, $F^{\mathbf{S}}$ tends to a limiting distribution the same as described in Theorem 3.6.*

Remark 3.8. The proofs will be separated into several steps. Note that the M-P law varies with the scale parameter σ^2. Therefore, in the proof we shall assume that $\sigma^2 = 1$, without loss of generality.

In most work in multivariate statistics, it is assumed that the means of the entries of $\mathbf{X}_n$ are zero. The centralization technique, which is Theorem A.44, relies on the interlacing property of eigenvalues of two matrices that differ by a rank-one matrix. One then sees that removing the common mean of the entries of $\mathbf{X}_n$ does not alter the LSD of sample covariance matrices.

Step 1. Truncation, Centralization, and Rescaling
Let C be a positive number, and define

$$\begin{aligned}
\hat{x}_{ij} &= x_{ij} I(|x_{ij}| \le C), \\
\tilde{x}_{ij} &= \hat{x}_{ij} - \mathrm{E}(\hat{x}_{11}), \\
\widehat{\mathbf{x}}_i &= (\hat{x}_{i1}, \cdots, \hat{x}_{ip})', \\
\widetilde{\mathbf{x}}_i &= (\tilde{x}_{i1}, \cdots, \tilde{x}_{ip})', \\
\widehat{\mathbf{S}}_n &= \frac{1}{n} \sum_{i=1}^{n} \widehat{\mathbf{x}}_i \widehat{\mathbf{x}}_i^* = \frac{1}{n} \widehat{\mathbf{X}} \widehat{\mathbf{X}}^*, \\
\widetilde{\mathbf{S}}_n &= \frac{1}{n} \sum_{i=1}^{n} \widetilde{\mathbf{x}}_i \widetilde{\mathbf{x}}_i^* = \frac{1}{n} \widetilde{\mathbf{X}} \widetilde{\mathbf{X}}^*.
\end{aligned}$$

Write the ESDs of $\widehat{\mathbf{S}}_n$ and $\widetilde{\mathbf{S}}_n$ as $F^{\widehat{\mathbf{S}}_n}$ and $F^{\widetilde{\mathbf{S}}_n}$, respectively. By Corollary A.42 and the strong law of large numbers, we have

$$\begin{aligned}
L^4(F^{\mathbf{S}}, F^{\widehat{\mathbf{S}}_n}) &\le \left(\frac{2}{np} \sum_{i,j} (|x_{ij}^2| + |\hat{x}_{ij}^2|) \right) \left(\frac{1}{np} \sum_{i,j} (|x_{ij} - \hat{x}_{ij}|^2) \right) \\
&\le \left(\frac{4}{np} \sum_{i,j} |x_{ij}^2| \right) \left(\frac{1}{np} \sum_{i,j} (|x_{ij}^2| I(|x_{ij}| > C)) \right) \\
&\to 4\mathrm{E}(|x_{ij}^2| I(|x_{ij}| > C)), \text{ a.s.} \qquad (3.1.3)
\end{aligned}$$

Note that the right-hand side of (3.1.3) can be made arbitrarily small by choosing C large enough.

Also, by Theorem A.44, we obtain

$$||F^{\widehat{\mathbf{S}}_n} - F^{\widetilde{\mathbf{S}}_n}|| \le \frac{1}{p} \mathrm{rank}(\mathrm{E}\widehat{\mathbf{X}}) = \frac{1}{p}. \qquad (3.1.4)$$

Write $\tilde{\sigma}^2 = \mathrm{E}(|\tilde{x}_{jk}|^2) \to 1$, as $C \to \infty$. Applying Corollary A.42, we obtain

$$\begin{aligned}
L^4(F^{\widetilde{\mathbf{S}}_n}, F^{\tilde{\sigma}^{-2}\widetilde{\mathbf{S}}_n}) &\le 2 \left(\frac{1 + \tilde{\sigma}^2}{np\tilde{\sigma}^2} \sum_{i,j} |\tilde{x}_{ij}|^2 \right) \left(\frac{1 - \tilde{\sigma}^2}{np\tilde{\sigma}^2} \sum_{i,j} |\tilde{x}_{ij(c)}|^2 \right) \\
&\to 2(1 - \tilde{\sigma}^4), \text{ a.s.} \qquad (3.1.5)
\end{aligned}$$

Note that the right-hand side of the inequality above can be made arbitrarily small by choosing C large. Combining (3.1.3), (3.1.4), and (3.1.5), in the proof of Theorem 3.7 we may assume that the variables x_{jk} are uniformly bounded with mean zero and variance 1. For abbreviation, in proofs given in the next step, we still use $\mathbf{S}_n$, $\mathbf{X}_n$ for the matrices associated with the truncated variables.

Step 2. Proof for the M-P Law by MCT

Now, we are able to employ the moment approach to prove Theorem 3.7. By elementary calculus, we have

$$\begin{aligned}\beta_k(\mathbf{S}_n) &= \int x^k F^{\mathbf{S}_n}(dx) \\ &= p^{-1}n^{-k} \sum_{\{i_1,\cdots,i_k\}} \sum_{\{j_1,\cdots,j_k\}} x_{i_1j_1}\bar{x}_{i_2j_1}x_{i_2j_2}\cdots x_{i_kj_k}\bar{x}_{i_1j_k} \\ &:= p^{-1}n^{-k}\sum_{\mathbf{i},\mathbf{j}} X_{G(\mathbf{i},\mathbf{j})},\end{aligned}$$

where the summation runs over all $G(\mathbf{i},\mathbf{j})$-graphs as defined in Subsection 3.1.2, the indices in $\mathbf{i} = (i_1,\cdots,i_k)$ run over $1,2,\cdots,p$, and the indices in $\mathbf{j} = (j_1,\cdots,j_k)$ run over $1,2,\cdots,n$.

To complete the proof of the almost sure convergence of the ESD of $\mathbf{S}_n$, we need only show the following two assertions:

$$\begin{aligned}\mathrm{E}(\beta_k(\mathbf{S}_n)) &= p^{-1}n^{-k}\sum_{\mathbf{i},\mathbf{j}} \mathrm{E}(x_{G(\mathbf{i},\mathbf{j})}) \\ &= \sum_{r=0}^{k-1} \frac{y_n^r}{r+1}\binom{k}{r}\binom{k-1}{r} + O(n^{-1}) \qquad (3.1.6)\end{aligned}$$

and

$$\begin{aligned}&\mathrm{Var}(\beta_k(\mathbf{S}_n)) \\ &= p^{-2}n^{-2k}\sum_{\mathbf{i}_1,\mathbf{j}_1,\mathbf{i}_2,\mathbf{j}_2} [\mathrm{E}(x_{G_1(\mathbf{i}_1,\mathbf{j}_1)}x_{G_2(\mathbf{i}_2,\mathbf{j}_2)} - \mathrm{E}(x_{G_1(\mathbf{i}_1,\mathbf{j}_1)})\mathrm{E}(x_{G_2(\mathbf{i}_2,\mathbf{j}_2)}))] \\ &= O(n^{-2}), \qquad (3.1.7)\end{aligned}$$

where $y_n = p/n$, and the graphs G_1 and G_2 are defined by $(\mathbf{i}_1,\mathbf{j}_1)$ and $(\mathbf{i}_2,\mathbf{j}_2)$, respectively.

The proof of (3.1.6). On the left-hand side of (3.1.6), two terms are equal if their corresponding graphs are isomorphic. Therefore, by Lemma 3.2, we may rewrite

$$\mathrm{E}(\beta_k(\mathbf{S}_n)) = p^{-1}n^{-k}\sum_{\Delta(k,r,s)} p(p-1)\cdots(p-r)n(n-1)\cdots(n-s+1)\mathrm{E}(X_{\Delta(k,r,s)}), \qquad (3.1.8)$$

where the summation is taken over canonical $\Delta(k,r,s)$-graphs. Now, split the sum in (3.1.8) into three parts according to $\Delta_1(k,r)$ and $\Delta_j(k,r,s)$, $j = 2,3$. Since the graph in $\Delta_2(k,r,s)$ contains at least one single edge, the corresponding expectation is zero. That is,

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

By Lemma 3.3, for a graph of $\Delta_3(k, r, s)$, we have $r + s < k$. Since the variable $x_{\Delta(k,r,s)}$ is bounded by $(2C/\tilde{\sigma})^{2k}$, we conclude that

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

Now let us evaluate S_1. For a graph in $\Delta_1(k, r)$ (with $s = k - r$), each pair of coincident edges consists of a down edge and an up edge; say, the edge (i_a, j_a) must coincide with the edge (j_a, i_a). This pair of coincident edges corresponds to the expectation $\mathrm{E}(|X_{i_a,j_a}|^2) = 1$. Therefore, $\mathrm{E}(X_{\Delta_1(k,r)}) = 1$. By Lemma 3.4,

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

where $y_n = p/n \to y \in (0, \infty)$. The proof of (3.1.6) is complete.

The proof of (3.1.7). Recall

$$\begin{aligned} &\mathrm{Var}(\beta_k(\mathbf{S}_n)) \\ &= p^{-2}n^{-2k} \sum_{\mathbf{i},\mathbf{j}} [\mathrm{E}(X_{G_1(\mathbf{i}_1,\mathbf{j}_1)} X_{G_2(\mathbf{i}_2,\mathbf{j}_2)}) - \mathrm{E}(X_{G_1(\mathbf{i}_1,\mathbf{j}_1)})\mathrm{E}(X_{G_2(\mathbf{i}_2,\mathbf{j}_2)})]. \end{aligned}$$

Similar to the proof of Theorem 2.5, if G_1 has no edges coincident with edges of G_2 or $G = G_1 \cup G_2$ has an overall single edge, then

$$\mathrm{E}(X_{G_1(\mathbf{i}_1,\mathbf{j}_1)} X_{G_2(\mathbf{i}_2,\mathbf{j}_2)}) - \mathrm{E}(X_{G_1(\mathbf{i}_1,\mathbf{j}_1)})\mathrm{E}(X_{G_2(\mathbf{i}_2,\mathbf{j}_2)}) = 0$$

by independence between X_{G_1} and X_{G_2}.

Similar to the arguments in Subsection 2.1.3, one may show that the number of noncoincident vertices of G is not more than $2k$. By the fact that the terms are bounded, we conclude that assertion (3.1.7) holds and consequently conclude the proof of Theorem 3.7.

Remark 3.9. The existence of the second moment of the entries is obviously necessary and sufficient for the Marčenko-Pastur law since the limiting distribution involves the parameter σ^2.

3.2 Generalization to the Non-iid Case

Sometimes it is of practical interest to consider the case where the entries of $\mathbf{X}_n$ depend on n and for each n they are independent but not necessarily identically distributed. As in Section 2.2, we shall briefly present a proof of the following theorem.

Theorem 3.10. *Suppose that, for each n, the entries of $\mathbf{X}$ are independent complex variables with a common mean μ and variance σ^2. Assume that $p/n \to y \in (0, \infty)$ and that, for any $\eta > 0$,*

$$\frac{1}{\eta^2 np} \sum_{jk} \mathrm{E}(|x_{jk}^{(n)}|^2 I(|x_{jk}^{(n)}| \geq \eta\sqrt{n})) \to 0. \tag{3.2.1}$$

Then, with probability one, $F^{\mathbf{S}}$ tends to the Marčenko-Pastur law with ratio index y and scale index σ^2.

Proof. We shall only give an outline of the proof of this theorem. The details are left to the reader. Without loss of generality, we assume that $\mu = 0$ and $\sigma^2 = 1$. Similar to what we did in the proof of Theorem 2.9, we may select a sequence $\eta_n \downarrow 0$ such that condition (3.2.1) holds true when η is replaced by η_n. In the following, once condition (3.2.1) is used, we always mean this condition with η replaced by η_n.

Applying Theorem A.44 and the Bernstein inequality, by condition (3.2.1), we may truncate the variables $x_{ij}^{(n)}$ at $\eta_n\sqrt{n}$. Then, applying Corollary A.42, by condition (3.2.1), we may recentralize and rescale the truncated variables. Thus, in the rest of the proof, we shall drop the superscript (n) from the variables for brevity. We further assume that

$$\begin{aligned} &1)\ |x_{ij}| < \eta_n\sqrt{n}, \\ &2)\ \mathrm{E}(x_{ij}) = 0 \ \text{ and } \ \mathrm{Var}(x_{ij}) = 1. \end{aligned} \tag{3.2.2}$$

By arguments to those in the proof of Theorem 2.9, one can show the following two assertions:

$$\mathrm{E}(\beta_k(\mathbf{S}_n)) = \sum_{r=0}^{k-1} \frac{y_n^r}{r+1} \binom{k}{r} \binom{k-1}{r} + o(1) \tag{3.2.3}$$

and

$$\mathrm{E}\left|\beta_k(\mathbf{S}_n) - \mathrm{E}\left(\beta_k(\mathbf{S}_n)\right)\right|^4 = o(n^{-2}). \tag{3.2.4}$$

The proof of Theorem 3.10 is then complete.

3.3 Proof of Theorem 3.10 by the Stieltjes Transform

As an illustration applying Stieltjes transforms to sample covariance matrices, we give a proof of Theorem 3.10 in this section. Using the same approach of truncating, centralizing, and rescaling as we did in the last section, we may assume the additional conditions given in (3.2.2).

3.3.1 Stieltjes Transform of the M-P Law

Let $z = u+iv$ with $v > 0$ and $s(z)$ be the Stieltjes transform of the M-P law.

Lemma 3.11.

$$s(z) = \frac{\sigma^2(1-y) - z + \sqrt{(z-\sigma^2-y\sigma^2)^2 - 4y\sigma^4}}{2yz\sigma^2}. \tag{3.3.1}$$

Proof. When $y < 1$, we have

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

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

Letting $x = \sigma^2(1+y+2\sqrt{y}\cos w)$ and then setting $\zeta = e^{iw}$, we have

$$\begin{aligned}
s(z) &= \int_0^\pi \frac{2}{\pi}\frac{1}{(1+y+2\sqrt{y}\cos w)(\sigma^2(1+y+2\sqrt{y}\cos w)-z)}\sin^2 w dw \\
&= \frac{1}{\pi}\int_0^{2\pi} \frac{((e^{iw}-e^{-iw})/2i)^2}{(1+y+\sqrt{y}(e^{iw}+e^{-iw}))(\sigma^2(1+y+\sqrt{y}(e^{iw}+e^{-iw}))-z)}dw \\
&= -\frac{1}{4i\pi}\oint_{|\zeta|=1} \frac{(\zeta-\zeta^{-1})^2}{\zeta(1+y+\sqrt{y}(\zeta+\zeta^{-1}))(\sigma^2(1+y+\sqrt{y}(\zeta+\zeta^{-1}))-z)}d\zeta \\
&= -\frac{1}{4i\pi}\oint_{|\zeta|=1} \frac{(\zeta^2-1)^2}{\zeta((1+y)\zeta+\sqrt{y}(\zeta^2+1))(\sigma^2(1+y)\zeta+\sqrt{y}\sigma^2(\zeta^2+1)-z\zeta)}d\zeta.
\end{aligned} \tag{3.3.2}$$

The integrand function has five simple poles at

$$\begin{aligned}
\zeta_0 &= 0, \\
\zeta_1 &= \frac{-(1+y)+(1-y)}{2\sqrt{y}}, \\
\zeta_2 &= \frac{-(1+y)-(1-y)}{2\sqrt{y}},
\end{aligned}$$

$$\zeta_3 = \frac{-\sigma^2(1+y) + z + \sqrt{\sigma^4(1-y)^2 - 2\sigma^2(1+y)z + z^2}}{2\sigma^2\sqrt{y}},$$
$$\zeta_4 = \frac{-\sigma^2(1+y) + z - \sqrt{\sigma^4(1-y)^2 - 2\sigma^2(1+y)z + z^2}}{2\sigma^2\sqrt{y}}.$$

By elementary calculation, we find that the residues at these five poles are

$$\frac{1}{y\sigma^2}, \quad \mp\frac{1-y}{yz} \quad \text{and} \quad \pm\frac{1}{\sigma^2 yz}\sqrt{\sigma^4(1-y)^2 - 2\sigma^2(1+y)z + z^2}.$$

Noting that $\zeta_3\zeta_4 = 1$ and recalling the definition for the square root of complex numbers, we know that both the real part and imaginary part of $\sqrt{\sigma^2(1-y)^2 - 2\sigma^2(1+y)z + z^2}$ and $-\sigma^2(1+y)+z$ have the same signs and hence $|\zeta_3| > 1$, $|\zeta_4| < 1$. Also, $|\zeta_1| = |-\sqrt{y}| < 1$ and $|\zeta_2| = |-1/\sqrt{y}| > 1$. By Cauchy integration, we obtain

$$\begin{aligned} s(z) &= -\frac{1}{2}\left(\frac{1}{y\sigma^2} - \frac{1}{\sigma^2 yz}\sqrt{\sigma^4(1-y)^2 - 2\sigma^2(1+y)z + z^2} - \frac{1-y}{yz}\right) \\ &= \frac{\sigma^2(1-y) - z + \sqrt{(z-\sigma^2-y\sigma^2)^2 - 4y\sigma^4}}{2yz\sigma^2}. \end{aligned}$$

This proves equation (3.3.1) when $y < 1$.

When $y > 1$, since the M-P law has also a point mass $1 - 1/y$ at zero, $s(z)$ equals the integral above plus $-(y-1)/yz$. In this case, $|\zeta_3| = |-\sqrt{y}| > 1$ and $|\zeta_4| = |-1/\sqrt{y}| < 1$, and thus the residue at ζ_4 should be counted into the integral. Finally, one finds that equation (3.3.1) still holds. When $y = 1$, the equation is still true by continuity in y.

3.3.2 Proof of Theorem 3.10

Let the Stieltjes transform of the ESD of $\mathbf{S}_n$ be denoted by $s_n(z)$. Define

$$s_n(z) = \frac{1}{p}\mathrm{tr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1}.$$

As in Section 2.3, we shall complete the proof by the following three steps:

(i) For any fixed $z \in \mathbb{C}^+$, $s_n(z) - \mathrm{E}s_n(z) \to 0$, a.s.
(ii) For any fixed $z \in \mathbb{C}^+$, $\mathrm{E}s_n(z) \to s(z)$, the Stieltjes transform of the M-P law.
(iii) Except for a null set, $s_n(z) \to s(z)$ for every $z \in \mathbb{C}^+$.

Similar to Section 2.3, the last step is implied by the first two steps and thus its proof is omitted. We now proceed with the first two steps.

Step 1. Almost sure convergence of the random part

$$s_n(z) - \mathrm{E}s_n(z) \to 0, \quad \text{a.s.} \tag{3.3.3}$$

Let $\mathrm{E}_k(\cdot)$ denote the conditional expectation given $\{\mathbf{x}_{k+1}, \cdots, \mathbf{x}_n\}$. Then, by the formula

$$(\mathbf{A} + \boldsymbol{\alpha}\boldsymbol{\beta}^*)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\boldsymbol{\alpha}\boldsymbol{\beta}^*\mathbf{A}^{-1}}{1 + \boldsymbol{\beta}^*\mathbf{A}^{-1}\boldsymbol{\alpha}} \tag{3.3.4}$$

we obtain

$$\begin{aligned} s_n(z) - \mathrm{E}s_n(z) &= \frac{1}{p}\sum_{k=1}^{n}[\mathrm{E}_k \mathrm{tr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1} - \mathrm{E}_{k-1}\mathrm{tr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1}] \\ &= \frac{1}{p}\sum_{k=1}^{n}\gamma_k, \end{aligned}$$

where, by Theorem A.5,

$$\begin{aligned} \gamma_k &= (\mathrm{E}_k - \mathrm{E}_{k-1})[\mathrm{tr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1} - \mathrm{tr}(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-1}] \\ &= -[\mathrm{E}_k - \mathrm{E}_{k-1}]\frac{\mathbf{x}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-2}\mathbf{x}_k}{1 + \mathbf{x}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-1}\mathbf{x}_k} \end{aligned}$$

and $\mathbf{S}_{nk} = \mathbf{S}_n - \mathbf{x}_k\mathbf{x}_k^*$. Note that

$$\begin{aligned} &\left|\frac{\mathbf{x}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-2}\mathbf{x}_k}{1 + \mathbf{x}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-1}\mathbf{x}_k}\right| \\ &\le \frac{\mathbf{x}_k^*((\mathbf{S}_{nk} - u\mathbf{I}_p)^2 + v^2\mathbf{I}_p)^{-1}\mathbf{x}_k}{\Im(1 + \mathbf{x}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_p)^{-1}\mathbf{x}_k)} = \frac{1}{v}. \end{aligned}$$

Noticing that $\{\gamma_k\}$ forms a sequence of bounded martingale differences, by Lemma 2.12 with $p = 4$, we obtain

$$\begin{aligned} \mathrm{E}|s_n(z) - \mathrm{E}s_n(z)|^4 &\le \frac{K_4}{p^4}\mathrm{E}\left(\sum_{k=1}^{n}|\gamma_k|^2\right)^2 \\ &\le \frac{4K_4 n^2}{v^4 p^4} = O(n^{-2}), \end{aligned}$$

which, together with the Borel-Cantelli lemma, implies (3.3.3). The proof is complete.

Step 2. Mean convergence

We will show that

$$\mathrm{E}s_n(z) \to s(z), \tag{3.3.5}$$

where $s(z)$ is defined in (3.3.1) with $\sigma^2 = 1$.

By Theorem A.4, we have

$$s_n(z) = \frac{1}{p}\sum_{k=1}^{p} \frac{1}{\frac{1}{n}\boldsymbol{\alpha}_k'\overline{\boldsymbol{\alpha}}_k - z - \frac{1}{n^2}\boldsymbol{\alpha}_k'\mathbf{X}_k^*(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1})^{-1}\mathbf{X}_k\overline{\boldsymbol{\alpha}}_k}, \quad (3.3.6)$$

where $\mathbf{X}_k$ is the matrix obtained from $\mathbf{X}$ with the k-th row removed and $\boldsymbol{\alpha}_k'$ $(n \times 1)$ is the k-th row of $\mathbf{X}$.

Set

$$\varepsilon_k = \frac{1}{n}\boldsymbol{\alpha}_k'\overline{\boldsymbol{\alpha}}_k - 1 - \frac{1}{n^2}\boldsymbol{\alpha}_k'\mathbf{X}_k^*\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1}\mathbf{X}_k\overline{\boldsymbol{\alpha}}_k + y_n + y_n z\mathrm{E}s_n(z), \quad (3.3.7)$$

where $y_n = p/n$. Then, by (3.3.6), we have

$$\mathrm{E}s_n(z) = \frac{1}{1 - z - y_n - y_n z\mathrm{E}s_n(z)} + \delta_n, \quad (3.3.8)$$

where

$$\delta_n = -\frac{1}{p}\sum_{k=1}^{p}\mathrm{E}\left(\frac{\varepsilon_k}{(1 - z - y_n - y_n z\mathrm{E}s_n(z))(1 - z - y_n - y_n z\mathrm{E}s_n(z) + \varepsilon_k)}\right). \quad (3.3.9)$$

Solving $\mathrm{E}s_n(z)$ from equation (3.3.8), we get two solutions:

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

Comparing this with (3.3.1), it suffices to show that

$$\mathrm{E}s_n(z) = s_1(z) \quad (3.3.10)$$

and

$$\delta_n \to 0. \quad (3.3.11)$$

We show (3.3.10) first. Making $v \to \infty$, we know that $\mathrm{E}s_n(z) \to 0$ and hence $\delta_n \to 0$ by (3.3.8). This shows that $\mathrm{E}s_n(z) = s_1(z)$ for all z with large imaginary part. If (3.3.10) is not true for all $z \in \mathbb{C}^+$, then by the continuity of s_1 and s_2, there exists a $z_0 \in \mathbb{C}^+$ such that $s_1(z_0) = s_2(z_0)$, which implies that

$$(1 - z_0 - y_n + y_n z\delta_n)^2 - 4y_n z_0(1 + \delta_n(1 - z_0 - y_n)) = 0.$$

Thus,

$$\mathrm{E}s_n(z_0) = s_1(z_0) = \frac{1 - z_0 - y_n + y_n z_0 \delta_n}{2y_n z_0}.$$

Substituting the solution δ_n of equation (3.3.8) into the identity above, we obtain

$$\mathrm{E}s_n(z_0) = \frac{1 - z_0 - y_n}{y_n z_0} + \frac{1}{y_n + z_0 - 1 + y_n z_0 \mathrm{E}s_n(z_0)}. \tag{3.3.12}$$

Noting that for any Stieltjes transform $s(z)$ of probability F defined on $\mathbb{R}^+$ and positive y, we have

$$\begin{aligned}\Im(y + z - 1 + yzs(z)) &= \Im\left(z - 1 + \int_0^\infty \frac{yxdF(x)}{x - z}\right) \\ &= v\left(1 + \int_0^\infty \frac{yxdF(x)}{(x-u)^2 + v^2}\right) > 0. \end{aligned} \tag{3.3.13}$$

In view of this, it follows that the imaginary part of the second term in (3.3.12) is negative. If $y_n \leq 1$, it can be easily seen that $\Im(1 - z_0 - y_n)/(y_n z_0) < 0$. Then we conclude that $\Im \mathrm{E}s_n(z_0) < 0$, which is impossible since the imaginary part of the Stieltjes transform should be positive. This contradiction leads to the truth of (3.3.10) for the case $y_n \leq 1$.

For the general case, we can prove it in the following way. In view of (3.3.12) and (3.3.13), we should have

$$y_n + z_0 - 1 + y_n z_0 \mathrm{E}s_n(z_0) = \sqrt{y_n z_0}. \tag{3.3.14}$$

Now, let $\underline{s}_n(z)$ be the Stieltjes transform of the matrix $\frac{1}{n}\mathbf{X}^*\mathbf{X}$. Noting that $\frac{1}{n}\mathbf{X}^*\mathbf{X}$ and $\mathbf{S}_n = \frac{1}{n}\mathbf{X}\mathbf{X}^*$ have the same set of nonzero eigenvalues, we have the relation between s_n and $\underline{s}_n$ given by

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

Note that the equation above is true regardless of whether $y_n > 1$ or $y_n \leq 1$. From this we have

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

Substituting this into (3.3.14), we obtain

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

which leads to a contradiction that the imaginary part of LHS is positive and that of the RHS is negative. Then, (3.3.10) is proved.

Now, let us consider the proof of (3.3.11). Rewrite

$$\delta_n = -\frac{1}{p}\sum_{k=1}^{p}\left(\frac{\mathrm{E}\varepsilon_k}{(1-z-y_n-y_nz\mathrm{E}s_n(z))^2}\right)$$
$$+\frac{1}{p}\sum_{k=1}^{p}\mathrm{E}\left(\frac{\varepsilon_k^2}{(1-z-y_n-y_nz\mathrm{E}s_n(z))^2(1-z-y_n-y_nz\mathrm{E}s_n(z)+\varepsilon_k)}\right)$$
$$= J_1 + J_2.$$

At first, by assumptions given in (3.2.2), we note that

$$\begin{aligned}|\mathrm{E}\varepsilon_k| &= \left|-\frac{1}{n^2}\mathrm{Etr}\mathbf{X}_k^*\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1}\mathbf{X}_k + y_n + y_nz\mathrm{E}s_n(z)\right| \\ &= \left|-\frac{1}{n}\mathrm{Etr}\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1}\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* + y_n + y_nz\mathrm{E}s_n(z)\right| \\ &\le \frac{1}{n} + \frac{|z|y_n}{n}\mathrm{E}\left|\mathrm{tr}\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1} - s_n(z)\right| \\ &\le \frac{1}{n} + \frac{|z|y_n}{nv} \to 0, \end{aligned} \tag{3.3.15}$$

which implies that $J_1 \to 0$.

Now we prove $J_2 \to 0$. Since

$$\begin{aligned} &\Im\left(1-z-y_n-y_nz\mathrm{E}s_n(z)+\varepsilon_k\right) \\ &= \Im\left(\frac{1}{n}\boldsymbol{\alpha}_k'\overline{\boldsymbol{\alpha}}_k - z - \frac{1}{n^2}\boldsymbol{\alpha}_k'\mathbf{X}_k^*\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1}\mathbf{X}_k\overline{\boldsymbol{\alpha}}_k\right) \\ &= -v\left(1 + \frac{1}{n^2}\boldsymbol{\alpha}_k'\mathbf{X}_k^*\left[\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - u\mathbf{I}_{p-1}\right)^2 + v^2\mathbf{I}_{p-1}\right]^{-1}\mathbf{X}_k\overline{\boldsymbol{\alpha}}_k\right) < -v, \end{aligned}$$

combining this with (3.3.13), we obtain

$$\begin{aligned} |J_2| &\le \frac{1}{pv^3}\sum_{k=1}^{p}\mathrm{E}|\varepsilon_k|^2 \\ &= \frac{1}{pv^3}\sum_{k=1}^{p}[\mathrm{E}|\varepsilon_k - \widetilde{\mathrm{E}}(\varepsilon_k)|^2 + \mathrm{E}|\widetilde{\mathrm{E}}\varepsilon_k - \mathrm{E}(\varepsilon_k)|^2 + (\mathrm{E}(\varepsilon_k))^2], \end{aligned}$$

where $\widetilde{\mathrm{E}}(\cdot)$ denotes the conditional expectation given $\{\boldsymbol{\alpha}_j, j = 1, ..., k-1, k+1, ..., p\}$. In the estimation of J_1, we have proved that

$$|\mathrm{E}(\varepsilon_k)| \le \frac{1}{n} + \frac{|z|y}{nv} \to 0.$$

Write $\mathbf{A} = (a_{ij}) = \mathbf{I}_n - \frac{1}{n}\mathbf{X}_k^*(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1})^{-1}\mathbf{X}_k$. Then, we have

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

By elementary calculation, we have

$$\begin{aligned}
&\frac{1}{n^2}\tilde{\mathrm{E}}|\varepsilon_k' - \tilde{\mathrm{E}}\varepsilon_k|^2 \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n} |a_{ii}|^2(\mathrm{E}|x_{ki}|^4 - 1) + \sum_{i \neq j}[|a_{ij}|^2 \mathrm{E}|x_{ki}|^2 \mathrm{E}|x_{kj}|^2 + a_{ij}^2 \mathrm{E}x_{ki}^2 \mathrm{E}x_{kj}^2]\right) \\
&\leq \frac{1}{n^2}\left(\sum_{i=1}^{n} |a_{ii}|^2(\eta_n^2 n) + 2\sum_{i \neq j} |a_{ij}|^2\right) \\
&\leq \frac{\eta_n^2}{v^2} + \frac{2}{nv^2}.
\end{aligned}$$

Here, we have used the fact that $|a_{ii}| \leq v^{-1}$.

Using the martingale decomposition method in the proof of (3.3.3), we can show that

$$\begin{aligned}
&\mathrm{E}|\widetilde{\mathrm{E}}\varepsilon_k - \mathrm{E}\varepsilon_k|^2 \\
&= \frac{|z|^2 y^2}{n^2}\mathrm{E}\left|\mathrm{tr}\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1} - \mathrm{E}\mathrm{tr}\left(\frac{1}{n}\mathbf{X}_k\mathbf{X}_k^* - z\mathbf{I}_{p-1}\right)^{-1}\right|^2 \\
&\leq \frac{|z|^2 y^2}{nv^2} \to 0.
\end{aligned}$$

Combining the three estimations above , we have completed the proof of the mean convergence of the Stieltjes transform of the ESD of $\mathbf{S}_n$.

Consequently, Theorem 3.10 is proved by the method of Stieltjes transforms.

Chapter 4
Product of Two Random Matrices

In this chapter, we shall consider the LSD of a product of two random matrices, one of them a sample covariance matrix and the other an arbitrary Hermitian matrix. This topic is related to two areas: The first is the study of the LSD of a multivariate F-matrix that is a product of a sample covariance matrix and the inverse of another sample covariance matrix, independent of each other. Multivariate F plays an important role in multivariate data analysis, such as two-sample tests, MANOVA (multivariate analysis of variance), and multivariate linear regression. The second is the investigation of the LSD of a sample covariance matrix when the population covariance matrix is arbitrary. The sample covariance matrix under a general setup is, as mentioned in Chapter 3, fundamental in multivariate analysis.

Pioneering work was done by Wachter [290], who considered the limiting distribution of the solutions to the equation

$$\det(\mathbf{X}_{1,n_1}\mathbf{X}'_{1,n_1} - \lambda\mathbf{X}_{2,n_2}\mathbf{X}'_{2,n_2}) = 0, \tag{4.0.1}$$

where $\mathbf{X}_{j,n_j}$ is a $p \times n_j$ matrix whose entries are iid $N(0,1)$ and $\mathbf{X}_{1,n_1}$ is independent of $\mathbf{X}_{2,n_2}$. When $\mathbf{X}_{2,n_2}\mathbf{X}'_{2,n_2}$ is of full rank, the solutions to (4.0.1) are n_2/n_1 times the eigenvalues of the multivariate F-matrix $(\frac{1}{n_1}\mathbf{X}_{1,n_1}\mathbf{X}'_{1,n_1})(\frac{1}{n_2}\mathbf{X}_{2,n_2}\mathbf{X}'_{2,n_2})^{-1}$.

Yin and Krishnaiah [304] established the existence of the LSD of the matrix sequence $\{\mathbf{S}_n\mathbf{T}_n\}$, where $\mathbf{S}_n$ is a standard Wishart matrix of dimension p and degrees of freedom n with $p/n \to y \in (0,\infty)$, $\mathbf{T}_n$ is a positive definite matrix satisfying $\beta_k(\mathbf{T}_n) \to H_k$, and the sequence H_k satisfies the Carleman condition (see (B.1.4)). In Yin [300], this result was generalized to the case where the sample covariance matrix is formed based on iid real random variables of mean zero and variance one. Using the result of Yin and Krishnaiah [304], Yin, Bai, and Krishnaiah [302] showed the existence of the LSD of the multivariate F-matrix. The explicit form of the LSD of multivariate F-matrices was derived in Bai, Yin, and Krishnaiah [40] and Silverstein

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_4,

[256]. Under the same structure, Bai, Yin, and Krishnaiah [41] established the existence of the LSD when the underlying distribution of $\mathbf{S}_n$ is isotropic.

Some further extensions were done in Silverstein [256] and Silverstein and Bai [266]. In this chapter, we shall introduce some recent developments in this direction.

Bai and Yin [39] considered the upper limit of the spectral moments of a power of $\mathbf{X}_n$ (i.e., the limits of $\beta_\ell((\frac{1}{\sqrt{n}}\mathbf{X}_n)^k(\frac{1}{\sqrt{n}}\mathbf{X}'_n)^k)$, where $\mathbf{X}_n$ is of order $n \times n$) when investigating the limiting behavior of solutions to a large system of linear equations. Based on this result, it is proved that the upper limit of the spectral radius of $\frac{1}{\sqrt{n}}\mathbf{X}_n$ is not larger than 1. The same result was obtained in Geman [117] at almost the same time but by different approaches and assuming stronger conditions.

4.1 Main Results

Here we present the following results.

Theorem 4.1. *Suppose that the entries of* $\mathbf{X}_n$ $(p \times n)$ *are independent complex random variables satisfying (3.2.1), that* $\mathbf{T}_n$ *is a sequence of Hermitian matrices independent of* $\mathbf{X}_n$*, and that the ESD of* $\mathbf{T}_n$ *tends to a nonrandom limit* F^T *in some sense (in probability or a.s.). If* $p/n \to y \in (0, \infty)$*, then the ESD of the product* $\mathbf{S}_n\mathbf{T}_n$ *tends to a nonrandom limit in probability or almost surely (accordingly), where* $\mathbf{S}_n = \frac{1}{n}\mathbf{X}_n\mathbf{X}_n^*$.

Remark 4.2. Note that the eigenvalues of the product matrix $\mathbf{S}_n\mathbf{T}_n$ are all real, although it is not symmetric, because the whole set of eigenvalues is the same as that of the symmetric matrix $\mathbf{S}_n^{1/2}\mathbf{T}_n\mathbf{S}_n^{1/2}$.

This theorem contains Yin's result as a special case. In Yin [300], the entries of $\mathbf{X}$ are assumed to be real and iid with mean zero and variance one and the matrix $\mathbf{T}_n$ real and positive definite and satisfying, for each fixed k,

$$\frac{1}{p}\mathrm{tr}(\mathbf{T}_n^k) \to H_k \quad \text{(in probability or a.s.)} \tag{4.1.1}$$

while the constant sequence $\{H_k\}$ satisfies the Carleman condition.

In Silverstein [256], Theorem 4.1 was established under the additional condition that $\mathbf{T}_n$ is nonnegative definte.

In Silverstein and Bai [266], the following theorem is proved.

Theorem 4.3. *Suppose that the entries of* $\mathbf{X}_n$ $(n \times p)$ *are complex random variables that are independent for each* n *and identically distributed for all* n *and satisfy* $\mathrm{E}(|x_{1\,1} - \mathrm{E}(x_{1\,1})|^2) = 1$*. Also, assume that* $\mathbf{T}_n = \mathrm{diag}(\tau_1, \ldots, \tau_p)$*,* τ_i *is real, and the empirical distribution function of* $\{\tau_1, \ldots, \tau_p\}$ *converges almost surely to a probability distribution function* H *as* $n \to \infty$*. The entries* *of both* $\mathbf{X}_n$ *and* $\mathbf{T}_n$ *may depend on* n*, which is suppressed for brevity. Set* $\mathbf{B}_n = \mathbf{A}_n + \frac{1}{n}\mathbf{X}_n\mathbf{T}_n\mathbf{X}_n^*$*, where* $\mathbf{A}_n$ *is Hermitian,* $n \times n$ *satisfying* $F^{\mathbf{A}_n} \to F^A$ *almost surely, where* F^A *is a distribution function (possibly defective) on the real line. Assume also that* $\mathbf{X}_n$*,* $\mathbf{T}_n$*, and* $\mathbf{A}_n$ *are independent. When* $p = p(n)$ *with* $p/n \to y > 0$ *as* $n \to \infty$*, then, almost surely,* $F^{\mathbf{B}_n}$*, the ESD of the eigenvalues of* $\mathbf{B}_n$*, converges vaguely, as* $n \to \infty$*, to a (nonrandom) d.f.* F*, where for any* $z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} : \Im z > 0\}$*, its Stieltjes transform* $s = s(z)$ *is the unique solution in* $\mathbb{C}^+$ *to the equation*

$$s = s_A\left(z - y\int \frac{\tau dH(\tau)}{1+\tau s}\right), \tag{4.1.2}$$

where s_A *is the Stieltjes transform of* F^A*.*

Remark 4.4. Note that Theorem 4.3 is more general than Yin's result in the sense that there is no requirement on the moment convergence of the ESD of $\mathbf{T}_n$ as well as no requirement on the positive definiteness of the matrix $\mathbf{T}_n$. Also, it allows a perturbation matrix $\mathbf{A}_n$ involved in $\frac{1}{n}\mathbf{X}_n^*\mathbf{T}_n\mathbf{X}_n$. However, it is more restrictive than Yin's result in the sense that it requires the matrix $\mathbf{T}_n$ to be diagonal. Weak convergence of (4.1.2) was established in Marčenko and Pastur [201] under higher moment conditions than assumed in Theorem 4.1 but with mild dependence between the entries of $\mathbf{X}_n$.

The proof of Theorem 4.3 uses the Stieltjes transform that will be given in Section 4.5.

4.2 Some Graph Theory and Combinatorial Results

In using the moment approach to establish the existence of the LSD of products of random matrices, we need some combinatorial results related to graph theory.

For a pair of vectors $\mathbf{i} = (i_1, \cdots, i_{2k})'$ $(1 \le i_\ell \le p,\ \ell \le 2k)$ and $\mathbf{j} = (j_1, \cdots, j_k)'$ $(1 \le j_u \le n,\ u \le k)$, construct a graph $Q(\mathbf{i}, \mathbf{j})$ in the following way. Draw two parallel lines, referred to as the I-line and J-line, Plot $i_1, \cdots, i_{2k}$ on the I-line and $j_1, \cdots, j_k$ on the J-line, called the I-vertices and J-vertices, respectively. Draw k down edges from $i_{2\ell-1}$ to j_ℓ, k up edges from j_ℓ to $i_{2\ell}$, and k horizontal edges from $i_{2\ell}$ to $i_{2\ell+1}$ (with the convention that $i_{2k+1} = i_1$). An example of a Q-graph is shown in Fig. 4.1.

Definition. The graph $Q(\mathbf{i}, \mathbf{j})$ defined above is called a ***Q*-graph**; i.e., its vertex set $V = V_i + V_j$, where V_i is the set of distinct numbers of $i_1, \cdots, i_{2k}$ and V_j are the distinct numbers of $j_1, \cdots, j_k$. The edge set $E = \{e_{d\ell}, e_{u\ell}, e_{h\ell}, \ell = 1, \cdots, k\}$, and the function F is defined by $F(e_{d\ell}) = (i_{2\ell-1}, j_\ell)$, $F(e_{u\ell}) = (j_\ell, i_{2\ell})$, and $F(e_{h\ell}) = (i_{2\ell}, i_{2\ell+1})$.

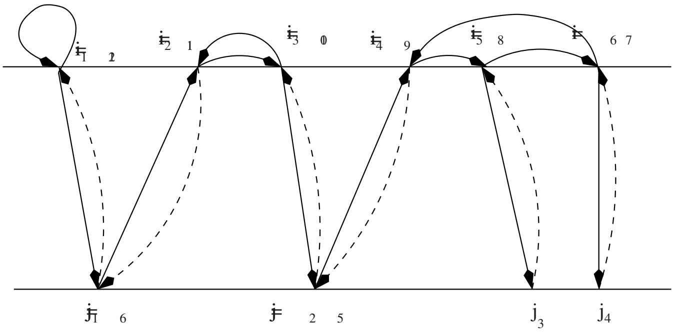

Fig. 4.1 A Q-graph with $k = 6$.

Definition. Let $Q = (V, E, F)$ be a Q-graph. The subgraph of all I-vertices and all horizontal edges of Q is called the **roof** of Q and is denoted by $H(Q)$. Let r equal 1 less than the number of connected components of $H(Q)$.

Definition. Let $Q = (V, E, F)$ be a Q-graph. The M**-minor or the pillar** of Q is defined as the **minor** of Q by contracting all horizontal edges, which means all horizontal edges are removed from Q and all I-vertices connected through horizontal edges are glued together.

Note that a pillar is a Δ-graph. Note that the number of noncoincident I-vertices of the pillar is $1+r$, the same as the number of connected components of the roof of Q. Also, we denote the number of noncoincident J-vertices by s.

We denote the M**-minor or the pillar** of Q by $M(Q)$. If two Q-graphs have isomorphic pillars, then the number of horizontal edges in corresponding connected components of their roofs is equal.

For a given Q-graph Q, glue all coincident vertical edges; namely, we regard all vertical edges with a common I-vertex and J-vertex as one edge. But coincident horizontal edges are still considered different edges. Then, we get an undirectional connected graph of k horizontal edges and m vertical edges. We shall call the resulting graph the **base** of the graph Q and denote it by $B(Q)$.

Definition. For a vertical edge e of $B(Q)$, the number of up (down) vertical edges of Q coincident with e is called the **up (down) multiplicity** of e. The up (down) multiplicity of the ℓ-th vertical edge of $B(Q)$ is denoted by $\mu_\ell(\nu_\ell)$.

We classify the Q-graphs into three categories. Category 1 (denoted by Q_1) contains all Q-graphs that have no single vertical edges and whose pillar $M(Q)$ is a Δ_1-graph. For the definition of Δ_1-graphs, see Subsection 3.1.2. From the definition, one can see that, for a Q_1-graph, each down edge must coincide with one and only one up edge and there are k noncoincident vertical edges. That implies that for each noncoincident vertical edge e_ℓ of a Q_1-graph Q, the multiplicities are $\mu_\ell = \nu_\ell = 1$.

Category 2 (Q_2) contains all graphs that have at least one single vertical edge.

Category 3 (Q_3) contains all other $Q(k, n)$ graphs.

In later applications, one will see that a Q_2-graph corresponds to a zero term and hence needs no further consideration. Let us look further into the graphs of Q_1 and Q_3.

Lemma 4.5. *If $Q \in Q_3$, then the degree of each vertex of $H(Q)$ is not less than 2. Denote the coincidence multiplicities of the ℓ-th noncoincident vertical edge by μ_ℓ and ν_ℓ, $\ell = 1, 2, \cdots, m$, where m is the number of noncoincident vertical edges. Then either there is a $\mu_\ell + \nu_\ell \geq 3$ with $r + s \leq m < k$ or all $\mu_\ell + \nu_\ell = 2$ with $r + s < m = k$.*

If $Q \in Q_1$, then the degree of each vertex of $H(Q)$ is even. In this case, for all ℓ, $\mu_\ell = \nu_\ell = 1$ and $r + s = k$.

Proof. Note that each I-vertex of Q must connect with a vertical edge and a horizontal edge. Therefore, if there is a vertex of $H(Q)$ having degree one, then this vertex connects with only one vertical edge, which is then single. This indicates that the graph Q belongs to Q_2. Since the graph Q is connected, there are at least $r + s$ noncoincident vertical edges to make the graph of $r+1$ disjoint components of $H(Q)$ and s J-vertices connected. This shows that $r + s \leq m$. It is trivial to see that $m \leq k$ because there are in total $2k$ vertical edges and there are no single edges. If, for all ℓ, $\mu_\ell + \nu_\ell = 2$, then $m = k$. If $r + s = k$, then the minor $M(Q)$ is a tree of noncoincident edges, which implies that Q is a Q_1-graph and $\mu_\ell = \nu_\ell = 1$. This violates the assumption that Q is a Q_3-graph. This proves the first conclusion of the lemma.

Note that each down edge of a Q_1-graph coincides with one and only one up edge. Thus, for each Q_1-graph, the degree of each vertex of $H(Q)$ is just twice the number of noncoincident vertical edges of Q connecting with this vertex. Since $M(Q) \in \Delta_1$, for all ℓ, $\mu_\ell = 2$ and $r + s = k$. The proof of the lemma is complete.

As proved in Subsection 3.1.2, for Δ_1-graphs, $s + r = k$. We now begin to count the number of various Δ_1-graphs. Because each edge has multiplicity 2, the degree of an I-vertex (the number of edges connecting to this vertex) must be an even number.

Lemma 4.6. *There are*

$$\frac{k!}{s! i_1! \cdots i_s!}$$

isomorphic classes of Δ_1-graphs that have s J-vertices, $r + 1 = k - s + 1$ I-vertices with degrees (the number of vertical edges connecting this I-vertex) $2\iota_\ell$, $\ell = 1, \cdots, r+1$, where $i_b = \#\{\ell, \iota_\ell = b\}$ denotes the number of I-vertices of degree $2b$ satisfying $i_1 + \cdots + i_s = r + 1$ and $i_1 + 2i_2 + \cdots + s i_s = k$.

Proof. Because $\iota_1 + \cdots + \iota_{r+1} = k$, we have

$$i_1 + \cdots + i_k = r + 1 \quad \text{and} \quad i_1 + 2i_2 + \cdots + ki_k = k.$$

For the canonical Δ_1-graph, the noncoincident edges form a tree. Therefore, there is at most one noncoincident vertical edge directly connecting to a given I-vertex and a given J-vertex; that is to say, an I-vertex of degree $2b$ must connect with b different J-vertices. Therefore, $b \le s$. Consequently, the integer b such that $i_b \neq 0$ is not larger than s, so we can rewrite the constraints above as

$$i_1 + \cdots + i_s = r + 1 \quad \text{and} \quad i_1 + 2i_2 + \cdots + si_s = k.$$

Since the canonical Δ_1-graph has $r+1$ I-vertices with degrees $2\iota_1, \cdots, 2\iota_{r+1}$, we can construct a characteristic sequence of integers while the graph is being formed. After drawing each up edge, place a 1 in the sequence. After drawing a down edge from the ℓ-th I-vertex, if this vertex is never visited again, then put $-\iota_\ell$ in the sequence. Otherwise, put nothing and go to the next up edge. We make the following convention: after drawing the last up edge, put a one and a $-\iota_1$. Then, we get a sequence of k ones and $r + 1$ negative numbers $\{-\iota_2, \cdots, -\iota_{r+1}, -\iota_1\}$. Then we obtain a sequence that consists of negative integers $\{-\iota_2, \cdots, -\iota_{r+1}, -\iota_1\}$ separated by k 1's, and its partial sums are all nonnegative (note the total sum is 0). As an example, for the graph given in Fig. 4.2, the characteristic sequence is

$$1, 1, 1, 1, -3, 1, -2, 1, -1.$$

Conversely, suppose that we are given a characteristic sequence of k ones and $r+1$ negative numbers for which all partial sums are nonnegative and the total sum is zero. We show that there is one and only one canonical Δ_1-graph having this sequence as its characteristic sequence.

In a canonical Δ_1-graph, each down edge must be an innovation except those that complete the preassigned degrees of its I-vertex (see e_{5d}, e_{6d} in Fig. 4.2). Also, all up edges must coincide with the down innovation just prior to it (see edges e_{3u}, e_{4u}, e_{5u}, and e_{6u} in Fig. 4.2), except those that lead to a new I-vertex; i.e., an up innovation. Therefore, if we can determine the up innovations and the down T_3 edges by the given characteristic sequence, then the Δ_1-graph is uniquely determined.

We shall prove the conclusion by induction on r. If $r = 0$ (that is, the characteristic sequence consists of k 1's and ends with $-k$), it is obvious that there is only one I-vertex, which is 1. Then, all down edges are innovations and all up edges are T_3 edges. That is, each up edge coincides with the previous (down) edge. This proves that, if $r = 0$, the Δ_1-graph is uniquely determined by the characteristic sequence.

Now, suppose that $r \ge 1$ and the first negative number is $-a_1$, before which there are p_1 1's. By the condition of nonnegative partial sums, we have

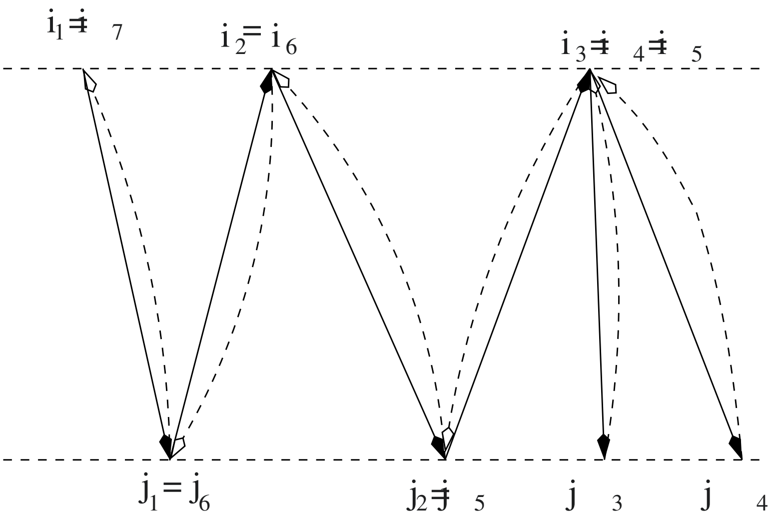

Fig. 4.2 An example of a characteristic sequence.

$p_1 \geq a_1$. By the definition of characteristic sequences, the $p_1 + 1$-st down edge must coincide with an up innovation leading to its end (I-)vertex. By the property of a Δ_1-graph, once the path leaves from an I-vertex through a T_3, the path can never revisit this I-vertex. Therefore, in between this pair of coincident edges, there should be $a_1 - 1$ down innovations and $a_1 - 1$ up T_3 edges that are coincident with the previous innovations. This shows that the $p_1 - a_1 + 1$-st up edge is the up innovation that leads to this I-vertex.

As an example, consider the characteristic sequence defined by Fig. 4.2. $a_1 = 3$ and $p_1 = 4$, by our arguments, the second up edge is an up innovation that leads to the I-vertex $i_3 = i_4 = i_5$, and the third and fourth up edges are T_3 edges.

Now, remove the negative number $-a_1$ and a_1 1's before it from the characteristic sequence. The remainder is still a characteristic sequence with $k-a_1$ ones and r negative numbers. By induction, the positions of up innovations and T_3 up edges can be uniquely determined by the sequence of $k - a_1$ 1's and r negative numbers. That is, there is a Δ_1-graph of $k - a_1$ down edges, and $k - a_1$ up edges, and having the remainder sequence as its characteristic sequence. As for the sequence

$$1, 1, 1, 1, -3, 1, -2, 1, -1,$$

the remainder sequence is

$$1, 1, -2, 1, -1.$$

The Δ_1-graph constructed from the remainder sequence is shown in Fig. 4.3.

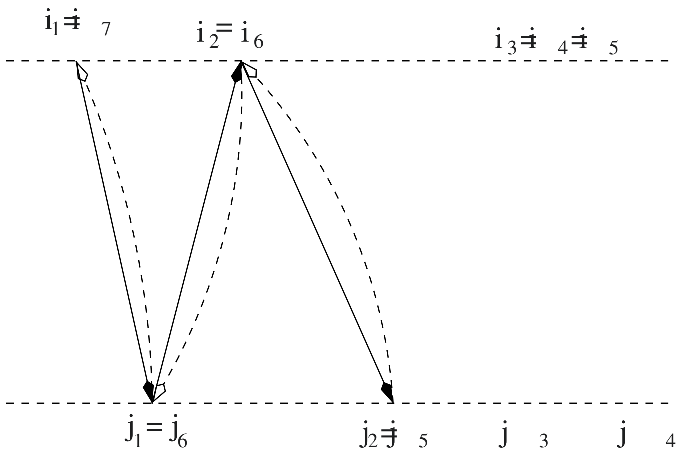

Fig. 4.3 Subgraph corresponding to a shortened sequence.

Then we cut off the graph at the J-vertex between the $\mu_1 - a_1 + 1$-st down edge and the $\mu_1 - a_1 + 1$-st up edge. Then insert an up innovation from this J-vertex, draw $a_1 - 1$ down innovations and T_3 up edges coincident with their previous edges, and finally return to the path through a down T_3 edge by connecting the $\mu_1 - a_1 + 1$-st up edge of the original graph. Then, it is easy to show that the new graph has the given sequence as its characteristic sequence.

Now, we are in a position to count the number of isomorphism classes of Δ_1-graphs with $r + 1$ I-vertices of degrees $2\iota_1, \cdots, 2\iota_{r+1}$, which is the same as the number of characteristic sequences. Place the $r + 1$ negative numbers into the k places after the k 1's. We get a sequence of k 1's and $r+1$ negative numbers. We need to exclude all sequences that do not satisfy the condition of nonnegative partial sums. Ignoring the requirement of nonnegative partial sums, to arrange $\{-\iota_2, \cdots, -\iota_{r+1}, -\iota_1\}$ into k places after the k 1's is equivalent to dropping k -1's into k boxes so that the number of nonempty boxes is $\{\iota_2, \cdots, \iota_{r+1}, \iota_1\}$. Since i_b is the number of b's in the set $\{\iota_1, \cdots, \iota_{r+1}\}$ and the number of empty boxes is $s - 1$, the total number of possible arrangements is

$$\frac{k!}{i_1! \cdots i_s!(s-1)!}.$$

Add a 1 behind the end of the sequence and make the sequence into a circle by connecting its two ends. Then, to complete the proof of the lemma, we need only show that in every s such sequences corresponding to a common circle there is one and only one sequence satisfying the condition that its partial sums be nonnegative. Note that in the circle there are $k + 1$ ones and $r + 1$ negative numbers separated by the ones. Therefore, there are s gaps between consecutive 1's. Cut off the cycle at these gaps. We get s different sequences led and ended by 1's. We show that there is one and only one sequence among the s sequences that has all nonnegative partial sums. Suppose we have a sequence

$$a_1, a_2, \cdots, a_{t-1}, a_t, a_{t+1}, \cdots, a_{k+r+2}$$

for which all partial sums are nonnegative. Obviously, $a_1 = a_{k+r+2} = 1$. Also, we assume that $a_t = a_{t+1} = 1$, which gives a pair of consecutive 1's. Cut the sequence off between a_t and a_{t+1} and construct a new sequence

$$a_{t+1}, \cdots, a_{k+r+2}, a_1, a_2, \cdots, a_{t-1}, a_t.$$

Since $\sum_{i=1}^{k+r+1} a_i = 0$ and $\sum_{i=1}^{t} a_i \geq 1$, the partial sum is

$$a_{t+1} + \cdots + a_{k+r+1} \leq -1.$$

This shows that corresponding to each circle of $k+1$ ones and the $r+1$ negative numbers $\{-\iota_1, \cdots, -\iota_{r+1}\}$ there is at most one sequence whose partial sums are nonnegative.

The final job to conclude the proof of the lemma is to show that for any sequence of $k+1$ ones and the $r+1$ negative numbers summing up to $-k$ where the two ends of the sequence are ones, there exists one sequence of the cut-off circle with nonnegative partial sums. Suppose that we are given the sequence

$$a_1(=1),\ a_2,\ \cdots,\ a_{t-1},\ a_t,\ a_{t+1},\ \cdots,\ a_{k+r+2}(=1),$$

where t is the largest integer such that the partial sum $a_1 + a_2 + \cdots + a_{t-1}$ is the minimum among all partial sums. By the definition of t, we conclude that $a_t = a_{t+1} = 1$. Then, the sequence

$$a_{t+1},\ \cdots,\ a_{k+r+2},\ a_1,\ a_2,\ \cdots,\ a_{t-1},\ a_t$$

satisfies the property that all partial sums be nonnegative. In fact, for any $m \leq k - t + r + 2$, we have

$$a_{t+1} + \cdots + a_{t+m} = (a_1 + \cdots + a_{t+m}) - (a_1 + \cdots + a_t) \geq 0,$$

and for any $k - t + r + 2 < m \leq k + r + 2$, we have

$$\begin{aligned} a_{t+1} + \cdots + a_{t+m} &= (a_1 + \cdots + a_{t+m}) - (a_1 + \cdots + a_t) \\ &= 1 + (a_1 + \cdots + a_{t+m-k-r-2}) - (a_1 + \cdots + a_t) \geq 0. \end{aligned}$$

The proof of the lemma is complete.

4.3 Proof of Theorem 4.1

Again, the proof will rely on the MCT, preceded by truncation and renormalization of the entries of $\mathbf{X}_n$. Additional steps will be taken to reduce the assumption on $\mathbf{T}_n$ to be nonrandom and to truncate its ESD.

4.3.1 Truncation of the ESD of $\mathbf{T}_n$

For brevity, we shall suppress the superscript n from the x-variables.

Step 1. Reducing to the case where $\mathbf{T}_n$'s are nonrandom
If the ESD of $\mathbf{T}_n$ converges to a limit F^T almost surely, we may consider the LSD of $\mathbf{S}_n\mathbf{T}_n$ conditioned on all $\mathbf{T}_n$ as given and hence may assume that $\mathbf{T}_n$ is nonrandom. Then the final result follows by Fubini's theorem. If the convergence is in probability, then we may use the subsequence method or use the strong representation theorem (see Skorohod [270] or Dudley [96]).[1] The strong representation theorem says that there is a probability space on which we can define a sequence of random matrices $(\widetilde{\mathbf{X}}_n, \widetilde{\mathbf{T}}_n)$ such that, for each n, the joint distribution of $(\widetilde{\mathbf{X}}_n, \widetilde{\mathbf{T}}_n)$ is identical to that of $(\mathbf{X}_n, \mathbf{T}_n)$ and the ESD of $\widetilde{\mathbf{T}}_n$ converges to F^T almost surely. Therefore, to prove Theorem 4.1, it suffices to show it for the case of a.s. convergence.

Now, suppose that $\mathbf{T}_n$ are nonrandom and that the ESD of $\mathbf{T}_n$ converges to F^T.

Step 2. Truncation of the ESD of $\mathbf{T}_n$
Suppose that the spectral decomposition of $\mathbf{T}_n$ is $\sum_{i=1}^p \lambda_{in}\mathbf{u}_i\mathbf{u}_i^*$. Define a matrix $\widetilde{\mathbf{T}}_n = \sum_{i=1}^p \tilde{\lambda}_{in}\mathbf{u}_i\mathbf{u}_i^*$, where $\tilde{\lambda}_{in} = \lambda_{in}$ or zero in accordance with whether $|\lambda_{in}| \le \tau_0$ or not, where τ_0 is prechosen to be constant such that both $\pm\tau_0$ are continuity points of F^T. Then, the ESD of $\widetilde{\mathbf{T}}_n$ converges to the limit

$$F_{T,\tau_0}(x) = \int_{-\infty}^x I_{[-\tau_0,\tau_0]}(u)F^T(du) + (F^T(-\tau_0) + 1 - F^T(\tau_0))I_{[0,\infty)}(x),$$

and (4.1.1) is true for $\widetilde{\mathbf{T}}_n$ with $\tilde{H}_k = \int_{|x|\le\tau_0} x^k dF^T(x)$.

Applying Theorem A.43, we obtain

---

[1] In an unpublished work by Bai et al [19], Skorohod's result was generalized to: *Suppose that μ_n is a probability measure defined on a Polish space (i.e., a complete and separable metric space) S_n and φ_n is a measurable mapping from S_n to another Polish space S_0. If $\mu_n\varphi_n^{-1}$ tends to μ_0 weakly, where μ_0 is a probability measure defined on the space S_0, then there exists a probability space $(\Omega, \mathcal{F}, P)$ on which we have random mappings $X_n : \Omega \mapsto S_n$, such that μ_n is the distribution of X_n and $\varphi_n(X_n) \to X_0$ almost surely.* Skorohod's result is the special case where all S_n are identical to S_0 and all $\varphi_n(x) = x$.

$$\left\| F^{\mathbf{S}_n \mathbf{T}_n} - F^{\mathbf{S}_n \widetilde{\mathbf{T}}_n} \right\| \leq \frac{1}{p} \text{rank}(\mathbf{T}_n - \widetilde{\mathbf{T}}_n) \to F^T(-\tau_0) + 1 - F^T(\tau_0) \quad (4.3.1)$$

as $n \to \infty$. Note that the right-hand side of the inequality above can be made arbitrarily small if τ_0 is large enough.

We claim that Theorem 4.1 follows if we can prove that, with probability 1, $F^{\mathbf{S}_n \widetilde{\mathbf{T}}_n}$ converge to a nondegenerate distribution F_{τ_0} for each fixed τ_0. We shall prove this assertion by the following two lemmas.

Lemma 4.7. *If the distribution family $\{F^{\mathbf{S}_n \widetilde{\mathbf{T}}_n}\}$ is tight for every $\tau_0 > 0$, then so is the distribution family $\{F^{\mathbf{S}_n \mathbf{T}_n}\}$.*

Proof. Since $F^{\mathbf{T}_n} \to F^T$, for each fixed $\varepsilon \in (0,1)$, we can select a $\tau_0 > 0$ such that, for all n, $F^{\mathbf{T}_n}(-\tau_0) + 1 - F^{\mathbf{T}_n}(\tau_0) < \varepsilon/3$. On the other hand, we can select $M > 0$ such that, for all n, $F^{\mathbf{S}_n \widetilde{\mathbf{T}}_n}(-M) + 1 - F^{\mathbf{S}_n \widetilde{\mathbf{T}}_n}(M) < \varepsilon/3$ because $\{F^{\mathbf{S}_n \widetilde{\mathbf{T}}_n}\}$ is tight. Thus, we have

$$\begin{aligned} & F^{\mathbf{S}_n \mathbf{T}_n}(M) - F^{\mathbf{S}_n \mathbf{T}_n}(-M) \\ \geq & F^{\mathbf{S}_n \widetilde{\mathbf{T}}_n}(M) - F^{\mathbf{S}_n \widetilde{\mathbf{T}}_n}(-M) - 2||F^{\mathbf{S}_n \mathbf{T}_n} - F^{\mathbf{S}_n \widetilde{\mathbf{T}}_n}|| \\ \geq & 1 - \varepsilon/3 - 2(F^{\mathbf{T}_n}(-\tau_0) + 1 - F^{\mathbf{T}_n}(\tau_0)) \geq 1 - \varepsilon. \end{aligned}$$

This proves that the family of $\{F^{\mathbf{S}_n \mathbf{T}_n}\}$ is tight.

Lemma 4.8. *If $F^{\mathbf{S}_n \widetilde{\mathbf{T}}_n} \to F_{\tau_0}$, a.s., for each $\tau_0 > 0$, then*

$$F^{\mathbf{S}_n \mathbf{T}_n} \to F, \ a.s.,$$

for some distribution F.

Proof. Since the convergence of $F^{\mathbf{S}_n \widetilde{\mathbf{T}}_n} \to F_{\tau_0}$ a.s. implies the tightness of the distribution family $\{F^{\mathbf{S}_n \mathbf{T}_n}\}$, by Lemma 4.7, the distribution family $\{F^{\mathbf{S}_n \mathbf{T}_n}\}$ is also tight. Therefore, for any subsequence of $\{F^{\mathbf{S}_n \mathbf{T}_n}\}$, there is a convergent subsequence of the previous subsequence of $\{F^{\mathbf{S}_n \mathbf{T}_n}\}$. Therefore, to complete the proof of Theorem 4.1, we need only show that the sequence $\{F^{\mathbf{S}_n \mathbf{T}_n}\}$ has a unique subsequence limit.

Suppose that $F^{(1)}$ and $F^{(2)}$ are two limiting distributions of two convergent subsequences of $\{F^{\mathbf{S}_n \mathbf{T}_n}\}$. Suppose that x is a common continuity point of $F^{(1)}$, $F^{(2)}$, and F_{τ_0} for all rational τ_0. Then, for any fixed $\varepsilon > 0$, we can select a rational τ_0 such that $F^{\mathbf{T}_n}(-\tau_0) + 1 - F^{\mathbf{T}_n}(\tau_0) < \varepsilon/5$ for all n. Since $F^{\mathbf{S}_n \widetilde{\mathbf{T}}_n} \to F_{\tau_0}$, there exists an n_0, such that for all $n_1, n_2 > n_0$, $|F^{\mathbf{S}_{n_1} \widetilde{\mathbf{T}}_{n_1}}(x) - F^{\mathbf{S}_{n_2} \widetilde{\mathbf{T}}_{n_2}}(x)| < \varepsilon/5$. Also, we can select $n_1, n_2 > n_0$ such that $|F^{(j)}(x) - F^{\mathbf{S}_{n_j} \mathbf{T}_{n_j}}(x)| < \varepsilon/5$, $j = 1, 2$. Thus,

$$|F^{(1)}(x) - F^{(2)}(x)|$$

$$\leq \sum_{j=1}^{2}[|F^{(j)}(x) - F^{\mathbf{S}_{n_j}\mathbf{T}_{n_j}}(x)| + \|F^{\mathbf{S}_{n_j}\mathbf{T}_{n_j}} - F^{\mathbf{S}_{n_j}\widetilde{\mathbf{T}}_{n_j}}\|]$$
$$+|F^{\mathbf{S}_{n_1}\widetilde{\mathbf{T}}_{n_1}}(x) - F^{\mathbf{S}_{n_2}\widetilde{\mathbf{T}}_{n_2}}(x)| < \varepsilon.$$

This shows that $F^{(1)} \equiv F^{(2)}$, and the proof of the lemma is complete. It is easy to see from the proof that $\lim_{\tau_0\to\infty} F_{\tau_0}$ exists and is equal to F, the a.s. limit of $F^{\mathbf{S}_n\mathbf{T}_n}$.

Therefore, we may truncate the ESD of $F^{\mathbf{T}_n}$ first and then proceed to the proof with the truncated matrix $\mathbf{T}_n$. For brevity, we still use $\mathbf{T}_n$ for the truncated matrix $\mathbf{T}_n$, that is; we shall assume that the eigenvalues of $\mathbf{T}_n$ are bounded by a constant, say τ_0.

4.3.2 Truncation, Centralization, and Rescaling of the X-variables

Following the truncation technique used in Section 3.2, let $\widetilde{\mathbf{X}}_n$ and $\widetilde{\mathbf{S}}_n$ denote the sample matrix and the sample covariance matrix defined by the truncated variables at the truncation location $\eta_n\sqrt{n}$. Note that $\mathbf{S}_n\mathbf{T}_n$ and $\frac{1}{n}\mathbf{X}_n^*\mathbf{T}_n\mathbf{X}_n$ have the same set of nonzero eigenvalues, as do the matrices $\widetilde{\mathbf{S}}_n\mathbf{T}_n$ and $\frac{1}{n}\widetilde{\mathbf{X}}_n^*\mathbf{T}_n\widetilde{\mathbf{X}}_n$. Thus,

$$\|F^{\mathbf{S}_n\mathbf{T}_n} - F^{\widetilde{\mathbf{S}}_n\mathbf{T}_n}\|$$
$$= \frac{n}{p}\|F^{\frac{1}{n}\mathbf{X}_n^*\mathbf{T}_n\mathbf{X}_n} - F^{\frac{1}{n}\widetilde{\mathbf{X}}_n^*\mathbf{T}_n\widetilde{\mathbf{X}}_n}\| = \frac{n}{p}\|F^{\mathbf{X}_n^*\mathbf{T}_n\mathbf{X}_n} - F^{\widetilde{\mathbf{X}}_n^*\mathbf{T}_n\widetilde{\mathbf{X}}_n}\|.$$

Then, by Theorem A.43, for any $\varepsilon > 0$, we have

$$\begin{aligned}
&\mathrm{P}\left(\left\|F^{\mathbf{S}_n\mathbf{T}_n} - F^{\widetilde{\mathbf{S}}_n\mathbf{T}_n}\right\| \geq \varepsilon\right)\\
&= \mathrm{P}\left(\left\|F^{\mathbf{X}_n^*\mathbf{T}_n\mathbf{X}_n} - F^{\widetilde{\mathbf{X}}_n^*\mathbf{T}_n\widetilde{\mathbf{X}}_n}\right\| \geq \varepsilon p/n\right)\\
&\leq \mathrm{P}(\mathrm{rank}(\mathbf{X}_n^*\mathbf{T}_n\mathbf{X}_n - \widetilde{\mathbf{X}}_n^*\mathbf{T}_n\widetilde{\mathbf{X}}_n) \geq \varepsilon p)\\
&\leq \mathrm{P}(2\mathrm{rank}(\mathbf{X}_n - \widetilde{\mathbf{X}}_n) \geq \varepsilon p)\\
&\leq \mathrm{P}\left(\sum_{ij} I_{\{|x_{ij}|\geq\eta_n\sqrt{n}\}} \geq \varepsilon p/2\right).
\end{aligned}$$

From the condition (3.2.1), one can easily see that

$$\mathrm{E}\left(\sum_{ij} I_{\{|x_{ij}|\geq\eta_n\sqrt{n}\}}\right) \leq \frac{1}{\eta_n^2 n}\sum_{ij}\mathrm{E}|x_{ij}|^2 I_{\{|x_{ij}|\geq\eta_n\sqrt{n}\}} = o(p)$$

and

$$\mathrm{Var}\left(\sum_{ij} I_{\{|x_{ij}|\geq\eta_n\sqrt{n}\}}\right) \leq \frac{1}{\eta_n^2 n}\sum_{ij}\mathrm{E}|x_{ij}|^2 I_{\{|x_{ij}|\geq\eta_n\sqrt{n}\}} = o(p).$$

Then, applying Bernstein's inequality, one obtains

$$\mathrm{P}\left(\left\|F^{\mathbf{S}_n\mathbf{T}_n} - F^{\widetilde{\mathbf{S}}_n\mathbf{T}_n}\right\| \geq \varepsilon\right) \leq 2\exp\left(-\frac{1}{8}\varepsilon^2 p\right), \tag{4.3.2}$$

which is summable. By the Borel-Cantelli lemma, we conclude that, with probability 1,

$$\left\|F^{\mathbf{S}_n\mathbf{T}_n} - F^{\widetilde{\mathbf{S}}_n\mathbf{T}_n}\right\| \to 0. \tag{4.3.3}$$

We may do the centralization and rescaling of the X-variables in the same way as in Section 3.2. We leave the details to the reader.

4.3.3 Completing the Proof

Therefore, the proof of Theorem 4.1 can be done under the following additional conditions:

$$\begin{aligned} &\|\mathbf{T}_n\| \leq \tau_0, \\ &|x_{jk}| \leq \eta_n\sqrt{n}, \\ &\mathrm{E}(x_{jk}) = 0, \\ &\mathrm{E}|x_{jk}|^2 = 1. \end{aligned} \tag{4.3.4}$$

Now, we will proceed in the proof of Theorem 4.1 by applying the MCT under the additional conditions above. We need to show the convergence of the spectral moments of $\mathbf{S}_n\mathbf{T}_n$. We have

$$\begin{aligned} \beta_k(\mathbf{S}_n\mathbf{T}_n) &= \frac{1}{p}\mathrm{E}[(\mathbf{S}_n\mathbf{T}_n)^k] \\ &= p^{-1}n^{-k}\sum x_{i_1j_1}\overline{x}_{i_2j_1}t_{i_2i_3}x_{i_3j_2}\cdots x_{i_{2k-1}j_k}\overline{x}_{i_{2k}j_k}t_{i_{2k}i_1} \\ &= p^{-1}n^{-k}\sum_{\mathbf{i},\mathbf{j}} T(H(\mathbf{i}))X(Q(\mathbf{i},\mathbf{j})), \end{aligned} \tag{4.3.5}$$

where $Q(\mathbf{i},\mathbf{j})$ is the Q-graph defined by $\mathbf{i} = (i_1,\cdots,i_{2k})$ and $\mathbf{j} = (j_1,\cdots,j_k)$ and $H(\mathbf{i})$ is the roof of $Q(\mathbf{i},\mathbf{j})$. We shall prove the theorem by showing the following lemma.

Lemma 4.9. *We have*

$$\mathrm{E}\beta_k(\mathbf{S}_n\mathbf{T}_n) \to \beta_k^{st} = \sum_{s=1}^{k} y^{k-s} \sum_{\substack{i_1+\cdots+i_s=k-s+1 \\ i_1+\cdots+si_s=k}} \frac{k!}{s!} \prod_{m=1}^{s} \frac{H_m^{i_m}}{i_m!}, \tag{4.3.6}$$

$$\mathrm{E}|\beta_k(\mathbf{S}_n\mathbf{T}_n) - \mathrm{E}\beta_k(\mathbf{S}_n\mathbf{T}_n)|^4 = O(n^{-2}), \tag{4.3.7}$$

and the β_k^{st}*'s satisfy the Carleman condition.*

Proof. We first prove (4.3.6). Write

$$\mathrm{E}\beta_k(\mathbf{S}_n\mathbf{T}_n) = p^{-1}n^{-k} \sum_{Q} \sum_{Q(\mathbf{i},\mathbf{j})\in Q} T(H(\mathbf{i}))X(Q(\mathbf{i},\mathbf{j})), \tag{4.3.8}$$

where the first summation is taken for all canonical graphs and the second for all graphs isomorphic to the given canonical graph Q. Glue all coincident vertical edges of Q, and denote the resulting graph as Q_{gl}. Let each horizontal edge associate with the matrix $\mathbf{T}_n$. If a vertical edge of Q_{gl} consists of μ up edges and ν down edges, then associated with this edge is the matrix

$$\mathbf{T}(\mu,\nu) = \left[\mathrm{E}x_{ij}^{\nu}\bar{x}_{ij}^{\mu}\right]_{p\times n}.$$

We call $\mu+\nu$ the multiplicity of the vertical edge of Q_{gl}.

Since the operator norm of a matrix is less than or equal to its Euclidean norm, it is easy to verify that

$$\|\mathbf{T}(\mu,\nu)\| \le \begin{cases} 0, & \text{if } \mu+\nu=1, \\ o(n^{(\mu+\nu)/2}), & \text{if } \mu+\nu>2, \\ \max(n,p), & \text{if } \mu+\nu=2. \end{cases} \tag{4.3.9}$$

One can also verify that $\|\mathbf{T}(\mu,\nu)\|_0$ satisfies the same inequality, where the definition of the norm $\|\cdot\|_0$ can be found in Theorem A.35.

Split the sum in (4.3.8) according to the three categories of the Δ-graphs. If $Q \in Q_2$ (i.e., it contains a single vertical edge), the corresponding term is 0. Hence, the sum corresponding to Q_2 is 0.

Next, we consider the sum corresponding to Q_3. For a given canonical graph $Q \in Q_3$, using the notation defined in Lemma 4.5, by Lemma 4.5 and Theorem A.35, we have

$$\begin{aligned} &\frac{1}{pn^k}\left|\sum_{Q(\mathbf{i},\mathbf{j})\in Q} T(H(\mathbf{i}))X(Q(\mathbf{i},\mathbf{j}))\right| \\ &\le \begin{cases} o(1)n^{-k-1}n^{\frac{1}{2}\left(\sum_{i=1}^{m}(\mu_i+\nu_i-2)\right)}n^{r+s+1} & \text{if for some } i,\ \mu_i+\nu_i>2 \\ Cn^{-k-1}n^{\frac{1}{2}\left(\sum_{i=1}^{m}(\mu_i+\nu_i-2)\right)}n^{r+s+1} & \text{if for all } i,\ \mu_i+\nu_i=2 \end{cases} \\ &= o(1), \end{aligned} \tag{4.3.10}$$

where we have used the fact that $\sum_{i=1}^{m}(\mu_i + \nu_i) = 2k$ and $r+s < m = k$ for the second case. Because the number of canonical graphs is bounded for fixed k, we have proved that the sum corresponding to Q_3 tends to 0.

Finally, we consider the sum corresponding to all Q_1-graphs, those terms corresponding to canonical graphs with vertical edge multiplicities $\mu = \nu = 1$. This condition implies that the expectation factor $X(Q(\mathbf{i},\mathbf{j})) \equiv 1$. Then, (4.3.5) reduces to

$$\beta_k(\mathbf{S}_n\mathbf{T}_n) = p^{-1}n^{-r}\sum_{\mathbf{i}} T(H(\mathbf{i})) + o(1), \tag{4.3.11}$$

where the summation runs over all possible heads of Q_1-graphs.

Denote the number of disjoint connected components of the head $H(Q)$ of a canonical Q_1-graph by $r+1$ and the sizes (the number of edges) of the connected components of $H(Q)$ by $\iota_1, \cdots, \iota_{r+1}$. We will show that

$$p^{-1}n^{-r}\sum_{H(\mathbf{i})_\iota} T(H(\mathbf{i})) \to y^r H_{\iota_1}\cdots H_{\iota_{r+1}}, \tag{4.3.12}$$

where the summation $\sum_{H(\mathbf{i})_\iota}$ runs over an isomorphic class of heads $H(\mathbf{i})$ of Q_1-graph with indices $\{\iota_1, \cdots, \iota_{r+1}\}$ and

$$H_\ell = \int_{-\tau_0}^{\tau_0} t^\ell dF^T(t).$$

By Lemma 4.6, we have

$$\beta_k(\mathbf{S}_n\mathbf{T}_n) = p^{-1}\sum_{r=0}^{k-1} n^{-r}\sum_{i}\frac{k!}{i_1!\cdots i_s! s!}\sum_{H(\mathbf{i})_\iota} H(\mathbf{i}) + o(1), \tag{4.3.13}$$

where the summation $\sum_i$ runs over all solutions of the equations

$$i_1 + \cdots + i_s = r+1 \quad \text{and} \quad i_1 + 2i_2 + \cdots + s i_s = k.$$

When a roof of a canonical Q_1-graph consists of $1+r$ connected components with sizes $\iota_1, \cdots, \iota_{r+1}$, by the inclusion-exclusion principle we conclude that

$$\sum_{H(\mathbf{i})_\iota}(t)_{H(\mathbf{i})} = \prod_{\ell=1}^{r+1}(\mathrm{tr}\mathbf{T}^{\nu_\ell})(1+o(1)) = p^{1+r}\left[\prod_{\ell=1}^{r+1} H_{\nu_\ell} + o(1)\right],$$

which proves (4.3.12). Combining (4.3.10), (4.3.12), and (4.3.13), we obtain

$$\frac{1}{p}\mathrm{E}[(\mathbf{ST})^k] = \sum_{s=1}^{k} y^{k-s} \sum_{\substack{i_1+\cdots+i_s=k-s+1\\ i_1+\cdots+si_s=k}} \frac{k!}{s!}\prod_{m=1}^{s}\frac{H_m^{i_m}}{i_m!} + o(1). \quad (4.3.14)$$

This completes the proof of (4.3.6).

Next, we prove (4.3.7). Similar to the proof of (4.3.5), for given $i_1, \cdots, i_{8k}$ taking values in $\{1, 2, \cdots, p\}$ and $j_1, \cdots, j_{4k}$ taking values in $\{1, 2, \cdots, n\}$, and for each $\ell = 1, 2, 3, 4$, we construct a Q-graph G_ℓ with the indices $\mathbf{i}_\ell = (i_{2(\ell-1)k+1}, \cdots, i_{2\ell k})$ and $\mathbf{j}_\ell = (j_{(\ell-1)k+1}, \cdots, j_{\ell_k})$. We then have

$$\begin{aligned}
&\mathrm{E}\left(\left|\frac{1}{p}\mathrm{tr}[(\mathbf{ST})^k] - \mathrm{E}\left(\frac{1}{p}\mathrm{tr}[(\mathbf{ST})^k]\right)\right|^4\right)\\
&= p^{-4}n^{-4k}\sum_{\mathbf{i}_1,\mathbf{j}_1,\cdots,\mathbf{i}_4,\mathbf{j}_4}\left[\mathrm{E}\left(\prod_{\ell=1}^{4}(tx)_{G_\ell(\mathbf{i}_\ell,\mathbf{j}_\ell)}\right) - \left(\prod_{\ell=1}^{4}\mathrm{E}((tx)_{G_\ell(\mathbf{i}_\ell,\mathbf{j}_\ell)})\right)\right], \quad (4.3.15)
\end{aligned}$$

where

$$(tx)_{G_\ell(\mathbf{i}_\ell;\mathbf{j}_\ell)} = \prod_{\ell=1}^{k}\left(t_{i_{f_\ell(2\ell)},i_{f_\ell(2\ell+1)}} x_{i_{f_\ell(2\ell-1)},j_{g((\ell-1)k+\ell)}} \overline{x}_{i_{f_\ell(2\ell)},j_{g((\ell-1)k+\ell)}}\right).$$

If, for some $\ell = 1, 2, 3, 4$, all vertical edges of G_ℓ do not coincide with any vertical edges of the other three graphs, then

$$\left[\mathrm{E}\left(\prod_{\ell=1}^{4}(tx)_{G_\ell}(\mathbf{i}_\ell,\mathbf{j}_\ell)\right) - \left(\prod_{\ell=1}^{4}\mathrm{E}((tx)_{G_\ell}(\mathbf{i}_\ell,\mathbf{j}_\ell))\right)\right] = 0$$

due to the independence of the X-variables. Therefore, $G = \cup G_\ell$ consists of either one or two connected components. Similar to the proof of (4.3.6), applying the second part of Theorem A.35, the sum of terms corresponding to graphs G of two connected components has the order of $O(n^{4k+2})$, while the sum of terms corresponding to a connected graphs G has the order of $O(n^{4k+1})$. From this, (4.3.7) follows.

Finally, we verify the Carleman condition. By elementary calculation, we have

$$\left|\beta_k^{st}\right| \le \tau_0^k(1+\sqrt{y})^{2k},$$

which yields the Carleman condition. The proof of Lemma 4.9 is complete.

From (4.3.14) and (4.3.7), it follows that with probability 1

$$\frac{1}{p}\mathrm{tr}[(\mathbf{ST})^k] \to \beta_k^{st} = \sum_{s=1}^{k} y^{k-s} \sum_{\substack{i_1+\cdots+i_s=k-s+1\\ i_1+\cdots+si_s=k}} \frac{k!}{s!}\prod_{m=1}^{s}\frac{H_m^{i_m}}{i_m!}.$$

Applying the MCT, we obtain that, with probability 1, the ESD of $\mathbf{ST}$ tends to the nonrandom distribution determined by the moments β_k^{st}.

The proof of the theorem is complete.

4.4 LSD of the *F*-Matrix

In this section, we shall derive the LSD of a multivariate F-matrix.

Theorem 4.10. *Let* $\mathbf{F} = \mathbf{S}_{n1}\mathbf{S}_{n2}^{-1}$*, where* $\mathbf{S}_{ni}$ $(i = 1, 2)$ *is a sample covariance matrix with dimension* p *and sample size* n_i *with an underlying distribution of mean 0 and variance 1. If* $\mathbf{S}_{n1}$ *and* $\mathbf{S}_{n2}$ *are independent,* $p/n_1 \to y \in (0, \infty)$ *and* $p/n_2 \to y' \in (0, 1)$*. Then the LSD* $F_{y,y'}$ *of* $\mathbf{F}$ *exists and has a density function given by*

$$F'_{y,y'}(x) = \begin{cases} \frac{(1-y')\sqrt{(b-x)(x-a)}}{2\pi x(y+xy')}, & \text{when } a < x < b, \\ 0, & \text{otherwise,} \end{cases} \tag{4.4.1}$$

where $a = \left(\frac{1-\sqrt{y+y'-yy'}}{1-y'}\right)^2$ *and* $b = \left(\frac{1+\sqrt{y+y'-yy'}}{1-y'}\right)^2$.

Further, if $y > 1$*, then* F_{st} *has a point mass* $1 - 1/y$ *at the origin.*

Remark 4.11. If $\mathbf{S}_{n2} = \frac{1}{n_2}\mathbf{X}_{n2}\mathbf{X}_{n2}^*$ and the entries of $\mathbf{X}_{n2}$ come from a double array of iid random variables having finite fourth moment, under the condition $y' \in (0, 1)$, it will be proven in the next chapter that, with probability 1, the smallest eigenvalue of $\mathbf{S}_{n2}$ has a positive limit and thus $\mathbf{S}_{n2}^{-1}$ is well defined. Then, the existence of F_{st} follows from Theorems 3.6 and 4.1. If the fourth moment does not exist, then $\mathbf{S}_{n2}^{-1}$ may not exist. In this case, $\mathbf{S}_{n2}^{-1}$ should be understood as the generalized Moore-Penrose inverse, and the conclusion of Theorem 4.10 remains true.

Proof. We first derive the generating function for the LSD of $\mathbf{S}_n\mathbf{T}_n$ in the next subsection. We use it to derive the Stieltjes transform of the LSD of multivariate F-matrices in the last subsection.

4.4.1 Generating Function for the LSD of $\mathbf{S}_n\mathbf{T}_n$

We compute the generating function $g(z) = 1 + \sum_{k=1}^{\infty} z^k \beta_k^{st}$ of the LSD F^{st} of the matrix sequence $\{\mathbf{S}_n\mathbf{T}_n\}$ where the moments β_k^{st} are given by (4.3.6). For $k \geq 1$, β_k^{st} is the coefficient of z^k in the Taylor expansion of

$$\sum_{s=0}^{k} y^{k-s} \frac{k!}{s!(k-s+1)!} \left(\sum_{\ell=1}^{\infty} z^\ell H_\ell\right)^{k-s+1} + \frac{1}{y(k+1)}$$

$$= \frac{1}{y(k+1)} \left[1 + y \sum_{\ell=1}^{\infty} z^\ell H_\ell \right]^{k+1}, \tag{4.4.2}$$

where H_k are the moments of the LSD H of $\mathbf{T}_n$. Therefore, β_k^{st} can be written as

$$\beta_k^{st} = \frac{1}{2\pi i y(k+1)} \oint_{|\zeta|=\rho} \zeta^{-k-1} \left[1 + y \sum_{\ell=1}^{\infty} \zeta^\ell H_\ell \right]^{k+1} d\zeta$$

for any $\rho \in (0, 1/\tau_0)$, which guarantees the convergence of the series $\sum \zeta^\ell H_\ell$.

Using the expression above, we can construct a generating function of β_k^{st} as follows. For all small z with $|z| < 1/\tau_0 b$, where $b = (1+\sqrt{y})^2$,

$$g(z) - 1 = \frac{1}{2\pi i y} \oint_{|\zeta|=\rho} \sum_{k=1}^{\infty} \frac{1}{k+1} z^k \zeta^{-1-k} \left(1 + y \sum_{\ell=1}^{\infty} \zeta^\ell H_\ell \right)^{k+1} d\zeta$$

$$= \frac{1}{2\pi i y} \oint_{|\zeta|=\rho} \left[-\zeta^{-1} - y \sum_{\ell=1}^{\infty} \zeta^{\ell-1} H_\ell - \frac{1}{z} \log \left(1 - z\zeta^{-1} - zy \sum_{\ell=1}^{\infty} \zeta^{\ell-1} H_\ell \right) \right] d\zeta$$

$$= -\frac{1}{y} - \frac{1}{2\pi i y z} \oint_{|\zeta|=\rho} \log \left(1 - z\zeta^{-1} - zy \sum_{\ell=1}^{\infty} \zeta^{\ell-1} H_\ell \right) d\zeta.$$

The exchange of summation and integral is justified provided that $|z| < \rho/(1 + y\sum \rho^\ell |H_\ell|)$. Therefore, we have

$$g(z) = 1 - \frac{1}{y} - \frac{1}{2\pi i y z} \oint_{|\zeta|=\rho} \log \left(1 - z\zeta^{-1} - zy \sum_{\ell=1}^{\infty} \zeta^{\ell-1} H_\ell \right) d\zeta. \tag{4.4.3}$$

Let $s_F(z)$ and $s_H(z)$ denote the Stieltjes transforms of F^{st} and H, respectively. It is easy to verify that

$$-\frac{1}{z} s_F\left(\frac{1}{z}\right) = 1 + \sum_{k=1}^{\infty} z^k \beta_k^{st},$$

$$-\frac{1}{z} s_H\left(\frac{1}{z}\right) = 1 + \sum_{k=1}^{\infty} z^k H_k.$$

Then, from (4.4.3) it follows that

$$\frac{1}{z} s_F\left(\frac{1}{z}\right) = \frac{1}{y} - 1 + \frac{1}{2\pi i y z} \oint_{|\zeta|=\rho} \log \left(1 - z\zeta^{-1} + \zeta^{-1} zy + \zeta^{-2} zy s_H\left(\frac{1}{\zeta}\right) \right) d\zeta. \tag{4.4.4}$$

4.4.2 Completing the Proof of Theorem 4.10

Now, let us use (4.4.4) to derive the LSD of general multivariate F-matrices. A multivariate F-matrix is defined as a product of $\mathbf{S}_n$ with the inverse of another covariance matrix; i.e., $\mathbf{T}_n$ is the inverse of another covariance matrix with dimension p and degrees of freedom n_2. To guarantee the existence of the inverse matrix, we assume that $p/n_2 \to y' \in (0,1)$. In this case, it is easy to verify that H will have a density function

$$H'(x) = \begin{cases} \frac{1}{2\pi y' x^2}\sqrt{(xb'-1)(1-a'x)}, & \text{if } \frac{1}{b'} < x < \frac{1}{a'}, \\ 0, & \text{otherwise}, \end{cases}$$

where $a' = (1-\sqrt{y'})^2$ and $b' = (1+\sqrt{y'})^2$. Noting that the k-th moment of H is the $-k$-th moment of the Marčenko-Pastur law with index y', one can verify that

$$\zeta^{-1} s_H\left(\frac{1}{\zeta}\right) = -\zeta s_{y'}(\zeta) - 1,$$

where $s_{y'}$ is the Stieltjes transform of the M-P law with index y'. Thus,

$$s_F(z) = \frac{1}{yz} - \frac{1}{z} + \frac{1}{2\pi i y}\oint_{|\zeta|=\rho} \log(z - \zeta^{-1} - y s_{y'}(\zeta)) d\zeta. \tag{4.4.5}$$

By (3.3.1), we have

$$s_{y'}(\zeta) = \frac{1 - y' - \zeta + \sqrt{(1+y'-\zeta)^2 - 4y'}}{2y'\zeta}. \tag{4.4.6}$$

By integration by parts, we have

$$\begin{aligned} &\frac{1}{2\pi i y}\oint_{|\zeta|=\rho} \log(z - \zeta^{-1} - y s_{y'}(\zeta)) d\zeta \\ &= -\frac{1}{2\pi i y}\oint_{|\zeta|=\rho} \zeta \frac{\zeta^{-2} - y s'_{y'}(\zeta)}{z - \zeta^{-1} - y s_{y'}(\zeta)} d\zeta \\ &= -\frac{1}{2\pi i y}\oint_{|\zeta|=\rho} \frac{1 - y\zeta^2 s'_{y'}(\zeta)}{z\zeta - 1 - y\zeta s_{y'}(\zeta)} d\zeta. \end{aligned} \tag{4.4.7}$$

For easy evaluation of the integral, we make a variable change from ζ to s.

Note that $s_{y'}$ is a solution of the equation (see (3.3.8) with $\delta = 0$)

$$s = \frac{1}{1 - \zeta - y' - \zeta y' s}. \tag{4.4.8}$$

From this, we have

$$\zeta = \frac{s - sy' - 1}{s + s^2 y'},$$
$$\frac{ds}{d\zeta} = \frac{s + s^2 y'}{1 - y' - \zeta - 2sy'\zeta} = \frac{s^2(1+sy')^2}{1 + 2sy' - s^2 y'(1-y')}.$$

Note that when ζ runs along $\zeta = \rho$ anticlockwise, s will also run along a contour $\mathcal{C}$ anticlockwise. Therefore,

$$\begin{aligned}
&-\frac{1}{2\pi i y}\oint_{|\zeta|=\rho} \frac{1 - y\zeta^2 (ds_{y'}(\zeta)/d\zeta)}{z\zeta - 1 - y\zeta s_{y'}(\zeta)} d\zeta \\
&= -\frac{1}{2\pi i y}\oint_{\mathcal{C}} \frac{1 + 2sy' - s^2 y'(1-y') - y(s - sy' - 1)^2}{s(1+sy')[z(s - sy' - 1) - s(1+sy') - ys(s - sy' - 1)]} ds \\
&= -\frac{1}{2\pi i y}\oint_{\mathcal{C}} \frac{(y' + y - yy')(1-y')s^2 - 2s(y' + y - yy') - 1 + y}{(s + s^2 y')[(y' + y - yy')s^2 + s((1-y) - z(1-y')) + z]} ds.
\end{aligned}$$

The integrand has 4 poles at $s = 0, -1/y'$ and

$$\begin{aligned}
s_1, s_2 &= \frac{-(1-y) + z(1-y') \pm \sqrt{((1-y) + z(1-y'))^2 - 4z}}{2(y + y' - yy')} \\
&= \frac{2z}{-(1-y) + z(1-y') \mp \sqrt{((1-y) + z(1-y'))^2 - 4z}}
\end{aligned}$$

(the convention being that the first function takes the top operation).

We need to decide which pole is located inside the contour $\mathcal{C}$. From (4.4.8), it is easy to see that when ρ is small, for all $|\zeta| \le \rho$, $s_{y'}(\zeta)$ is close to $\frac{1}{1-y}$; that is, the contour $\mathcal{C}$ and its inner region are around $\frac{1}{1-y}$. Hence, 0 and $-1/y'$ are not inside the contour $\mathcal{C}$.

Let $z = u + iv$ with large u and $v > 0$. Then we have

$$\Im(((1-y) + z(1-y'))^2 - 4z) = 2v[(1-y)(u(1-y') + (1-y)) - 2] > 0.$$

By the convention for the square root of complex numbers, both real and imaginary parts of $\sqrt{((1-y) + z(1-y'))^2 - 4z}$ are positive. Therefore, $|s_1| > |s_2|$ and s_1 may take very large values. Also, s_2 will stay around $1/(1-y')$. We conclude that only s_2 is the pole inside the contour $\mathcal{C}$ for all z with large real part and positive imaginary part.

Now, let us compute the residue at s_2. By using $s_1 s_2 = z/(y + y' - yy')$, the residue is given by

$$\begin{aligned}
R &= \frac{(y' + y - yy')(1-y')s_2^2 - 2s_2(y' + y - yy') - 1 + y}{(s_2 + s_2^2 y')(y' + y - yy')(s_2 - s_1)} \\
&= \frac{(1-y')zs_2 s_1^{-1} - 2zs_1^{-1} - 1 + y}{(zs_1^{-1} + zs_2 s_1^{-1} y')(s_2 - s_1)}
\end{aligned}$$

$$= \frac{z(1-y')s_2 - 2z - (1-y)s_1}{z(1+s_2y')(s_2-s_1)}$$
$$= \frac{[(1-y+z-zy') - \sqrt{((1-y)+z(1-y'))^2-4z}](y+y'-yy')}{z(2y+y'-yy'+zy'(1-y')-y'\sqrt{((1-y)+z(1-y'))^2-4z})}.$$

Multiplying both the numerator and denominator by $2y+y'-yy'+zy'(1-y')+y'\sqrt{((1-y)+z(1-y'))^2-4z}$, after simplification we obtain

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

So, for all large $z \in \mathbb{C}^+$,

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

Since $s_F(z)$ is analytic on $\mathbb{C}^+$, the identity above is true for all $z \in \mathbb{C}^+$. Now, using Theorem B.10, letting $z \downarrow x + i0$, $\pi^{-1}\Im s_F(z)$ tends to the density function of the LSD of multivariate F-matrices; that is,

$$\begin{cases} \frac{\sqrt{4x-((1-y)+x(1-y'))^2}}{2\pi x(y+y'x)}, & \text{when } 4x-((1-y)+x(1-y'))^2 > 0, \\ 0, & \text{otherwise.} \end{cases}$$

This is equivalent to (4.4.1). Now we determine the possible atom at 0 by the fact that as $z = u+iv \to 0$ with $v > 0$, $zs_F(z) \to -F(\{0\})$. We have

$$\Im((1-y+z(1-y'))^2-4z) = 2v[(1-y+u(1-y'))(1-y')-2] < 0.$$

Hence $\Re(\sqrt{(1-y+z(1-y'))^2-4z}) < 0$. Thus $\sqrt{(1-y+z(1-y'))^2-4z} \to -|1-y|$. Consequently,

$$F(\{0\}) = -\lim_{z\to 0} zs_F(z) = 1 - \frac{1}{y} + \frac{1-y+|1-y|}{2y}$$
$$= \begin{cases} 1-\frac{1}{y}, & \text{if } y > 1 \\ 0, & \text{otherwise.} \end{cases}$$

This conclusion coincides with the intuitive observation that the matrix $\mathbf{S}_n\mathbf{T}_n$ has $p-n$ 0 eigenvalues.

This completes the proof of the theorem.

4.5 Proof of Theorem 4.3

In this section, we shall present a proof of Theorem 4.3 by using Stieltjes transforms. We shall prove it under a weaker condition that the entries of $\mathbf{X}_n$ satisfy (3.2.1). Steps in the proof follow along the same way as earlier proofs, with the additional step of verifying the uniqueness of solutions to (4.1.2). We first handle truncation and centralization.

4.5.1 Truncation and Centralization

Using similar arguments as in the proof of Theorem 4.1, we may assume $\mathbf{A}_n$ and $\mathbf{T}_n$ are nonrandom. Also, using the truncation approach given in the proof of Theorem 4.1, we may truncate the diagonal entries of the matrix $\mathbf{T}_n$ and thus we may assume additionally that $|\tau_k^{(n)}| \le \tau_0$.

Now, let us proceed to truncate and centralize the x-variables. Choose $\{\eta_n\}$ such that $\eta_n \to 0$ and

$$\frac{1}{n^2\eta_n^8}\sum_{ij} \mathrm{E}|X_{ij}^2|I(|x_{ij}| \ge \eta_n\sqrt{n}) \to 0. \tag{4.5.1}$$

Set $\widehat{x}_{ij} = x_{ij}I(|x_{ij}| < \eta_n\sqrt{n})$ and $\tilde{x}_{ij} = [\widehat{x}_{ij} - \mathrm{E}(\widehat{x}_{ij})]$, and define $\widehat{\mathbf{X}}_n$, $\widetilde{\mathbf{X}}_n$, $\widehat{\mathbf{B}}_n$, and $\widetilde{\mathbf{B}}_n$ as analogues of $\mathbf{X}_n$ and $\mathbf{B}_n$ by the corresponding $\widehat{x}_{ij}$ and $\tilde{x}_{ij}$, respectively. At first, by the second conclusion of Theorem A.44, we have

$$\begin{aligned}\|F^{\mathbf{B}_n} - F^{\widehat{\mathbf{B}}_n}\| &\le \frac{2}{p}\mathrm{rank}(\mathbf{X}_n - \widehat{\mathbf{X}}_n)\\ &\le \frac{2}{p}\sum_{ij} I(|x_{ij}| \ge \eta_n\sqrt{n}).\end{aligned}$$

Applying Bernstein's inequality, one may easily show that

$$\|F^{\mathbf{B}_n} - F^{\widehat{\mathbf{B}}_n}\| \to 0, \quad \text{a.s.}$$

Then, we will show that

$$L(F^{\widehat{\mathbf{B}}_n}, F^{\widetilde{\mathbf{B}}_n}) \to 0, \quad \text{a.s.} \tag{4.5.2}$$

By Theorem A.46, we have

$$\begin{aligned}L(F^{\widehat{\mathbf{B}}_n}, F^{\widetilde{\mathbf{B}}_n}) &\le \max_k |\lambda_k(\widehat{\mathbf{B}}_n) - \lambda_k(\widetilde{\mathbf{B}}_n)|\\ &\le \frac{1}{n}\|\widehat{\mathbf{X}}_n\mathbf{T}_n\widehat{\mathbf{X}}_n^* - \widetilde{\mathbf{X}}_n\mathbf{T}_n\widetilde{\mathbf{X}}_n^*\|\end{aligned}$$

$$\leq \frac{2}{n}\|(\mathrm{E}\widehat{\mathbf{X}}_n)\mathbf{T}_n\widetilde{\mathbf{X}}_n^*\| + \frac{1}{n}\|(\mathrm{E}\widehat{\mathbf{X}}_n)\mathbf{T}_n(\mathrm{E}\widehat{\mathbf{X}}_n^*)\|.$$

At first, we have

$$\begin{aligned}\frac{1}{n}\|(\mathrm{E}\widehat{\mathbf{X}}_n)\mathbf{T}_n(\mathrm{E}\widehat{\mathbf{X}}_n)^*\| &= \frac{1}{n}\|\mathrm{E}\widehat{\mathbf{X}}_n\|^2\|\mathbf{T}_n\| \\ &\leq \tau_0 n^{-1}\sum_{ij}|\mathrm{E}x_{ij}I(|x_{ij}| \leq \eta_n\sqrt{n})|^2 \\ &\leq \frac{\tau_0}{n^2\eta_n}\sum_{ij}\mathrm{E}|x_{ij}^2|I(|x_{ij}| \geq \eta_n\sqrt{n}) \to 0.\end{aligned}$$

Then, we shall complete the proof of (4.5.2) by showing that

$$\frac{1}{n}\|(\mathrm{E}\widehat{\mathbf{X}}_n)\mathbf{T}_n\widetilde{\mathbf{X}}_n^*\| \to 0, \text{ a.s.} \tag{4.5.3}$$

We have

$$\begin{aligned}\left(\frac{1}{n}\|(\mathrm{E}\widehat{\mathbf{X}}_n)\mathbf{T}_n\widetilde{\mathbf{X}}_n^*\|\right)^2 &\leq \frac{1}{n^2}\sum_{ik}\left|\sum_{j=1}^{p}(\mathrm{E}\widehat{x}_{ij})\tau_j\bar{\tilde{x}}_{kj}\right|^2 \\ &\leq J_1 + J_2 + J_3,\end{aligned}$$

where

$$\begin{aligned}J_1 &= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{p}\sum_{i=1}^{n}|\mathrm{E}\widehat{x}_{ij}\tau_j|^2|\tilde{x}_{kj}|^2, \\ J_2 &= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{j_1<j_2}^{p}\left(\sum_{i=1}^{n}\mathrm{E}\widehat{x}_{ij_1}\mathrm{E}\bar{\widehat{x}}_{ij_2}\tau_{j_1}\tau_{j_2}\right)\tilde{x}_{kj_1}\bar{\tilde{x}}_{kj_2}, \\ J_3 &= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{j_1>j_2}^{p}\left(\sum_{i=1}^{n}\mathrm{E}\widehat{x}_{ij_1}\mathrm{E}\bar{\widehat{x}}_{ij_2}\tau_{j_1}\tau_{j_2}\right)\tilde{x}_{kj_1}\bar{\tilde{x}}_{kj_2}.\end{aligned}$$

Using (4.5.1), we can prove

$$\begin{aligned}\mathrm{E}J_1 &= \frac{1}{n^2}\sum_{ik}\sum_{j=1}^{p}|\mathrm{E}\widehat{x}_{ij}\tau_j|^2\mathrm{E}|\tilde{x}_{kj}|^2 \\ &= \frac{\tau_0^2}{n^2\eta_n^2}\sum_{ij}|\mathrm{E}|x_{ij}|^2I(|x_{ij}| \geq \eta_n\sqrt{n}) \to 0\end{aligned}$$

and

$$\mathrm{E}|J_1 - \mathrm{E}J_1|^4 \le \frac{\tau_0^8}{n^8}\left[\sum_{kj}\mathrm{E}\left||\tilde{x}_{kj}^2|-1\right|^4\left|\sum_{i=1}^n|\mathrm{E}\widehat{x}_{ij}|^2\right|^4\right.$$
$$\left.+3\left(\sum_{kj}\mathrm{E}\left||\tilde{x}_{kj}^2|-1\right|^2\left|\sum_{i=1}^n|\mathrm{E}\widehat{x}_{ij}|^2\right|^2\right)^2\right]$$
$$= O(n^{-2}).$$

The preceding two formulas imply that $J_1 \to 0$, a.s.

Furthermore, we have

$$\mathrm{E}|J_2|^4 \le \frac{\tau_0^8}{n^8}\left[\sum_{k=1}^n\sum_{j_1<j_2}\mathrm{E}|\tilde{x}_{kj_1}^4|\mathrm{E}|\tilde{x}_{kj_2}^4|\left|\sum_{i=1}^n\mathrm{E}\widehat{x}_{ij_1}\mathrm{E}\bar{\widehat{x}}_{ij_2}\right|^4\right.$$
$$\left.+3\left(\sum_{k=1}^n\sum_{j_1<j_2}\mathrm{E}|\tilde{x}_{kj_1}^2|\mathrm{E}|\tilde{x}_{kj_2}^2|\left|\sum_{i=1}^n\mathrm{E}\widehat{x}_{ij_1}\mathrm{E}\bar{\widehat{x}}_{ij_2}\right|^2\right)^2\right]$$
$$= O(n^{-2})$$

and similarly

$$\mathrm{E}|J_3|^4 = O(n^{-2}).$$

These two imply that $J_2, J_3 \to 0$, a.s. Thus we have proved (4.5.3). Consequently, (4.5.2) follows.

Therefore, in what follows, we shall assume that:

(i) For each n, x_{ij} are independent.
(ii) $|x_{ij}| \le \eta_n\sqrt{n}$.
(iii) $\mathrm{E}x_{ij} = 0$.
(iv) $1 \ge \frac{1}{np}\sum_{ij}\mathrm{E}|x_{ij}|^2 \to 1$.

The details of the proofs are given in the next section.

4.5.2 Proof by the Stieltjes Transform

Let

$$s_n(z) = \frac{1}{n}\mathrm{tr}(\mathbf{B}_n - z\mathbf{I})^{-1}.$$

Assume first that F^A, the LSD of the sequence $\mathbf{A}_n$, is the zero measure. Then $s_A(z) \equiv 0$ for any $z \in \mathbb{C}^+$. In this case, the proof of (4.1.2) reduces to showing that $s_n(z) \to 0$. We have that, except for $o(n)$ eigenvalues of $\mathbf{A}_n$, all other eigenvalues of $\mathbf{A}_n$ will tend to $\pm$ infinity. Let $\lambda_k(\mathbf{A})$ denote the k-th largest singular value of matrix $\mathbf{A}$. By Theorem A.8, we have

$$\lambda_{i+j+1}(\mathbf{A}_n) \le \lambda_{i+1}(\mathbf{B}_n) + \lambda_{j+1}\left(-\frac{1}{n}\mathbf{X}_n\mathbf{T}_n\mathbf{X}_n^*\right). \tag{4.5.4}$$

Since $\|\mathbf{T}_n\| \le \tau_0$,

$$\lambda_{j+1}\left(-\frac{1}{n}\mathbf{X}_n\mathbf{T}_n\mathbf{X}_n^*\right) \le \tau_0\lambda_{j+1}\left(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n^*\right).$$

Let j_0 be an integer such that $\lambda_{j_0+1}(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n^*) < b+1 \le \lambda_{j_0}(\frac{1}{n}\mathbf{X}_n\mathbf{X}_n^*)$. Since the ESD of $\frac{1}{n}\mathbf{X}_n\mathbf{X}_n^*$ converges a.s. to the Marčenko-Pastur law with spectrum bounded in $[0,b]$, it follows that $j_0 = o(n)$. For any $M > 0$, define ν_0 to be such that $\lambda_{n-\nu_0+2}(\mathbf{A}_n) < M \le \lambda_{n-\nu_0+1}(\mathbf{A}_n)$. By the assumption that F^A is a zero measure, we conclude that $\nu_0 = o(n)$. Define $i_0 = n - \nu_0 - j_0$. By (4.5.4), we have

$$\lambda_{i_0+1}(\mathbf{B}_n) \ge \lambda_{n-\nu_0+1} - \lambda_{j_0+1}\left(-\frac{1}{n}\mathbf{X}_n\mathbf{T}_n\mathbf{X}_n^*\right) \ge M - b - 1.$$

This shows that except $o(n)$ eigenvalues of $\mathbf{B}_n$, the absolute values of all its other eigenvalues will be larger than $M - b - 1$. By the arbitrariness of M, we conclude that except for $o(n)$ eigenvalues of $\mathbf{B}_n$, all its other eigenvalues tend to infinity. This shows that $F^{\mathbf{B}_n}$ tends to a zero measure or $s_n(z) \to 0$, a.s., so that (4.1.2) holds trivially.

Now, we assume that $F^A \ne 0$. We claim that for any subsequence n', with probability one, $F^{\mathbf{B}_{n'}} \not\to 0$. Otherwise, using the same arguments given above, one may show that $F^{\mathbf{A}_{n'}} \to 0 = F^A$, a contradiction. This shows that, with probability 1, there is a constant m such that

$$\inf_n F^{\mathbf{B}_n}([-m,m]) > 0,$$

which simply implies that

$$\delta = \inf_n |\mathrm{E}s_n(z)| \ge \inf_n \mathrm{E}\Im(s_n(z)) \ge \mathrm{E}\inf_n \int \frac{v dF^{\mathbf{B}_n}(x)}{(x-u)^2 + v^2} > 0. \tag{4.5.5}$$

Now, we shall complete the proof of Theorem 4.3 by showing:

(a) $s_n(z) - \mathrm{E}s_n(z) \to 0$, a.s. (4.5.6)

(b) $\mathrm{E}s_n(z) \to s(z)$, (4.5.7)
which satisfies (4.1.2).

(c) The equation (4.1.2) has a unique solution in $\mathbb{C}^+$. (4.5.8)

Step 1. Proof of (4.5.6)

Let $\mathbf{x}_k$ denote the k-th column of $\mathbf{X}_n$, and set

$$\mathbf{q}_k = \frac{1}{\sqrt{n}}\mathbf{x}_k,$$
$$\mathbf{B}_{k,n} = \mathbf{B}_n - \tau_k \mathbf{q}_k \mathbf{q}_k^*.$$

Write E_k to denote the conditional expectation given $\mathbf{x}_{k+1}, \cdots, \mathbf{x}_p$. With this notation, we have $s_n(z) = \mathrm{E}_0(s_n(z))$ and $\mathrm{E}(s_n(z)) = \mathrm{E}_p(s_n(z))$. Therefore, we have

$$\begin{aligned} s_n(z) - \mathrm{E}(s_n(z)) &= \sum_{k=1}^{p}[\mathrm{E}_{k-1}(s_n(z)) - \mathrm{E}_k(s_n(z))] \\ &= \frac{1}{n}\sum_{k=1}^{p}[\mathrm{E}_{k-1} - \mathrm{E}_k](\mathrm{tr}(\mathbf{B}_n - z\mathbf{I})^{-1}) - \mathrm{tr}(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}) \\ &= \frac{1}{n}\sum_{k=1}^{p}[\mathrm{E}_{k-1} - \mathrm{E}_k]\gamma_k, \end{aligned}$$

where

$$\gamma_k = \frac{\tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-2}\mathbf{q}_k}{1 + \tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k}.$$

By (A.1.11), we have

$$|\gamma_k| \le \frac{|\tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-2}\mathbf{q}_k|}{|\Im(1 + \tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k)|} \le v^{-1}. \tag{4.5.9}$$

Note that $\{[\mathrm{E}_{k-1} - \mathrm{E}_k]\gamma_k\}$ forms a bounded martingale difference sequence. By applying Burkholder's inequality (see Lemma 2.12), one can easily show that, for any $\ell > 1$,

$$\begin{aligned} \mathrm{E}|s_n(z) - \mathrm{E}s_n(z)|^\ell &\le K_p n^{-\ell}\mathrm{E}\left(\sum_{k=1}^{n}|(\mathrm{E}_{k-1} - \mathrm{E}_k)\gamma_k|^2\right)^{\ell/2} \\ &\le K_\ell (2/v)^\ell n^{-\ell/2}. \end{aligned} \tag{4.5.10}$$

From this, with $p > 2$, it follows easily that

$$\frac{1}{n}\sum_{k=1}^{p}[\mathrm{E}_{k-1} - \mathrm{E}_k]\gamma_k \to 0, \text{ a.s.}$$

Then, what is to be shown follows.

Step 2. Proof of (4.5.7)

Let $s_{\mathbf{A}_n}(z)$ denote the Stieltjes transform of the ESD of $\mathbf{A}_n$. Write

$$x = x_n = \frac{1}{n}\sum_{k=1}^{p}\frac{\tau_k}{1 + \tau_k \mathrm{E}s_n(z)}.$$

It is easy to verify that $\Im x \le 0$. Write

$$\mathbf{B}_n - z\mathbf{I} = \mathbf{A}_n - (z-x)\mathbf{I} + \sum_{k=1}^{p} \tau_k \mathbf{q}_k \mathbf{q}_k^* - x\mathbf{I}.$$

Then, we have

$$\begin{aligned}&(\mathbf{A}_n - (z-x)\mathbf{I})^{-1} - (\mathbf{B}_n - z\mathbf{I})^{-1} \\ &= (\mathbf{A}_n - (z-x)\mathbf{I})^{-1}\left(\sum_{k=1}^{p} \tau_k \mathbf{q}_k \mathbf{q}_k^* - x\mathbf{I}\right)(\mathbf{B}_n - z\mathbf{I})^{-1}.\end{aligned}$$

From this and the definition of the Stieltjes transform of the ESD of random matrices, using the formula

$$\mathbf{q}_k^*(\mathbf{B}_n - z\mathbf{I})^{-1} = \frac{\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}}{1 + \tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k}, \tag{4.5.11}$$

we have

$$\begin{aligned} s_{\mathbf{A}_n}(z-x) - s_n(z) &= \frac{1}{n}\mathrm{tr}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}\left(\sum_{k=1}^{p} \tau_k \mathbf{q}_k \mathbf{q}_k^* - x\mathbf{I}\right)(\mathbf{B}_n - z\mathbf{I})^{-1} \\ &= \frac{1}{n}\mathrm{tr}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}\sum_{k=1}^{p} \tau_k \mathbf{q}_k \mathbf{q}_k^*(\mathbf{B}_n - z\mathbf{I})^{-1} \\ &\quad -\frac{x}{n}\mathrm{tr}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}(\mathbf{B}_n - z\mathbf{I})^{-1} \\ &= \frac{1}{n}\sum_{k=1}^{p} \frac{\tau_k d_k}{1 + \tau_k \mathrm{E} s_n(z)}, \end{aligned}$$

where

$$\begin{aligned} d_k = &\frac{1 + \tau_k \mathrm{E} s_n(z)}{1 + \tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k}\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}\mathbf{q}_k \\ &-\frac{1}{n}\mathrm{tr}(\mathbf{B} - z\mathbf{I})^{-1}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}. \end{aligned}$$

Write $d_k = d_{k1} + d_{k2} + d_{k3}$, where

$$\begin{aligned} d_{k1} &= \frac{1}{n}\mathrm{tr}(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1} - \frac{1}{n}\mathrm{tr}(\mathbf{B}_n - z\mathbf{I})^{-1}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}, \\ d_{k2} &= \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}\mathbf{q}_k - \frac{1}{n}\mathrm{tr}(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}, \\ d_{k3} &= \frac{\tau_k(\mathrm{E} s_n(z) - \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k)(\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}\mathbf{q}_k)}{1 + \tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k}. \end{aligned}$$

Noting that $\|(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}\| \le v^{-1}$, we have

$$|d_{k1}| = \frac{1}{n}\left| -\frac{\tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k}{1 + \tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k}\right|$$
$$\le n^{-1}v^{-1}\frac{|\tau_k|\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}(\mathbf{B}_{k,n} - \bar{z}\mathbf{I})^{-1}\mathbf{q}_k}{|\Im(1 + \tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k)|}$$
$$\le \frac{1}{nv^2}.$$

Therefore, by (4.5.5), we obtain

$$\frac{1}{n}\sum_{k=1}^{p}\frac{|\tau_k d_{k1}|}{|1 + \tau_k \mathrm{E}s_n(z)|} \le \frac{1}{nv^2\delta} \to 0.$$

Obviously, $\mathrm{E}d_{k2} = 0$.

To estimate $\mathrm{E}d_{k3}$, we first show that

$$\left|\frac{\tau_k(\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}\mathbf{q}_k)}{1 + \tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k}\right| \le 2\tau_0 v^{-2}\|\mathbf{q}_k\|^2. \tag{4.5.12}$$

One can consider $\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k/\|\mathbf{q}_k\|^2$ as the Stieltjes transform of a distribution. Thus, by Theorem B.11, we have

$$|\Re(\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k)| \le v^{-1/2}\|\mathbf{q}_k\|\sqrt{\Im(\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k)}.$$

Thus, if

$$\tau_0 v^{-1/2}\|\mathbf{q}_k\|\sqrt{\Im(\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k)} \le 1/2,$$

then

$$|1 + \tau_k(\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k)| \ge 1 - \tau_0|\Re(\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k)| \ge 1/2.$$

Hence,

$$\left|\frac{\tau_k(\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}\mathbf{q}_k)}{1 + \tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k}\right| \le 2\tau_0 v^{-2}\|\mathbf{q}_k\|^2.$$

Otherwise, we have

$$\left|\frac{\tau_k(\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}\mathbf{q}_k)}{1 + \tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k}\right|$$
$$\le \frac{|\tau_k\|(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k\|\|(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}\mathbf{q}_k\||}{|\Im(1 + \tau_k \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k)|}$$
$$= \frac{\|(\mathbf{A}_n - (z-x)\mathbf{I})^{-1}\mathbf{q}_k\|}{\sqrt{v\Im(\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k)}}$$

$$\leq 2\tau_0 v^{-2}\|\mathbf{q}_k\|^2.$$

Therefore, for some constant C,

$$|\mathrm{E}d_{k3}|^2 \leq C\mathrm{E}|\mathrm{E}s_n(z) - \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k|^2\mathrm{E}\|\mathbf{q}_k\|^4. \tag{4.5.13}$$

At first, we have

$$\begin{aligned}
\mathrm{E}\|\mathbf{q}_k\|^4 &= \frac{1}{n^2}\mathrm{E}\left(\sum_{i=1}^n |x_{ik}|^2\right)^2 \\
&= \frac{1}{n^2}\left[\sum_{i=1}^n \mathrm{E}|x_{ik}|^4 + \sum_{i\neq j}\mathrm{E}|x_{ik}|^2\mathrm{E}|x_{jk}|^2\right] \\
&\leq \frac{1}{n^2}[n^2\eta_n^2 + n(n-1)] \leq 1 + \eta_n^2.
\end{aligned}$$

To complete the proof of the convergence of $\mathrm{E}s_n(z)$, we need to show that

$$\frac{1}{n}\sum_{k=1}^p (\mathrm{E}|\mathrm{E}s_n(z) - \mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k|^2)^{1/2} \to 0. \tag{4.5.14}$$

Write $(\mathbf{B}_{k,n} - z\mathbf{I})^{-1} = (b_{ij})$. Then, we have

$$\begin{aligned}
&\mathrm{E}\left|\mathbf{q}_k^*(\mathbf{B}_{k,n} - z\mathbf{I})^{-1}\mathbf{q}_k - \frac{1}{n}\sum_{i=1}^n \sigma_{ik}^2 b_{ii}\right|^2 \\
&\leq \frac{1}{n^2}\left[\sum_{i=1}^n \mathrm{E}|x_{ik}^2 - \sigma_{ik}^2|^2 + 2\sum_{i\neq j}\mathrm{E}|x_{ik}^2|\mathrm{E}|x_{jk}^2||b_{ij}|^2\right] \\
&\leq v^{-2}\eta_n^2 + \frac{2}{n^2}\mathrm{tr}((\mathbf{B}_{k,n} - u\mathbf{I})^2 + v^2\mathbf{I})^{-1} \\
&\leq v^{-2}[\eta_n^2 + n^{-1}] \to 0.
\end{aligned}$$

By noting that $1 - \sigma_{ik}^2 \geq 0$,

$$\left|\frac{1}{n}\sum_{i=1}^n (\sigma_{ik}^2 - 1)b_{ii}\right| \leq \frac{1}{nv}\sum_{i=1}^n (1 - \sigma_{ik}^2),$$

$$\left|\frac{1}{n}\mathrm{tr}(\mathbf{B}_{k,n} - z\mathbf{I})^{-1} - s_n(z)\right| \leq 1/nv.$$

By Step 1, we have

$$\mathrm{E}|s_n(z) - \mathrm{E}(s_n(z))|^2 \leq \frac{1}{nv^2}.$$

Then, (4.5.14) follows from the estimates above.

Up to the present, we have proved that, for any $z \in \mathbb{C}^+$,

$$s_{\mathbf{A}_n}(z-x) - \mathrm{E}s_n(z) \to 0.$$

For any subsequence n' such that $\mathrm{E}s_n(z)$ tends to a limit, say s, by assumption of the theorem, we have

$$x = x_{n'} = \frac{1}{n}\sum_{k=1}^{p} \frac{\tau_k}{1+\tau_k \mathrm{E}s_{n'}(z)} \to y\int \frac{\tau dH(\tau)}{1+\tau s}.$$

Therefore, s will satisfy (4.1.2). We have thus proved (4.5.7) if equation (4.1.2) has a unique solution $s \in \mathbb{C}^+$, which is done in the next step.

Step 3. Uniqueness of the solution of (4.1.2)

If F^A is a zero measure, the unique solution is obviously $s(z) = 0$. Now, suppose that $F^A \neq 0$ and we have two solutions $s_1, s_2 \in \mathbb{C}^+$ of equation (4.1.2) for a common $z \in \mathbb{C}^+$; that is,

$$s_j = \int \frac{dF^A(\lambda)}{\lambda - z + y\int \frac{\tau dH(\tau)}{1+\tau s_j}}, \tag{4.5.15}$$

from which we obtain

$$\begin{aligned} &s_1 - s_2 \\ &= y\int \frac{(s_1-s_2)\lambda^2 dH(\tau)}{(1+\tau s_1)(1+\tau s_2)} \int \frac{dF^A(\lambda)}{\left(\lambda - z + y\int \frac{\tau dH(\tau)}{1+\tau s_1}\right)\left(\lambda - z + y\int \frac{\tau dH(\tau)}{1+\tau s_2}\right)}. \end{aligned}$$

If $s_1 \neq s_2$, then

$$\int \frac{y\int \frac{\lambda^2 dH(\tau)}{(1+\tau s_1)(1+\tau s_2)} dF^A(\lambda)}{\left(\lambda - z + y\int \frac{\tau dH(\tau)}{1+\tau s_1}\right)\left(\lambda - z + y\int \frac{\tau dH(\tau)}{1+\tau s_2}\right)} = 1.$$

By the Cauchy-Schwarz inequality, we have

$$1 \le \left(\int \frac{y\int \frac{\lambda^2 dH(\tau)}{|1+\tau s_1|^2} dF^A(\lambda)}{\left|\lambda - z + y\int \frac{\tau dH(\tau)}{1+\tau s_1}\right|^2} \int \frac{y\int \frac{\lambda^2 dH(\tau)}{|1+\tau s_2|^2} dF^A(\lambda)}{\left|\lambda - z + y\int \frac{\tau dH(\tau)}{1+\tau s_2}\right|^2} \right)^{1/2}.$$

From (4.5.15), we have

$$\Im s_j = \int \frac{v + y\Im s_j \int \frac{\tau^2 dH(\tau)}{|1+\tau s_j|^2} dF^A(\lambda)}{\left|\lambda - z + y\int \frac{\tau dH(\tau)}{1+\tau s_j}\right|^2} > \int \frac{y\Im s_j \int \frac{\tau^2 dH(\tau)}{|1+\tau s_j|^2} dF^A(\lambda)}{\left|\lambda - z + y\int \frac{\tau dH(\tau)}{1+\tau s_j}\right|^2},$$

which implies that, for both $j = 1$ and $j = 2$,

$$1 > \int \frac{y \int \frac{\tau^2 dH(\tau)}{|1+\tau s_j|^2} dF^A(\lambda)}{|\lambda - z + y \int \frac{\tau dH(\tau)}{1+\tau s_j}|^2}.$$

The inequality is strict even if F^A is a zero measure, which leads to a contradiction. The contradiction proves that $s_1 = s_2$ and hence equation (4.1.2) has at most one solution. The existence of the solution to (4.1.2) has been seen in Step 2. The proof of this theorem is then complete.

Chapter 5
Limits of Extreme Eigenvalues

In multivariate analysis, many statistics involved with a random matrix can be written as functions of integrals with respect to the ESD of the random matrix. When the LSD is known, one may want to apply the Helly-Bray theorem to find approximate values of the statistics. However, the integrands are usually unbounded. For instance, the integrand in Example 1.2 is $\log x$, which is unbounded both from below and above. Thus, one cannot use the LSD and Helly-Bray theorem to find approximate values of the statistics. This would render the LSD useless. Fortunately, in most cases, the supports of the LSDs are compact intervals. Still, this does not mean that the Helly-Bray theorem is applicable unless one can prove that the extreme eigenvalues of the random matrix remain in certain bounded intervals.

The investigation on limits of extreme eigenvalues is important not only in making the LSD useful when applying the Helly Bray theorem, but also for its own practical interests. In signal processing, pattern recognition, edge detection, and many other areas, the support of the LSD of the population covariance matrices consists of several disjoint pieces. It is important to know whether or not the LSD of the sample covariance matrices is also separated into the same number of disjoint pieces, under what conditions this is true, and whether or not there are eigenvalues falling into the spacings outside the support of the LSD of the sample covariance matrices.

The first work in this direction is due to Geman [118]. He proved that the largest eigenvalue of a sample covariance matrix tends to b $(= \sigma^2(1 + \sqrt{y})^2)$ when $p/n \to y \in (0, \infty)$ under a restriction on the growth rate of the moments of the underlying distribution. This work was generalized by Yin, Bai, and Krishnaiah [301] under the assumption of the existence of the fourth moment of the underlying distribution. In Bai, Silverstein, and Yin [33], it is further proved that if the fourth moment of the underlying distribution is infinite, then, with probability 1, the limsup of the largest eigenvalue of a sample covariance matrix is infinity. Combining the two results, we have in fact established the necessary and sufficient conditions for the existence of the limit of the largest eigenvalue of a large dimensional sample covariance

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_5,

matrix. In Bai and Yin [38], the necessary and sufficient conditions for the a.s. convergence of the extreme eigenvalues of a large Wigner matrix were found. The most difficult problem in this direction concerns the limit of the smallest eigenvalue of a large sample covariance matrix. In Yin, Bai, and Krishnaiah [302], it is proved that the lower limit of the smallest eigenvalue of a Wishart matrix has a positive lower bound if $p/n \to y \in (0, 1/2)$. Silverstein [262] extended this work to allow $y \in (0, 1)$. Further, Silverstein [261] showed that the smallest eigenvalue of a standard Wishart matrix almost surely tends to a $(= (1 - \sqrt{y})^2)$ if $p/n \to y \in (0, 1)$. The most current result is due to Bai and Yin [36], in which it is proved that the smallest (nonzero) eigenvalue of a large dimensional sample covariance matrix tends to $a = \sigma^2(1 - \sqrt{y})^2$ when $p/n \to y \in (0, \infty)$ under the existence of the fourth moment of the underlying distribution.

In Bai and Silverstein [32], it is shown that in any closed interval outside the support of the LSD of a sequence of large dimensional sample covariance matrices (when the population covariance matrix is not a multiple of the identity), with probability 1, there are no eigenvalues for all large n. This work will be introduced in Chapter 6. In this chapter, we introduce some results in this direction by using the moment approach.

5.1 Limit of Extreme Eigenvalues of the Wigner Matrix

The following theorem is a generalization of Bai and Yin [38], where the real case is considered. What we state here is for the complex case because we were questioned by researchers in electrical and electronic engineering on many occasions as to whether the result is true with complex entries.

Theorem 5.1. *Suppose that the diagonal elements of the Wigner matrix* $\sqrt{n}\mathbf{W}_n = (\sqrt{n} w_{ij}) = (x_{ij})$ *are iid real random variables, the elements above the diagonal are iid complex random variables and all these variables are independent. Then, the largest eigenvalue of* $\mathbf{W}$ *tends to* $c > 0$ *with probability 1 if and only if the five conditions*

$$\begin{array}{ll} \text{(i)} & \mathrm{E}((x_{11}^+)^2) < \infty, \\ \text{(ii)} & \mathrm{E}(x_{12}) \ \textit{is real and} \ \leq 0, \\ \text{(iii)} & \mathrm{E}(|x_{12} - \mathrm{E}(x_{12})|^2) = \sigma^2, \\ \text{(iv)} & \mathrm{E}(|x_{12}^4|) < \infty, \\ \text{(v)} & c = 2\sigma, \end{array} \tag{5.1.1}$$

where $x^+ = \max(x, 0)$*, are true.*

By the symmetry of the largest and smallest eigenvalues of a Wigner matrix, one can easily derive the necessary and sufficient conditions for the existence of the limit of smallest eigenvalues of a Wigner matrix. Combining these conditions, we obtain the following theorem.

Theorem 5.2. *Suppose that the diagonal elements of the Wigner matrix* $\mathbf{W}_n$ *are iid real random variables, the elements above the diagonal are iid complex random variables, and all these variables are independent. Then, the largest eigenvalue of* $\mathbf{W}$ *tends to* c_1 *and the smallest eigenvalue tends to* c_2 *with probability 1 if and only if the following five conditions are true:*

$$\begin{array}{ll} \text{(i)} & \mathrm{E}(x_{11}^2) < \infty, \\ \text{(ii)} & \mathrm{E}(x_{12}) = 0, \\ \text{(iii)} & \mathrm{E}(|x_{12}|^2) = \sigma^2, \\ \text{(iv)} & \mathrm{E}(|x_{12}^4|) < \infty, \\ \text{(v)} & c_1 = 2\sigma \ \ \textit{and} \ \ c_2 = -2\sigma. \end{array} \tag{5.1.2}$$

From the proof of Theorem 5.1, it is easy to see the following weak convergence theorem on the extreme eigenvalue of a large Wigner matrix.

Theorem 5.3. *Suppose that the diagonal elements of the Wigner matrix* $\sqrt{n}\mathbf{W}_n = (x_{ij})$ *are iid real random variables, the elements above the diagonal are iid complex random variables, and all these variables are independent. Then, the largest eigenvalue of* $\mathbf{W}$ *tends to* $c > 0$ *in probability if and only if the following five conditions are true:*

$$\begin{array}{ll} \text{(i)} & \mathrm{P}(x_{11}^+ > \sqrt{n}) = o(n^{-1}), \\ \text{(ii)} & \mathrm{E}(x_{12}) \ \ \textit{is real and} \ \ \leq 0, \\ \text{(iii)} & \mathrm{E}(|x_{12} - \mathrm{E}(x_{12})|^2) = \sigma^2, \\ \text{(iv)} & \mathrm{P}(|x_{12}| > \sqrt{n}) = o(n^{-2}), \\ \text{(v)} & c = 2\sigma. \end{array} \tag{5.1.3}$$

5.1.1 Sufficiency of Conditions of Theorem 5.1

It is obvious that we can assume $\sigma = 1$ without loss of generality.

The conditions of Theorem 5.1 imply that the assumptions of Theorem 2.5 are satisfied. By the latter, we have

$$\liminf_{n\to\infty} \lambda_n(\mathbf{W}) \geq 2, \text{ a.s.} \tag{5.1.4}$$

Thus, the proof of the sufficiency reduces to showing that

$$\limsup_{n\to\infty} \lambda_n(\mathbf{W}) \leq 2, \text{ a.s.} \tag{5.1.5}$$

The key in proving (5.1.5) is the bound given in (5.1.9) below for an appropriate sequence of k_n's. A combinatorial argument is required. Before this bound is used, the assumptions on the entries of $\mathbf{W}$ are simplified.

Condition (i) implies that $\limsup \frac{1}{\sqrt{n}} \max_{k\leq n} x_{kk}^+ = 0$, a.s. By condition (ii) and the relation

$$
\begin{aligned}
\lambda_{\max}(\mathbf{W}) &= \frac{1}{\sqrt{n}} \max_{\|\mathbf{z}\|=1} \left(\sum_{j,k} z_j \bar{z}_k x_{jk} \right) \\
&= \max_{\|\mathbf{z}\|=1} \left[\frac{1}{\sqrt{n}} \sum_{j \neq k} z_j \bar{z}_k (x_{jk} - \mathrm{E}(x_{jk})) + \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |z_k|^2 x_{kk} \right. \\
&\qquad \left. + \Re(\mathrm{E}(x_{12})) \frac{1}{\sqrt{n}} \sum_{j \neq k} z_j \bar{z}_k \right] \\
\leq &\quad \max_{\|\mathbf{z}\|=1} \left(\frac{1}{\sqrt{n}} \sum_{j \neq k} z_j \bar{z}_k (x_{jk} - \mathrm{E}(x_{jk})) + \frac{1}{\sqrt{n}} \max_k (x_{kk}^+ - \Re(\mathrm{E}(x_{12}))) \right) \\
\leq &\quad \lambda_{\max}(\widetilde{\mathbf{W}}) + o_{\text{a.s.}}(1), \qquad (5.1.6)
\end{aligned}
$$

where $\widetilde{\mathbf{W}}_n$ denotes the matrix whose diagonal elements are zero and whose off-diagonal elements are $\frac{1}{\sqrt{n}}(x_{ij} - \mathrm{E}(x_{ij}))$. By (5.1.4)–(5.1.6), we only need show that

$$\limsup_{n \to \infty} \lambda_{\max}(\widetilde{\mathbf{W}}) \leq 2, \text{a.s.}$$

That means we may assume that the diagonal elements and the mean of the off-diagonal elements of $\mathbf{W}_n$ are zero in the proof of (5.1.5).

We first truncate the off-diagonal elements. By condition (iv), for any $\delta > 0$, we have

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

Then, we can select a slowly decreasing sequence of constants $\delta_n \to 0$ such that

$$\sum_{k=1}^{\infty} \delta_{2^k}^{-2} 2^k \mathrm{E}|x_{12}|^2 I(|x_{12}| \geq \delta_{2^k} 2^{k/2}) < \infty. \qquad (5.1.7)$$

Let $\widetilde{\mathbf{W}} = \frac{1}{\sqrt{n}}(x_{jk} I(|x_{jk}| \leq \delta_n \sqrt{n}))$. Then, by (5.1.7), we have

$$
\begin{aligned}
\mathrm{P}(\mathbf{W} \neq \widetilde{\mathbf{W}}, \text{ i.o.}) &= \lim_{k \to \infty} \mathrm{P} \left(\bigcup_{n=2^k} \bigcup_{1 \leq i < j \leq n} (|x_{jk}| \geq \delta_n \sqrt{n}) \right) \\
&\leq \lim_{k \to \infty} \sum_{m=k}^{\infty} \mathrm{P} \left(\bigcup_{2^m < n \leq 2^{m+1}} \bigcup_{1 \leq i < j \leq n} (|x_{jk}| \geq \delta_n \sqrt{n}) \right) \\
&\leq \lim_{k \to \infty} \sum_{m=k}^{\infty} \mathrm{P} \left(\bigcup_{2^m < n \leq 2^{m+1}} \bigcup_{1 \leq i < j \leq 2^{m+1}} (|x_{jk}| \geq \delta_{2^m} 2^{m/2}) \right)
\end{aligned}
$$

$$\leq \lim_{k\to\infty} \sum_{m=k}^{\infty} 2^{2(m+1)} \mathrm{P}\left(|x_{12}| \geq \delta_{2^m} 2^{m/2}\right) = 0.$$

By the selection of δ_n, we have

$$\begin{aligned}\lambda_{\max}(\mathrm{E}(\widetilde{\mathbf{W}})) &\leq n^{1/2}|\mathrm{E}(x_{12}I(|x_{12}| \leq \delta_n\sqrt{n}))| \\ &\leq n^{1/2}\mathrm{E}(|x_{12}|I(|x_{12}| \geq \delta_n\sqrt{n})) \to 0. \end{aligned} \tag{5.1.8}$$

Therefore, we need only consider the upper limit of the largest eigenvalue of $\widetilde{\mathbf{W}} - \mathrm{E}(\widetilde{\mathbf{W}})$. For simplicity, we still use $\mathbf{W}$ to denote the truncated and recentralized matrix. That means we assume that $\sqrt{n}\mathbf{W}_n = (x_{ij})$ and the following conditions are true:

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

We shall prove (5.1.5) under the four assumptions above and the independence of the entries. For any even integer k and real number $\eta > 2$, we have

$$\mathrm{P}(\lambda_{\max}(\mathbf{W}_n) \geq \eta) \leq \mathrm{P}(\mathrm{tr}[(\mathbf{W}_n)^k] \geq \eta^k) \leq \eta^{-k}\mathrm{E}(\mathrm{tr}(\mathbf{W}_n)^k). \tag{5.1.9}$$

To complete the proof of the sufficient part of the theorem, select a sequence of even integers $k = k_n = 2s$ with the properties $k/\log n \to \infty$ and $k\delta_n^{1/3}/\log n \to 0$, and show that the right-hand side of (5.1.9) is summable. To this end, we shall estimate

$$\begin{aligned}\mathrm{E}(\mathrm{tr}(\mathbf{W}^k)) &= n^{-k/2} \sum_{i_1,\cdots,i_k} \mathrm{E}(x_{i_1i_2}x_{i_2i_3}\cdots x_{i_ki_1}) \\ &= n^{-k/2} \sum_G \sum_{\mathbf{i}} \mathrm{E}(x_G(\mathbf{i})), \end{aligned} \tag{5.1.10}$$

where the graphs G are $\Gamma(k)$-graphs as defined in Subsection 2.1.2. Classify the edges into several types:

1. If $f(a+1) = \max(f(1), \cdots, f(a)) + 1$, the edge $e_a = (f(a), f(a+1))$ is called an innovation or a Type 1 (T_1) edge. A T_1 edge leads to a new vertex in the path $e_1, \cdots, e_a$.
2. An edge is called a T_3 if it coincides with an innovation that is single until the T_3 edge appears. A T_3 edge $(f(a), f(a+1))$ is said to be **irregular** if there is only one innovation single up to a (an edge e is said to be single up to a if it does not coincide with any other edges in the chain $(f(1), \cdots, f(a))$). All other T_3 edges are called **regular** T_3 edges.
3. All other edges are called T_4 edges.

4. The first appearance of a T_4 edge is called a T_2 edge. There are two cases: the first is the first appearance of a single noninnovation, and the second is the first appearance of an edge that coincides with a T_3 edge.

To estimate the right-hand side of (5.1.10), we need the following lemmas. We remind the reader that the following lemmas are true for any connected graphs, not only for the Γ-graphs.

Lemma 5.4. *Let $(f(a), \cdots, f(c))$ be a chain such that the edge $(f(a), f(a+1))$ is an innovation single up to c (i.e., $f(a+1) \notin \{f(1), \cdots, f(a), f(a+2), \cdots, f(c)\}$) and $f(c) \in \{f(1), \cdots, f(a)\}$. Then there is a T_2 edge contained in the chain $(f(a), \cdots, f(c))$.*

Proof. Since $(f(a), f(a+1))$ is an innovation and $f(a+1) \notin \{f(1), \cdots, f(a)\}$, let $a < d < c$ be the smallest value such that $f(d) \notin \{f(1), \cdots, f(a)\}$ but $f(d+1) \in \{f(1), \cdots, f(a)\}$. Since $f(a+1) \notin \{f(1), \cdots, f(a)\}$ and $f(c) \in \{f(1), \cdots, f(a)\}$, the value d is well defined. Then, the edge $(f(d), f(d+1))$ must be a T_2 edge (see Fig. 5.1). The proof of lemma is complete.

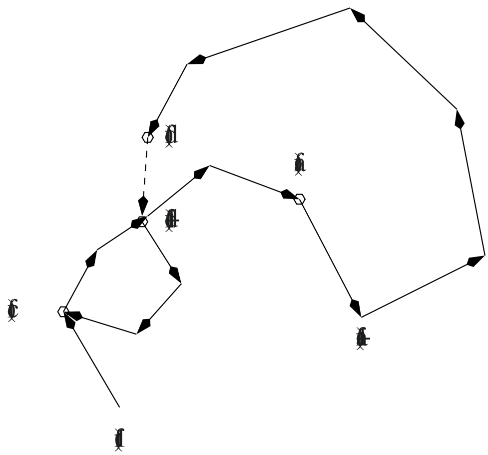

Fig. 5.1 The broken arrow is a T_2 edge.

Lemma 5.5. *Let t denote the number of T_2 edges and s denote the number of innovations in the chain $(f(1), f(2), \cdots, f(a))$ that are single up to a and have a vertex coincident with $f(a)$. Then $s \leq t+1$.*

Proof. If $s = 1$, then the lemma is trivially true. Thus, we need only consider the case $s > 1$. Except for the first innovation, which leads to the vertex $f(a)$, all other innovations start from the vertex $f(a)$. Write the remaining $s-1$ innovations single up to a by $(f(b_1), f(b_1+1)), \cdots, (f(b_{s-1}), f(b_{s-1}+1))$ with $f(b_1) = \cdots = f(b_{s-1}) = f(a)$ and $b_1 < b_2 < \cdots < b_{s-1} < a = b_s$.

For any $u \leq s-1$, we consider the cycle $(f(b_u), \cdots, f(b_{u+1}))$. Since $(f(b_u), f(b_u+1))$ is a single innovation up to $e_{b_{u+1}}$, by Lemma 5.4, there is a T_2 edge $(f(c_u), f(c_u+1))$ with $f(c_u) \notin \{f(1), \cdots, f(b_u)\}$. This property guarantees that the T_2 edges in the $s-1$ cycles are distinct (see Fig. 5.2). This proves the lemma.

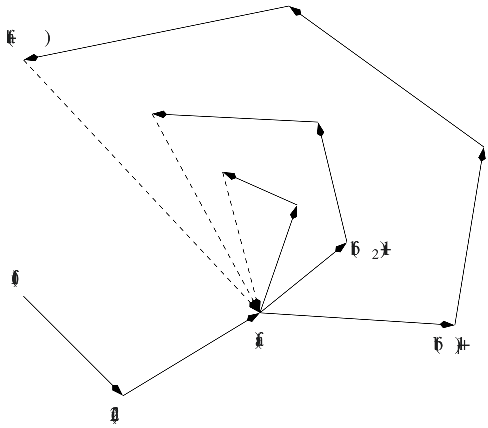

Fig. 5.2 Each cycle contains a T_2 edge (broken arrow).

Lemma 5.6. *The number of regular T_3 edges is not greater than twice the number of T_2 edges.*

To prove this lemma, we need the following concepts.

Definition. A chain $(f(a), \cdots, f(b))$ is called a $*$-cycle up to c $(\geq b)$ if:

1. $f(a) = f(b) \notin \{f(a+1), \cdots, f(b-1)\}$. The vertex $f(a) = f(b)$ is called the **head** of the $*$-cycle.
2. The edge $(f(a), f(a+1))$ is an innovation single up to c. This innovation is called the **leader** of the $*$-cycle.
3. In the chain $(f(1), \cdots, f(a))$, there is at least one innovation single up to c and having a vertex coincident with $f(a)$. The latest such innovation is called the **preleader** of the $*$-cycle.

By definition, the leader of a $*$-cycle can be the preleader of the next $*$-cycle with the same head.

Definition. A $*$-cycle is said to be of the **first type** if there are no T_2 edges from its

preleader to its leader (see Fig. 5.3). Otherwise, it is said to be of the **second type** (see Fig. 5.4).

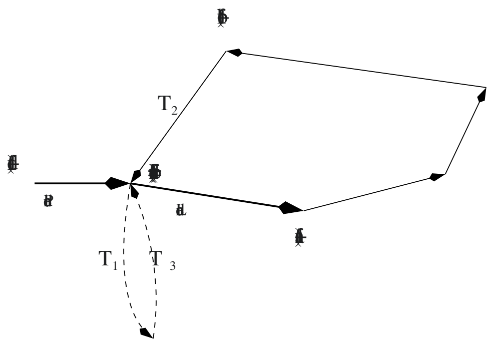

Fig. 5.3 Definition of a $*$-cycle of the first type.

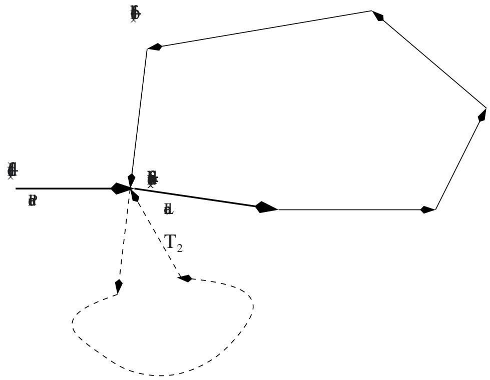

Fig. 5.4 Definition of a $*$-cycle of the second type.

If $(f(a), \cdots, f(b))$ is a first type $*$-cycle and its preleader is $(f(d-1), f(d))$, then we must have $f(d) = f(a)$. Otherwise, $f(d-1) = f(a) \in \{f(1), \cdots, f(d-1)\}$ and $(f(d-1), f(d))$ is an innovation single up to a, which, together with Lemma 5.4, implies that there must be a T_2 edge between its preleader and leader, which contradicts the definition of the first-type $*$-cycle. Furthermore, if $d < a$, then between $f(d)$ and $f(a)$ all edges are either T_1 or T_3 and no T_1 edges are single up to a, for otherwise there would be a T_2 by Lemma 5.4. Thus, $a - d$ is even and $(a - d)/2$ edges between $f(d)$ and $f(a)$ are T_1, which are coincident with the other $(a - d)/2$ T_3 edges. Therefore, the preleader $(f(d-1), f(d))$ is the only edge that links the chains $(f(d), f(d+1), \cdots, f(a))$ and $(f(1), \cdots, f(d-1))$. Moreover, we claim that $(f(b-1), f(b))$ must be a T_2. At first, $f(b-1) \neq f(d-1)$, for otherwise $(f(d-1), f(d))$ cannot be single up to b, which violates the definition of preleader. Also, $f(a) = f(b) \notin \{f(d-1), f(a+1), \cdots, f(b-1)\}$, so the edge $(f(b-1), f(b))$ is either single or coincident with a T_3 edge between $f(d)$ and $f(a)$. In either case, it is a T_2.

Proof of Lemma 5.6. Suppose that $(f(a), f(a+1))$ is a regular T_3 edge. Then, in the chain $(f(1), \cdots, f(a))$, there is at least one $*$-cycle up to a with head $f(a)$, whereas in the chain $(f(1), \cdots, f(a+1))$, one such cycle disappears because one single innovation has been coincident with $(f(a), f(a+1))$. Therefore, the number of regular T_3 edges does not exceed the number of $*$-cycles. Therefore, to complete the proof of this lemma, it is sufficient to show that the number of $*$-cycles does not exceed twice the number of T_2 edges.

Define a mapping Φ^* from the $*$-cycles to the T_2 edges in such a way that the Φ^*-image of a first-type $*$-cycle is the last T_2 edge in the cycle and the Φ^*-image of a second-type $*$-cycle is the first T_2 edge in the cycle. By Lemma 5.4, there is at least one T_2 edge in each $*$-cycle. Hence, the mapping Φ^* is well defined.

Then, the proof of this lemma amounts to showing that any three $*$-cycles cannot have a common Φ^*-image. If there are three $*$-cycles that have a common Φ^*-image, then among the three $*$-cycles either two of them are of the first type or two of them are of the second type. We will derive a contradiction for both cases.

Suppose that the $*$-cycles are $(f(a_1), \cdots, f(b_1))$ and $(f(a_2), \cdots, f(b_2))$ with $a_1 < a_2$. It is evident that the two $*$-cycles have different heads since otherwise the two cycles do not have common edges, which contradicts the assumption that they have a common Φ^*-image.

Case 1. Suppose that the two $*$-cycles are both of the first-type. As discussed earlier for first-type $*$-cycles, the last edges of the cycles are the Φ^* image of the two $*$-cycles, namely $(f(b_1 - 1), f(b_1))$ and $(f(b_2 - 1), f(b_2))$. Because the two $*$-cycles have different heads, $b_1 \neq b_2$. Therefore, the later of the two edges does not belong to both $*$-cycles and hence they cannot be the same edges. Therefore, Case 1 is impossible.

Case 2. Suppose that the two $*$-cycles are both of the second type. If the preleader $(f(d_2 - 1), f(d_2))$ appears after the head $f(a_1)$ of the first $*$-cycle, then the ϕ^* image of the first $*$-cycle is in the chain $(f(a_1 + 1), \cdots, f(a_2))$ because there is a T_2 edge in the chain $(f(d_2), \cdots, f(a_2))$. This edge is not in the second $*$-cycle (see Fig. 5.5).

If the preleader $(f(d_2 - 1), f(d_2))$ appears before the head $f(a_1)$ of the first $*$-cycle (that is, $d_2 < a_1$), then, by Lemma 5.4, there is a T_2 edge in the chain $(f(a_1 + 1), C, f(a_2))$ and hence the Φ^* image of the first $*$-cycle is also in this chain. This shows that this case is also impossible (see Fig. 5.6). The proof of the lemma is complete.

Continuing the proof of Theorem 5.1. Now, we begin to estimate the right-hand side of (5.1.10). At first, if the graph G has a single edge, then the corresponding terms are zero. Therefore, we need only estimate those terms corresponding to Γ_1- and Γ_3-graphs. Suppose that there are r innovations ($r \leq s$) and t T_2 edges in the graph G. Then there are r T_3 edges, $k - 2r$ T_4 edges and $r + 1$ noncoincident vertices.

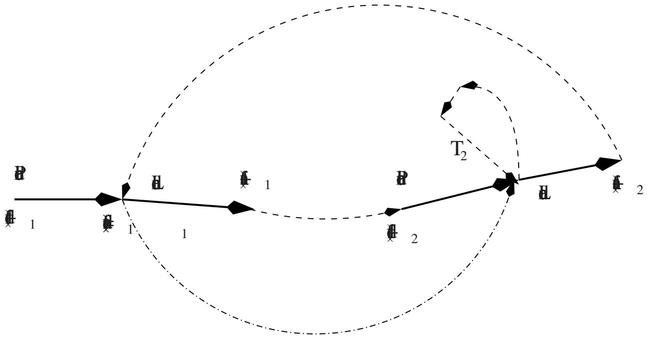

Fig. 5.5 The preleader of the second ∗-cycle is later than the leader of the first ∗-cycle.

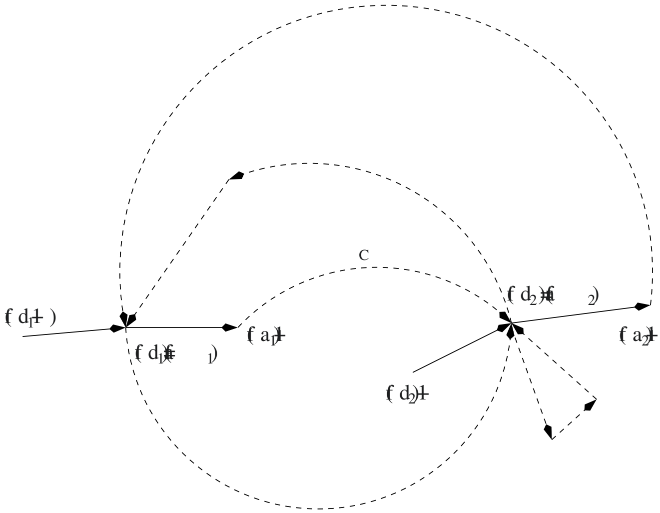

Fig. 5.6 The preleader of the second ∗-cycle is earlier than the leader of the first ∗-cycle.

Therefore, the number of graphs of each isomorphic class is less than n^{r+1}, and the expectation corresponding to each canonical graph is not larger than $b^t(\delta_n\sqrt{n})^{k-2r-t}$. Then we need to estimate the number of canonical graphs. At first, there are at most $\binom{k}{r}$ ways to select r edges out of the total k edges to be the r innovations. Then, there are at most $\binom{k-r}{r}$ ways to select r edges out of the rest of the $k-r$ edges to arrange the r T_3 edges. Then, the rest of the $k-2r$ edges are assigned for the T_4 edges. For an innovation, by the relation $f(\ell) = \max\{f(1), \ldots, f(\ell+1)\} + 1$, there is only one way to plot the innovation once the subgraph prior to this innovation is plotted. For an irregular T_3 edge, since there is only one single innovation to be matched, there is thus only one way to plot it when the subgraph prior to this T_3 edge is plotted. By Lemma 5.5, there are at most $t+1$ single innovations to be matched by a regular T_3 edge. That is, each regular T_3 edge has at most $t+1$ ways to plot it. By Lemma 5.6, there are at most $2t$ regular T_3 edges. Hence, there are at most $(t+1)^{2t} \le (t+1)^{2(k-2r)}$ ways to plot the regular T_3-edges. Finally, we consider the T_4 edges. For each T_4 edge, there are at most $(r+1)^2 < k^2$ ways to determine its two vertices. Therefore, there are at most $\binom{k^2}{t}$ ways to plot the t T_2 edges. After the t positions of T_4 are determined, there are at most $t^{k-2r} < (t+1)^{k-2r}$ ways to distribute the $k-2r$ T_4 edges. Finally, from (5.1.10), we obtain

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

In the above, the third inequality follows from the elementary inequality

$$a^bc^{-a} \le (b/\log c)^b, \text{ for all } a \ge 1,\ b>0, \text{ and } c>1, \tag{5.1.11}$$

with $a=t+1$, and the last equality follows from the fact that $\delta_n^{1/3}k/\log n \to 0$. Finally, substituting this into (5.1.9), we obtain

$$\mathrm{P}(\lambda_{\max}(\mathbf{W}_n) \ge \eta) \le n^2(2+o(1)/\eta)^{-k}. \tag{5.1.12}$$

The right-hand side of (5.1.12) is summable due to the fact that $k/\log n \to \infty$. The sufficiency is proved.

5.1.2 Necessity of Conditions of Theorem 5.1

Suppose $\limsup \lambda_{\max}(\mathbf{W}) \le a$, $(a>0)$ a.s. Then, by (A.2.4), $\lambda_{\max}(\mathbf{W}) \ge x_{nn}/\sqrt{n}$. Therefore, $\frac{1}{\sqrt{n}}x_{nn}^+ \le \max\{0, \lambda_{\max}(\mathbf{W})\}$. Hence, for any $\eta > a$, $\mathrm{P}(x_{nn}^+ \ge \eta\sqrt{n}, \text{i.o.}) = 0$. An application of the Borel-Cantelli lemma yields

$$\sum_{n=1}^{\infty}\mathrm{P}(x_{11}^+ \ge \eta\sqrt{n}) < \infty,$$

which implies condition (i).

Define $\mathcal{N}_\ell = \{j; 2^\ell < j \le 2^{\ell+1}; |x_{jj}| \le 2^{\ell/4}\}$ and $p = \mathrm{P}(|x_{11}| \le 2^{\ell/4})$. Write $N_\ell = \#(\mathcal{N})$.

When $n \in (2^{\ell+1}, 2^{\ell+2}]$, for $x_{jk} \ne 0$ and $j,k \in \mathcal{N}_\ell$, construct a unit complex vector $\mathbf{z}$ by taking $z_k = x_{jk}/(\sqrt{2}|x_{jk}|)$, $z_j = 1/\sqrt{2}$, and the remaining elements zero. Substituting $\mathbf{z}$ into the first identity of (5.1.6), we have $\lambda_{\max}(\mathbf{W}) \ge \frac{1}{\sqrt{n}}[|x_{jk}| + \frac{1}{2}(x_{kk} + x_{jj})]$. Thus, we have

$$\lambda_{\max}(\mathbf{W}) \ge 2^{-\ell/2-1} \max_{j,k\in\mathcal{N}_\ell} \{|x_{jk}|\} - 2^{-\ell/4}.$$

The above is trivially true when $x_{jk} = 0$. Thus, for any $\eta > a$, by assumption, we have

$$\mathrm{P}\left(\max_{2^{\ell+1}<n\le 2^{\ell+2}} \lambda_{\max}(\mathbf{W}_n) \ge \eta, \text{ i.o.} \right) = 0.$$

The equality above trivially implies that

$$\mathrm{P}\left(\max_{j,k\in\mathcal{N}_\ell} \{|x_{jk}|\} \ge \eta 2^{\ell/2+2}, \text{ i.o.} \right) = 0.$$

Applying the Borel-Cantelli lemma, we conclude that

$$\sum_{\ell=1}^{\infty} \mathrm{P}\left(\max_{j,k\in\mathcal{N}_\ell} \{|x_{jk}|\} \ge \eta 2^{\ell/2+2} \right) < \infty. \tag{5.1.13}$$

By the independence of N_ℓ and x_{jk}, we have

$$\begin{aligned} &\mathrm{P}\left(\max_{j,k\in\mathcal{N}_\ell} |x_{jk}| \ge \eta 2^{\ell/2+2} \,\middle|\, N_\ell = r \right) \\ &= \mathrm{P}\left(\max_{1\le j<k\le r} |x_{jk}| \ge \eta 2^{\ell/2+2} \right) \\ &= 1 - \left(1 - \mathrm{P}(|x_{12}| \ge \eta 2^{\ell/2+2})\right)^{r(r-1)/2}. \end{aligned}$$

Since N_ℓ has a binomial distribution with success probability p and number of trials 2^ℓ, by noting that p is close to 1 for all large ℓ, we have

$$\begin{aligned} &\mathrm{P}\left(\max_{j,k\in\mathcal{N}_\ell} \{|x_{jk}|\} \ge \eta 2^{\ell/2+2} \right) \\ &\ge \sum_{r=2^{\ell-1}+1}^{2^\ell} \binom{2^\ell}{r} p^r (1-p)^{2^\ell - r} [1 - (1 - \mathrm{P}(|x_{12}| \ge \eta 2^{\ell/2+2}))^{r(r-1)/2}] \\ &\ge \frac{1}{2}(1 - (1 - \mathrm{P}(|x_{12}| \ge \eta 2^{\ell/2+2}))^{2^{2\ell-3}}). \end{aligned}$$

From this and (5.1.13), we obtain

$$\sum_{\ell=1}^{\infty}(1-(1-\mathrm{P}(|x_{12}|\geq\eta 2^{\ell/2+2}))^{2^{2\ell-3}})<\infty.$$

From the relationship of convergence between infinite series and products, we obtain

$$\prod_{\ell=1}^{\infty}\left(1-\mathrm{P}(|x_{12}|\geq\eta 2^{\ell/2+2})\right)^{2^{2\ell-3}}>0,$$

which then implies

$$\sum_{\ell=1}^{\infty}2^{2\ell-3}\mathrm{P}\left(|x_{12}|\geq\eta 2^{\ell/2+2}\right)<\infty,$$

which implies condition (iv).

Now, suppose that $a=\mathrm{E}(\Re(x_{12}))>0$. Define $\mathcal{D}_n=\{j\leq n,|x_{jj}|<n^{1/4}\}$. Write $N=\#(\mathcal{D}_n)$. Define a unit vector $\mathbf{z}=(z_1,\cdots,z_n)$ with $z_j=\frac{1}{\sqrt{N}}$ if $j\in\mathcal{D}_n$ and $z_j=0$ otherwise. Substituting $\mathbf{z}$ into (A.2.4), we get

$$\begin{aligned}
&\lambda_{\max}(\mathbf{W})\geq\mathbf{z}^*(\mathbf{W})\mathbf{z}\\
=&\frac{a(N-1)}{\sqrt{n}}+\frac{1}{N\sqrt{n}}\sum_{i\in\mathcal{D}_n}x_{ii}+\mathbf{z}^*\left(\widetilde{\mathbf{W}}-\mathrm{E}(\widetilde{\mathbf{W}})\right)\mathbf{z}\\
\geq&\frac{a(N-1)}{\sqrt{n}}+\lambda_{\min}\left(\widetilde{\mathbf{W}}-\mathrm{E}(\widetilde{\mathbf{W}})\right)-n^{-1/4}\\
\geq&\frac{aN}{\sqrt{n}}+O(1)\to\infty,
\end{aligned}$$

where $\widetilde{\mathbf{W}}$ is the matrix obtained from $\mathbf{W}$ by replacing its diagonal elements with zero. Here the last limit follows from the fact that $\lambda_{\min}\left(\widetilde{\mathbf{W}}-\mathrm{E}(\widetilde{\mathbf{W}})\right)\to -2\sigma$ almost surely, which is a consequence of the sufficiency part of the theorem, and that $N/n\to 1$ almost surely because N has a binomial distribution with success probability $p=\mathrm{P}(|w_{11}|\leq n^{1/4})\to 1$. Thus, we have derived a contradiction to the assumption that $\limsup\lambda_{\max}(\mathbf{W})=c$ almost surely. This proves that $\Re(\mathrm{E}x_{12})\leq 0$, the second assertion of condition (ii).

To complete the proof of necessity of condition (ii), we need to show that the imaginary part of the expectation of the off-diagonal elements is zero.

Suppose that $b=\Im(\mathrm{E}(x_{12}))\neq 0$. Define a vector $\mathbf{u}=(u_1,\cdots,u_n)'$ by

$$\{u_j,\, j\in\mathcal{D}_n\}=N^{-1/2}\{1,e^{i\pi\mathrm{sign}(b)(2k-1)/N},\cdots,e^{i\pi\mathrm{sign}(b)(2k-1)(N-1)/N}\}$$

and $u_j=0, j\notin\mathcal{D}_n$. By Lemma 2.7,

$$i\mathbf{u}^*\Im(\mathrm{E}(\mathbf{W}_n))\mathbf{u}=\frac{1}{\sqrt{n}}|b|\mathrm{ctan}(\pi(2k-1)/2N).$$

Also,

$$\begin{aligned}\mathbf{u}^*\mathbf{J}\mathbf{u} &= \frac{1}{N}\left|\sum_{j=0}^{N-1} e^{i\pi \text{sign}(b) j(2k-1)/N}\right|^2 \\ &= \frac{1}{N}\left|\frac{1-e^{i\pi \text{sign}(b)(2k-1)}}{1-e^{i\pi \text{sign}(b)(2k-1)/N}}\right|^2 \\ &\le \frac{4}{N\sin^2(\pi(2k-1)/2N)},\end{aligned}$$

where $\mathbf{J}$ is the $n \times n$ matrix of 1's.

Write $a = \mathrm{E}(\Re(x_{12})) \le 0$. Then, by (A.2.4), we have

$$\begin{aligned}&\lambda_{\max}(\mathbf{W}) \ge \mathbf{u}^*\mathbf{W}_n\mathbf{u} \\ &\ge -\frac{4|a|}{\sqrt{n}N\sin^2(\pi(2k-1)/2N)} \\ &\quad +\frac{|b|}{\sqrt{n}\sin(\pi(2k-1)/2N)} + \lambda_{\min}((\widetilde{\mathbf{W}} - \mathrm{E}(\widetilde{\mathbf{W}}))) - n^{-1/4} \\ &:= I_1 + I_2 + I_3 - n^{-1/4}.\end{aligned} \tag{5.1.14}$$

Take $k = [n^{1/3}]$. Then, by the fact that $N/n \to 1$, a.s., we have

$$\begin{aligned}I_1 &\sim -\frac{2|a|\sqrt{n}}{\pi k^2} \to 0, \\ I_2 &\sim \frac{|b|\sqrt{n}}{\pi k} \to \infty, \\ |I_3| &\to -2\sigma.\end{aligned}$$

Thus, the necessity of condition (ii) is proved. Conditions (iii) and (v) follow by applying the sufficiency part. The proof of the theorem is complete.

Remark 5.7. In the proof of Theorem 5.1, if the entries of $\mathbf{W}$ depend on n but satisfy

$$\mathrm{E}(x_{jk}) = 0,\ \mathrm{E}(|x_{jk}^2|) \le \sigma^2,\ \mathrm{E}(|x_{jk}^\ell|) \le b(\delta_n\sqrt{n})^{\ell-3},\ (\ell \ge 3) \tag{5.1.15}$$

for some $b > 0$, then for fixed $\varepsilon > 0$ and $x > 0$,

$$\mathrm{P}(\lambda_{\max}(\mathbf{W}) \le 2\sigma + \varepsilon + x) = o(n^{-\ell}(2\sigma + \varepsilon + x)^{-2}). \tag{5.1.16}$$

This implies that the conclusion of $\limsup \lambda_{\max}(\mathbf{W}) \le 2\sigma$, a.s., is still true.

5.2 Limits of Extreme Eigenvalues of the Sample Covariance Matrix

We first introduce the following theorem.

Theorem 5.8. *Suppose that* $\{x_{jk},\ j,k = 1,2,\cdots\}$ *is a double array of iid random variables with mean zero and variance* σ^2 *and finite fourth moment. Let* $\mathbf{X}_n = (x_{jk},\ j \le p, k \le n)$ *and* $\mathbf{S}_n = \frac{1}{n}\mathbf{X}\mathbf{X}^*$. *Then the largest eigenvalue of* $\mathbf{S}_n$ *tends to* $\sigma^2(1+\sqrt{y})^2$ *almost surely.*

If the fourth moment of the underlying distribution is not finite, then with probability 1, the limsup of the largest eigenvalue of $\mathbf{S}_n$ *is infinity.*

The real case of the first conclusion is due to Yin, Bai, and Krishnaiah [301], and the real case of the second conclusion is proved in Bai, Silverstein, and Yin [33]. The proof of this theorem is almost the same as that of Theorem 5.1, and the proof for the real case can be found in these papers. Thus the details are omitted and left as an exercise for the reader. Here, for our future use, we remark that the proof of the theorem above can be extended to the following.

Theorem 5.9. *Suppose that the entries of the matrix* $\mathbf{X}_n = (x_{jkn},\ j \le p, k \le n)$ *are independent (not necessarily identically distributed) and satisfy*

1. $\mathrm{E}(x_{jkn}) = 0$,
2. $|x_{jkn}| \le \sqrt{n}\delta_n$,
3. $\max_{j,k} |\mathrm{E}|X_{jkn}|^2 - \sigma^2| \to 0$ *as* $n \to \infty$, *and*
4. $\mathrm{E}|x_{jkn}|^\ell \le b(\sqrt{n}\delta_n)^{\ell-3}$ *for all* $\ell \ge 3$,

where $\delta_n \to 0$ *and* $b > 0$. *Let* $\mathbf{S}_n = \frac{1}{n}\mathbf{X}_n\mathbf{X}_n^*$. *Then, for any* $x > \varepsilon > 0$ *and integers* $j,k \ge 2$, *we have*

$$\mathrm{P}(\lambda_{\max}(\mathbf{S}_n) \ge \sigma^2(1+\sqrt{y})^2 + x) \le Cn^{-k}(\sigma^2(1+\sqrt{y})^2 + x - \varepsilon)^{-k}$$

for some constant $C > 0$.

In this section, we shall present a generalization to a result of Bai and Yin [36]. Assume that $\mathbf{X}_n$ is a $p \times n$ complex matrix and $\mathbf{S}_n = \frac{1}{n}\mathbf{X}_n\mathbf{X}_n^*$.

Theorem 5.10. *Assume that the entries of* $\{x_{ij}\}$ *are a double array of iid complex random variables with mean zero, variance* σ^2, *and finite fourth moment. Let* $\mathbf{X}_n = (x_{ij};\, i \le p, j \le n)$ *be the* $p \times n$ *matrix of the upper-left corner of the double array. If* $p/n \to y \in (0,1)$, *then, with probability 1, we have*

$$\begin{aligned} -2\sqrt{y}\sigma^2 &\le \liminf_{n\to\infty} \lambda_{\min}(\mathbf{S}_n - \sigma^2(1+y)\mathbf{I}_n) \\ &\le \limsup_{n\to\infty} \lambda_{\max}(\mathbf{S}_n - \sigma^2(1+y)\mathbf{I}_n) \le 2\sqrt{y}\sigma^2. \end{aligned} \tag{5.2.1}$$

From Theorem 5.10, one immediately gets the following theorem.

Theorem 5.11. *Under the assumptions of Theorem 5.10, we have a.s.*

$$\lim_{n\to\infty} \lambda_{\min}(\mathbf{S}_n) = \sigma^2(1-\sqrt{y})^2 \tag{5.2.2}$$

and

$$\lim_{n\to\infty} \lambda_{\max}(\mathbf{S}_n) = \sigma^2(1+\sqrt{y})^2. \tag{5.2.3}$$

Denote the eigenvalues of $\mathbf{S}_n$ by $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. Write $\lambda_{\max} = \lambda_n$ and

$$\lambda_{\min} = \begin{cases} \lambda_1, & \text{if } p \le n, \\ \lambda_{p-n+1}, & \text{if } p > n. \end{cases}$$

Using the convention above, Theorem 5.11 is true for all $y \in (0, \infty)$.

5.2.1 Proof of Theorem 5.10

We split our tedious proof of Theorem 5.10 into several lemmas. The key idea is to estimate the spectral norm of the power matrix $(\mathbf{S}_n - \sigma^2(1+y)\mathbf{I})^\ell$. In the first step, we split the power matrix into several matrices, among which the most significant matrix is the one called $\mathbf{T}_n(\ell)$ defined below. Lemma 5.12 is devoted to the estimate of the norm of $\mathbf{T}_n(\ell)$. The aim of the subsequent lemmas is to estimate of the norm of $(\mathbf{S}_n - \sigma^2(1+y)\mathbf{I})^\ell$ by using the estimate on $\mathbf{T}_n(\ell)$.

Lemma 5.12. *Under the conditions of Theorem 5.10, we have*

$$\limsup_{n\to\infty} \|\mathbf{T}_n(\ell)\| \le (2\ell+1)(\ell+1)y^{(\ell-1)/2}\sigma^{2\ell} \quad a.s., \tag{5.2.4}$$

where

$$\mathbf{T}_n(\ell) = n^{-\ell}\left(\sum{}' x_{av_1}\overline{x}_{u_1v_1}x_{u_1v_2}\overline{x}_{u_2v_2}\cdots x_{u_{\ell-1}v_\ell}\overline{x}_{bv_\ell}\right),$$

the summation $\sum'$ *runs over for* $v_1, \cdots, v_\ell = 1, 2, \cdots, n$, *and* $u_1, \cdots, u_{\ell-1} = 1, 2, \cdots, p$ *subject to the restriction*

$$a \ne u_1,\ u_1 \ne u_2, \cdots, u_{\ell-1} \ne b \quad \text{and} \quad v_1 \ne v_2,\ v_2 \ne v_3, \cdots, v_{\ell-1} \ne v_\ell.$$

Proof. Without loss of generality, we assume $\sigma = 1$. We first truncate the x-variables. Since $\mathrm{E}|x_{11}|^4 < \infty$, we can select a sequence of slowly decreasing constants $\delta_n \to 0$ such that $n\delta_n$ is increasing and

$$\sum_k \delta_{2^k}^{-2} 2^k \mathrm{E}|x_{11}|^2(|x_{11}| \ge 2^{k/2}\delta_{2^k}) < \infty. \tag{5.2.5}$$

Then, define $x_{ijn} = x_{ij}I(|x_{ij}| \le \delta_n\sqrt{n})$. Next, construct a matrix $\widehat{\mathbf{T}}_n(\ell)$ with the same structure of $\mathbf{T}_n(\ell)$ by replacing x_{ij} with x_{ijn}. Then, we have

$$
\begin{aligned}
&\mathrm{P}\left(\widehat{\mathbf{T}}_n(\ell) \neq \mathbf{T}_n(\ell), \text{i.o.}\right) \\
&\leq \lim_{K\to\infty} \sum_{k=K}^{\infty} \mathrm{P}\left(\bigcup_{2^k<n\leq 2^{k+1}} \bigcup_{i\leq p, j\leq n} (x_{ijn} \neq x_{ij}) \right) \\
&\leq \lim_{K\to\infty} \sum_{k=K}^{\infty} \mathrm{P}\left(\bigcup_{2^k<n\leq 2^{k+1}} \bigcup_{i,j\leq 2^{k+1}} (|x_{ij}| \geq \delta_{2^k} 2^{k/2}) \right) \\
&= \lim_{K\to\infty} \sum_{k=K}^{\infty} \mathrm{P}\left(\bigcup_{i,j\leq 2^{k+1}} (|x_{ij}| \geq \delta_{2^k} 2^{k/2}) \right) \\
&\leq \lim_{K\to\infty} \sum_{k=K}^{\infty} 2^{2k+2}\mathrm{P}\left(|x_{11}| \geq \delta_{2^k} 2^{k/2}\right) = 0.
\end{aligned}
$$

This shows that we need only to show (5.2.4) for the matrix $\widehat{\mathbf{T}}_n(\ell)$.

Let $\widetilde{\mathbf{T}}_n(\ell)$ be the matrix constructed from the variables $x_{ijn} - \mathrm{E}(x_{ijn})$. We shall show that, for all $\ell \geq 0$,

$$
\|\widetilde{\mathbf{T}}_n(\ell) - \widehat{\mathbf{T}}_n(\ell)\| \to 0 \quad \text{a.s.} \tag{5.2.6}
$$

if (5.2.4) is true for the matrix $\widetilde{\mathbf{T}}_n(\ell)$ and all fixed ℓ.

Write $\widehat{\mathbf{Y}}_n = \frac{1}{\sqrt{n}}(x_{ijn})$ and $\widetilde{\mathbf{Y}}_n = \frac{1}{\sqrt{n}}(x_{ijn} - \mathrm{E}(x_{ijn})) = \widehat{\mathbf{Y}}_n - \mathrm{E}(\widehat{\mathbf{Y}}_n)$. Using the notation of $\odot$ products (see Page 483 for the definition of $\odot$), we have

$$
\widehat{\mathbf{T}}_n(\ell) = \widehat{\mathbf{Y}}_n \odot (\widehat{\mathbf{Y}}_n^*) \odot \cdots \odot \widehat{\mathbf{Y}}_n \odot (\widehat{\mathbf{Y}}_n^*) \tag{5.2.7}
$$

and

$$
\widetilde{\mathbf{T}}_n(\ell) = \widetilde{\mathbf{Y}}_n \odot (\widetilde{\mathbf{Y}}_n^*) \odot \cdots \odot \widetilde{\mathbf{Y}}_n \odot (\widetilde{\mathbf{Y}}_n^*). \tag{5.2.8}
$$

Suppose that (5.2.4) is true for the matrix $\widetilde{\mathbf{T}}_n(\ell)$. We will show that

$$
\limsup \|\widetilde{\mathbf{Y}}_n\|^2 \leq 7, \text{a.s.}
$$

In fact, with probability 1,

$$
\begin{aligned}
\limsup \|\widetilde{\mathbf{Y}}_n\|^2 &= \limsup \left\| \widetilde{\mathbf{T}}_n(1) + \mathrm{diag}\left(\frac{1}{n}\sum_{j=1}^{n} |x_{ijn} - \mathrm{E}(x_{ijn})|^2, i \leq p \right) \right\| \\
&\leq 6 + \limsup \frac{1}{n} \max_{i\leq p} \sum_{j=1}^{n} [|x_{ij}|^2 + 2|\mathrm{E}x_{11n}||x_{ij}| + |\mathrm{E}x_{11n}|^2] \\
&\leq 7.
\end{aligned}
$$

Here, the first inequality follows from (5.2.4) for $\widetilde{\mathbf{T}}(\ell)$ and the second follows from Lemma B.25 given in Appendix B and the fact that $\mathrm{E}x_{11n} \to 0$.

Note that $|\mathrm{E}(x_{ijn})| \le (\delta_n\sqrt{n})^{-3}\mathrm{E}(|x_{11}^4|) \to 0$. We obtain $\left\|\mathrm{E}(\widehat{\mathbf{Y}}_n)\right\| \le \sqrt{n}|\mathrm{E}(x_{11n})| = o(1)$.

By (5.2.7) and (5.2.8), the difference $\widetilde{\mathbf{T}}_n(\ell)-\widehat{\mathbf{T}}_n(\ell)$ can be written as a sum of $\odot$ products of matrices $\widetilde{\mathbf{Y}}_n$, $\mathrm{E}(\widehat{\mathbf{Y}}_n)$ or their complex conjugate transpose. Each product has 2ℓ factor matrices, and at least one of them is $\mathrm{E}(\widehat{\mathbf{Y}}_n)$ or its complex conjugate transpose. The number of terms in the sum is $2^\ell - 1$.

Then, assertion (5.2.6) follows by applying Theorem A.23. Therefore, the proof of the lemma reduces to proving (5.2.4) for the matrix $\widetilde{\mathbf{T}}_n(\ell)$. For brevity, we still use $\mathbf{T}_n(\ell)$ and x_{ij} to denote the matrix and variables after truncation and centralization. Namely, we shall proceed with our proof under the following additional assumptions:

$$\begin{aligned} &1.\ \mathrm{E}(x_{ij}) = 0,\quad \mathrm{E}(|x_{ij}|^2) \le 1,\quad \text{and}\quad \mathrm{E}(|x_{ij}|^2) \to 1. \\ &2.\ \mathrm{E}(|x_{ij}|^\ell) \le (\delta_n\sqrt{n})^{\ell-3}\quad \text{for all } \ell \ge 3. \end{aligned} \tag{5.2.9}$$

Select an integer m such that $m/\log n \to \infty$ and $m\delta_n^{1/3}/\log n \to 0$. Then, we have

$$\mathrm{tr}(\mathrm{E}(\mathbf{T}_n^{2m}(\ell))) = n^{-2m\ell}\sum \mathrm{E}(x_{i_1j_1}\overline{x}_{i_2j_1}x_{i_2j_2}\overline{x}_{i_3j_2}\cdots x_{i_{2m\ell}j_{2m\ell}}\overline{x}_{i_1j_{2m\ell}}), \tag{5.2.10}$$

where the summation runs over all integers $i_1,\cdots,i_{2m\ell}$ from $\{1,2,\cdots,p\}$ and $j_1,\cdots,j_{2m\ell}$ from $\{1,2,\cdots,n\}$ subject to the conditions that, for any $v = 0,1,\cdots,2m-1$,

$$\begin{aligned} &i_{v\ell+1} \ne i_{v\ell+2},\ i_{v\ell+2} \ne i_{v\ell+3},\ \cdots,\ i_{(v+1)\ell} \ne i_{(v+1)\ell+1}, \\ &j_{v\ell+1} \ne j_{v\ell+2},\ j_{v\ell+2} \ne j_{v\ell+3},\ \cdots,\ j_{(v+1)\ell-1} \ne j_{(v+1)\ell}. \end{aligned} \tag{5.2.11}$$

Given $i_1,\cdots,i_{2m\ell}$ and $j_1,\cdots,j_{2m\ell}$, define functions f and g by $f(1) = g(1) = 1$ and, for $1 < \ell \le 2m\ell$,

$$f(\ell) = \begin{cases} f(u), & \text{if } i_\ell = i_u \text{ for some } u < \ell, \\ \max\{f(1),\cdots,f(\ell-1)\}+1, & \text{otherwise,} \end{cases}$$

and

$$g(\ell) = \begin{cases} g(u), & \text{if } j_\ell = j_u \text{ for some } u < \ell, \\ \max\{g(1),\cdots,g(\ell-1)\}+1, & \text{otherwise.} \end{cases}$$

Similar to what we did in Subsection 3.1.2, construct a Δ-graph G of $2m\ell$ down edges and $2m\ell$ up edges, by using these two functions, called a canonical graph. Then, (5.2.10) can be rewritten as

$$\mathrm{tr}(\mathrm{E}(\mathbf{T}_n^{2m}(\ell))) = n^{-2m\ell}\sum_G\sum_{\mathbf{i},\mathbf{j}} \mathrm{E}(x_{i_1j_1}\overline{x}_{i_2j_1}x_{i_2j_2}\overline{x}_{i_3j_2}\cdots x_{i_{2m\ell}j_{2m\ell}}\overline{x}_{i_1j_{2m\ell}}). \tag{5.2.12}$$

Obviously, if G has a single edge, the terms corresponding to this graph are zero. Thus, we need only to estimate the sum of all those terms whose G has no single edge.

We split the graph G into $2m$ subgraphs $G_1, \cdots, G_{2m}$, where the graph G_v consists of the ℓ down edges $e_{d,(v-1)\ell+1} \cdots, e_{d,v\ell}$ and the ℓ up edges $e_{u,(v-1)\ell+1} \cdots, e_{u,v\ell}$ and their vertices.

Due to condition (5.2.9), within each subgraph, all down edges do not coincide with their adjacent (prior to or behind) up edges, but those between different subgraphs can be coincident.

Now, we begin to estimate the right-hand side of (5.2.12). Let k denote the total number of innovations and t denote the number of T_2 edges. Then, we have the following.

1. By noting that G has no single edge, the expectation is not greater than $(\delta_n \sqrt{n})^{4m\ell-2k-t}$.

2. Since G has no single edge, $1 \le k \le 2m\ell$.

3. For the same reason, $t \le 2m\ell$.

4. Let a_i denote the number of pairs of consecutive edges (e, e') in the subgraph G_i in which e is an innovation but e' is not. Then the number of sequences of consecutive innovations in G_i is either a_i or $a_i + 1$ (the latter happens when the last edge of G_i is an innovation). Hence, the number of ways to arrange the consecutive innovation sequences is not more than

$$\binom{2\ell}{2a_i} + \binom{2\ell}{2a_i+1} = \binom{2\ell+1}{2a_i+1}.$$

5. Given the positions of innovations, there are at most $\binom{4m\ell-k}{k}$ ways to arrange the T_3 edges.

6. Given the positions of innovations and T_3 edges, the T_4 edges will be at the rest $4m\ell - 2k$ positions.

7. For canonical graphs, there is only one way to plot the innovations and irregular T_3-edges. By Lemmas 5.5 and 5.6, there are at most $(t+1)^{2(4m\ell-2k)}$ ways to plot the regular T_3 edges.

8. Because there are $k+1$ vertices, there are at most $\binom{k+1}{2} (< (k+1)^2)$ positions to plot the T_2 edges. Hence, there are at most $\binom{(k+1)^2}{t}$ ways to plot the t T_2 edges. And there are at most $t^{4m\ell-2k}$ ways to distribute the $4m\ell - 2k$ T_4-edges into the t positions.

9. Let r and s denote the number of up and down innovations. Then, we have $k = r + s$ and the number of terms for each canonical graph is not more than $n^s p^{r+1} = n^{k+1} y_n^{r+1}$, where $y_n = p/n \to y$.

Combining the above, (5.2.12) can be estimated by

$$\mathrm{tr}(\mathrm{E}(\mathbf{T}_n^{2m}(\ell))) \le n \sum{}^* \left(\prod_{i=1}^{2m} \binom{2\ell+1}{2a_i+1} \right) \binom{4m\ell-k}{k} \binom{(k+1)^2}{t}$$

$$\cdot (t+1)^{3(4m\ell-2k)} y_n^{r+1} \delta_n^{4m\ell-2k} (\sqrt{n}\delta_n)^{-t}, \qquad (5.2.13)$$

where the summation is taken subject to restrictions $1 \le k \le 2m\ell$, $0 \le t \le 2m\ell$, and $0 \le a_i \le \ell$.

Suppose that in the pair (e, e') of the graph G_i, e is an innovation and e' is not. Then, e' must be a T_2 edge since it cannot coincide with e. Therefore,

$$a_1 + \cdots + a_{2m} \le \text{the number of } T_2 \text{ edges} \le t. \qquad (5.2.14)$$

In each consecutive sequence of innovations, the difference in the number of up and down innovations is at most 1. Since in G_i there are at most $a_i + 1$ consecutive innovation sequences, we obtain

$$|r - s| \le a_1 + \cdots + a_{2m} + 2m. \qquad (5.2.15)$$

From the estimations above and the relation $k = r + s$, we have

$$r \ge \frac{1}{2}(k - t) - m.$$

Since $y_n \le 1$, we have

$$y_n^{r+1} \le y_n^{(k-t-2m)/2}. \qquad (5.2.16)$$

By the trivial inequality $\binom{2\ell+1}{2a_i+1} \le (2\ell+1)^{2a_i+1}$, we have

$$\prod_{i=1}^{2m} \binom{2\ell+1}{2a_i+1} \le (2\ell+1)^{2\sum a_i + 2m} \le (2\ell+1)^{2t+2m}. \qquad (5.2.17)$$

Because each a_i may take values from 0 to ℓ, there are at most $(\ell+1)^{2m}$ ways to arrange various $a_1, \cdots, a_{2m}$. Then, applying inequality (5.1.11), we further get

$$\begin{aligned}
\operatorname{tr}(\mathrm{E}(\mathbf{T}_n^{2m}(\ell))) &\le n \sum_{k=1}^{2m\ell} \sum_{t=0}^{4m\ell-2k} (2\ell+1)^{2m} (\ell+1)^{2m} \binom{4m\ell-k}{k} \\
&\quad \cdot \left(\frac{(k+1)^2}{\sqrt{n}\delta_n} \right)^t (t+1)^{3(4m\ell-2k)} y_n^{\frac{1}{2}(k-t-2m)} \delta_n^{4m\ell-2k} \\
&\le n^2 (2\ell+1)^{2m} (\ell+1)^{2m} y_n^{-m} \sum_{k=1}^{2m\ell} \binom{4m\ell-k}{k} \\
&\quad \cdot \left(\frac{3(4m\ell-2k)\delta_n^{1/3}}{\log(\sqrt{ny_n}\delta_n/(k+1)^2)} \right)^{3(4m\ell-2k)} y_n^{k/2} \delta_n^{4m\ell-2k} \\
&\le n^2 (2\ell+1)^{2m} (\ell+1)^{2m} y_n^{-m} \left[y_n^{1/4} + \left(\frac{24m\ell\delta_n^{1/3}}{\frac{1}{2}\log n} \right)^3 \right]^{4m\ell}
\end{aligned}$$

$$= n^2(2\ell+1)^{2m}(\ell+1)^{2m}y_n^{m(\ell-1)}(1+o(1))^{4m\ell}. \tag{5.2.18}$$

Thus, for any $\eta > (2\ell+1)(\ell+1)y^{(\ell-1)/2}$, we have

$$\begin{aligned} &\mathrm{P}(\|\mathbf{T}_n(\ell)\| \geq \eta) \leq \eta^{-2m}\mathrm{E}(\|\mathbf{T}_n(\ell)\|^{2m}) \\ &\leq \eta^{-2m}\mathrm{E}(\mathrm{tr}(\mathbf{T}_n(\ell))^{2m}) \\ &\leq \left(\eta^{-1}(2\ell+1)(\ell+1)y^{(\ell-1)/2}\right)^{2m} n^2(1+o(1))^{4m\ell} \end{aligned} \tag{5.2.19}$$

which is summable due to the assumption $m/\log n \to \infty$. The lemma then follows by applying the Borel-Cantelli lemma.

Define $\mathbf{Y}_n^{(2f+1)} = (n^{-f-1/2}|x_{ij}|^{2f}x_{ij})$ and $\mathbf{Y}_n^{(2f+2)} = (n^{-f-1}|x_{ij}|^{2f+2})$, $f = 0, 1, 2, \cdots$. Then, we have the following lemma.

Lemma 5.13. *Under the conditions of Theorem 5.10, we have*

$$\limsup \|\mathbf{Y}_n^{(1)}\| \leq \sqrt{7}\sigma, \quad \text{a.s.},$$

$$\limsup \|\mathbf{Y}_n^{(2)}\| \leq \sqrt{\mathrm{E}|x_{11}|^4}, \quad \text{a.s.},$$

$$\limsup \|\mathbf{Y}_n^{(f)}\| = 0, \quad \text{a.s., for all } f > 2.$$

Proof. We have

$$\|\mathbf{Y}_n^{(1)}\|^2 \leq \|\mathbf{T}_n(1)\| + \frac{1}{n}\max_{i\leq p}\sum_{j=1}^n |x_{ij}|^2.$$

Then, the first conclusion of the lemma follows from Lemmas 5.12 and B.25.

By $\|\mathbf{Y}_n^{(2)}\|^2 \leq \mathrm{tr}(\mathbf{Y}_n^{(2)}\mathbf{Y}_n^{(2)*})$, we have

$$\|\mathbf{Y}_n^{(2)}\|^2 \leq n^{-2}\sum_{ij}|x_{ij}|^4 \to y\mathrm{E}(|x_{11}|^4), \text{ a.s.}$$

For $f > 2$, by Lemma B.25, we have

$$\|\mathbf{Y}_n^{(f)}\|^2 \leq n^{-f}\sum_{ij}|x_{ij}|^{2f} \to 0 \;\text{ a.s.}$$

Lemma 5.14. *Under the conditions of Theorem 5.10, we have*

$$\mathbf{T}_n\mathbf{T}_n(k) = \mathbf{T}_n(k+1) + y\sigma^2\mathbf{T}_n(k) + y\sigma^4\mathbf{T}_n(k-1) + o(1) \;\text{ a.s.} \tag{5.2.20}$$

Proof. We can assume $\sigma = 1$ without loss of generality. By relation (A.3.6) and Lemma 5.13,

$$\mathbf{T}_n(k) = \mathbf{Y}_n\left(\overbrace{\mathbf{Y}_n^* \odot \mathbf{Y}_n^* \odot \cdots \odot \mathbf{Y}_n^*}^{k\ \mathbf{Y}_n^*\text{'s}}\right)$$

$$-[\mathrm{diag}(\mathbf{Y}_n\mathbf{Y}_n^*)]\mathbf{T}_n(k-1) + \mathbf{Y}_n^{(3)} \odot (\mathbf{Y}_n^* \odot \cdots \odot \mathbf{Y}_n^*)$$
$$= \mathbf{Y}_n\,(\mathbf{Y}_n^* \odot \mathbf{Y}_n^* \odot \cdots \odot \mathbf{Y}_n^*) - \mathbf{T}_n(k-1) + o(1) \quad \text{a.s.}, \quad (5.2.21)$$

and similarly

$$\begin{aligned}
\mathbf{T}_n(k+1) &= \mathbf{Y}_n \left(\overbrace{\mathbf{Y}_n^* \odot \mathbf{Y}_n \odot \cdots \odot \mathbf{Y}_n^*}^{k+1\ \mathbf{Y}_n^*\text{'s}} \right) - [\mathrm{diag}(\mathbf{Y}_n\mathbf{Y}_n^*)]\mathbf{T}_n(k) + o(1) \quad \text{a.s.} \\
&= \mathbf{Y}_n\mathbf{Y}_n^*\mathbf{T}_n(k) - \mathbf{Y}_n \mathrm{diag}(\mathbf{Y}_n^*\mathbf{Y}_n^*)\,(\mathbf{Y}_n^* \odot \cdots \odot \mathbf{Y}_n^*) \\
&\quad -\mathrm{diag}(\mathbf{Y}_n\mathbf{Y}_n^*)\mathbf{T}_n(k) + o(1) \quad \text{a.s.} \\
&= \mathbf{T}_n\mathbf{T}_n(k) - y\mathbf{Y}_n\,(\mathbf{Y}_n^* \odot \cdots \odot \mathbf{Y}_n^*) + o(1) \quad \text{a.s.} \\
&= \mathbf{T}_n\mathbf{T}_n(k) - y(\mathbf{T}_n(k) + \mathbf{T}_n(k-1)) + o(1) \quad \text{a.s.} \qquad (5.2.22)
\end{aligned}$$

The proof of the lemma is complete.

Lemma 5.15. *Under the conditions of Theorem 5.10, we have*

$$(\mathbf{T}_n - y\sigma^2\mathbf{I}_p)^k = \sum_{r=0}^{k} (-1)^{r+1}\sigma^{2(k-r)}\mathbf{T}(r) \sum_{i=0}^{[(k-r)/2]} C_i(k,r)y^{k-r-i} + o(1), \qquad (5.2.23)$$

where the constants $|C_i(k,r)| \le 2^k$.

Proof. When $k = 1$, with the convention that $\mathbf{T}(0) = \mathbf{I}$, the lemma is trivially true with $C_0(1,1) = 1$ and $C_0(1,0) = 1$. The general case can easily be proved by induction and Lemma 5.14. The details are omitted.

We are now in a position to prove Theorem 5.10.

Proof of Theorem 5.10. Again, we assume that $\sigma^2 = 1$ without loss of generality. By Lemma B.25, we have

$$\|\mathbf{S}_n - \mathbf{I}_p - \mathbf{T}_n\| \le \max_{i \le p} \left| \sum_{j=1}^{n} (|x_{ij}|^2 - 1) \right| \to 0 \quad \text{a.s.}$$

Therefore, to prove Theorem 5.10, we need only to show that

$$\limsup \|\mathbf{T}_n - y\mathbf{I}_p\| \le 2\sqrt{y} \quad \text{a.s.}$$

By Lemmas 5.12 and 5.15, for any fixed k, we have

$$\limsup \|\mathbf{T}_n - y\mathbf{I}_p\|^k \le Ck^4 2^k y^{(k-1)/2}.$$

Therefore,

$$\limsup \|\mathbf{T}_n - y\mathbf{I}_p\| \le C^{1/k}k^{4/k}2y^{(k-1)/(2k)}.$$

Letting $k \to \infty$, we conclude the proof of Theorem 5.10.

5.2.2 Proof of Theorem 5.11

By Theorem 3.6, with probability 1, we have

$$\limsup \lambda_{\min}(\mathbf{S}_n) \le \sigma^2(1-\sqrt{y})^2 \quad \text{and} \quad \liminf \lambda_{\max}(\mathbf{S}_n) \ge \sigma^2(1+\sqrt{y})^2.$$

Then, by Theorem 5.10,

$$\begin{aligned}\limsup \lambda_{\max}(\mathbf{S}_n) &= \sigma^2(1+y) + \limsup \lambda_{\max}(\mathbf{S}_n - \sigma^2(1+y)\mathbf{I}_p)\\ &\le \sigma^2(1+y) + 2\sigma^2\sqrt{y}\end{aligned}$$

and

$$\begin{aligned}\liminf \lambda_{\min}(\mathbf{S}_n) &= \sigma^2(1+y) + \liminf \lambda_{\min}(\mathbf{S}_n - \sigma^2(1+y)\mathbf{I}_p)\\ &\ge \sigma^2(1+y) - 2\sigma^2\sqrt{y}.\end{aligned}$$

This completes the proof of the theorem.

5.2.3 Necessity of the Conditions

By the elementary inequality $\lambda_{\max}(\mathbf{A}) \ge \max_{i\le p} a_{ii}$, we have

$$\lambda_{\max}(\mathbf{S}_n) \ge \max_{i\le p} \frac{1}{n}\sum_{j=1}^{n} |x_{ij}|^2.$$

By Lemma B.25, if $\mathrm{E}(|x_{11}|^4) = \infty$, then

$$\limsup_{n\to\infty} \max_{i\le p} \frac{1}{n}\sum_{j=1}^{n} |x_{ij}|^2 \to \infty, \text{ a.s.}$$

This shows that the finiteness of the fourth moment of the underlying distribution is necessary for the almost sure convergence of the largest eigenvalue of a sample covariance matrix.

If $\mathrm{E}(|x_{11}|^4) < \infty$ but $\mathrm{E}(x_{11}) = a \ne 0$, then

$$\left\|\frac{1}{\sqrt{n}}\mathbf{X}_n\right\| \ge \left\|\frac{1}{\sqrt{n}}(a\mathbf{J})\right\| - \left\|\frac{1}{\sqrt{n}}(\mathbf{X}_n - \mathrm{E}(\mathbf{X}_n))\right\|$$

$$\ge |a|p/\sqrt{n} - \left\|\frac{1}{\sqrt{n}}(\mathbf{X}_n - \mathrm{E}(\mathbf{X}_n))\right\| \to \infty, \text{a.s.}$$

Combining the above, we have proved that the necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a large dimensional sample covariance matrix are that the underlying distribution has a zero mean and finite fourth moment.

Remark 5.16. It seems that the finiteness of the fourth moment is also necessary for the almost sure convergence of the smallest eigenvalue of the large dimensional sample covariance matrix. However, at this point we have no idea how to prove it.

5.3 Miscellanies

5.3.1 Spectral Radius of a Nonsymmetric Matrix

Let $\mathbf{X}$ be an $n \times n$ matrix of iid complex random variables with mean zero and variance σ^2. In Bai and Yin [39], large systems of linear equations and linear differential equations are considered. There, the norm of the matrix $(\frac{1}{\sqrt{n}}\mathbf{X})^k$ plays an important role in the stability of the solutions to those systems. The following theorem is established.

Theorem 5.17. *If* $\mathrm{E}(|x_{11}^4|) < \infty$, *then*

$$\limsup_{n\to\infty} \left\| \left(\frac{1}{\sqrt{n}} \mathbf{X} \right)^k \right\| \le (1+k)\sigma^k, \quad \text{a.s.} \tag{5.3.1}$$

The proof of this theorem, after truncation and centralization, relies on the estimation of $\mathrm{E}([\mathrm{tr}(\frac{1}{\sqrt{n}}\mathbf{X})^k(\frac{1}{\sqrt{n}}\mathbf{X}^*)^k]^\ell)$. The details are omitted. Here, we introduce an important consequence on the spectral radius of $\frac{1}{\sqrt{n}}\mathbf{X}$, which plays an important role in establishing the circular law (see Chapter 11). This was also independently proved by Geman [117] under additional restrictions on the growth of moments of the underlying distribution.

Theorem 5.18. *If* $\mathrm{E}(|x_{11}^4|) < \infty$, *then*

$$\limsup_{n\to\infty} \left| \lambda_{\max} \left(\frac{1}{\sqrt{n}} \mathbf{X} \right) \right| \le \sigma, \quad \text{a.s.} \tag{5.3.2}$$

Theorem 5.18 follows from the fact that, for any k,

$$\begin{aligned} &\limsup_{n\to\infty} \left| \lambda_{\max} \left(\tfrac{1}{\sqrt{n}} \mathbf{X} \right) \right| = \limsup_{n\to\infty} \left| \lambda_{\max} \left[\left(\tfrac{1}{\sqrt{n}} \mathbf{X} \right)^k \right] \right|^{1/k} \\ &\le \limsup_{n\to\infty} \left\| (\tfrac{1}{\sqrt{n}} \mathbf{X})^k \right\|^{1/k} \le (1+k)^{1/k}\sigma \to \sigma \end{aligned}$$

by making $k \to \infty$.

Remark 5.19. Checking the proof of Theorem 5.11, one finds that, after truncation and centralization, the conditions for guaranteeing (5.3.1) are $|x_{jk}| \leq \delta_n\sqrt{n}$, $\mathrm{E}(|x_{jk}^2|) \leq \sigma^2$, and $\mathrm{E}(|x_{jk}^3|) \leq b$, for some $b > 0$. This is useful in extending the circular law to the case where the entries are not identically distributed.

5.3.2 TW Law for the Wigner Matrix

In multivariate analysis, certain statistics are defined in terms of the extreme eigenvalues of random matrices, which makes the limiting distribution of normalized extreme eigenvalues of special interest. In [279], Tracy and Widom derived the limiting distribution of the largest eigenvalue of a Wigner matrix when the entries are Gaussian distributed. The limiting law is named the Tracy-Widom (TW) law in RMT. We shall introduce the TW law for the Gaussian Wigner matrix. Under the normality assumption, the density function of the ensemble is given by

$$P(\mathbf{w})d\mathbf{w} = C_\beta \exp\left(-\frac{\beta}{4}\mathrm{tr}\,\mathbf{w}^*\mathbf{w}\right)d\mathbf{w}$$

and the joint density of the eigenvalues is given by

$$p_{n\beta}(\lambda_1, \cdots, \lambda_n) = C_{n\beta}e^{-\frac{1}{2}\beta\sum\lambda_j^2}\prod_{j<k}|\lambda_j - \lambda_k|^\beta,\ -\infty < \lambda_1 < \cdots < \lambda_n < \infty,$$

where

$$\beta = \begin{cases} 1 & \text{for GOE,} \\ 2 & \text{for GUE,} \\ 4 & \text{for GSE,} \end{cases}$$

where *GOE* stands for *Gaussian orthogonal ensemble*, for which all entries of the matrix are real normal random variables and whose distribution is invariant under real orthogonal similarity transformations; *GUE* stands for *Gaussian unitary ensemble,* for which all entries of the matrix are complex normal random variables and whose distribution is invariant under complex unitary similarity transformations; while *GSE* stands for *Gaussian symplectic ensemble*, for which all entries of the matrix are normal quaternion random variables and whose distribution is invariant under symplectic transformations.

It is necessary here to provide a note on quaternions for the Wigner matrix and GSE. We define 2×2 matrices

$$\mathbf{e} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\quad \mathbf{i} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},\quad \mathbf{j} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\quad \text{and}\quad \mathbf{k} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

It is easy to verify that $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -\mathbf{e}$. For any real numbers a, b, c, d, the 2×2 matrix of the linear combination

$$x = a\mathbf{e} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} = \begin{pmatrix} u & v \\ -\bar{v} & \bar{u} \end{pmatrix}$$

is called a quaternion, where $u = a + bi$ and $v = c + di$. The quaternion conjugate of x is defined by

$$\bar{x} = a\mathbf{e} - b\mathbf{i} - c\mathbf{j} - d\mathbf{k} = \begin{pmatrix} \bar{u} & -v \\ \bar{v} & u \end{pmatrix}$$

which is a complex conjugate transpose of the quaternion x. An $n \times n$ quaternion matrix consists of n^2 quaternion entries and thus is in fact a $2n \times 2n$ complex matrix. By the property of quaternion conjugation, the quaternion conjugate transpose of a quaternion matrix is the same as the usual complex conjugate transpose of its $2n \times 2n$ complex matrix version. By this property, we can similarly define a quaternion Hermitian matrix by $\mathbf{X} = \mathbf{X}^*$, where $\mathbf{X}^*$ stands for quaternion conjugate transpose of the quaternion matrix $\mathbf{X}$.

A GSE is a $2n \times 2n$ Hermitian matrix $\mathbf{X} = (x_{ij})_{i,j=1}^n$, where x_{ij} is a quaternion with four coefficients being iid $N(0, 1/4)$ for $i > j$ and the diagonal elements of $x_{ii} = a_i\mathbf{e}$ with $a_i \overset{\text{iid}}{\sim} N(0, 1)$ and the quaternions above or on the diagonal are independent. Such a matrix is called GSE because its distribution is invariant under symplectic transformations. We shall not introduce these transformations here. Interested readers are referred to Section 2.4 of Mehta [212].

It is well known that all eigenvalues of a GSE are real and have multiplicities 2 and thus GSEs have n distinct eigenvalues.

In Tracy and Widom [279], the following theorem is proved.

Theorem 5.20. *Let λ_n denote the largest eigenvalue of an order n GOE, GUE, or GSE. Then*

$$n^{2/3}(\lambda_n - 2) \xrightarrow{\mathscr{D}} T_\beta,$$

where T_β is a random variable whose distribution function F_β is given by

$$F_2(x) = \exp\left(-\int_x^\infty (t - x) q^2(t) dt \right),$$
$$F_1(x) = \exp\left(-\frac{1}{2}\int_x^\infty q(t) dt \right) [F_2(x)]^{1/2},$$
$$F_4(2^{-1/2}x) = \cosh\left(-\frac{1}{2}\int_x^\infty q(t) dt \right) [F_2(x)]^{1/2},$$

and $q(t)$ is the solution to the differential equation

$$q'' = tq + 2q^3$$

(solutions to which are called Painlevé functions of type II) satisfying the marginal condition

$$q(t) \sim Ai(t), \text{ as } t \to \infty$$

and Ai is the Airy function.

The descriptions of the Airy and the TW distribution functions are complicated. For an intuitive understanding of the TW distributions, we present a graph of their densities (see Fig. 5.7).

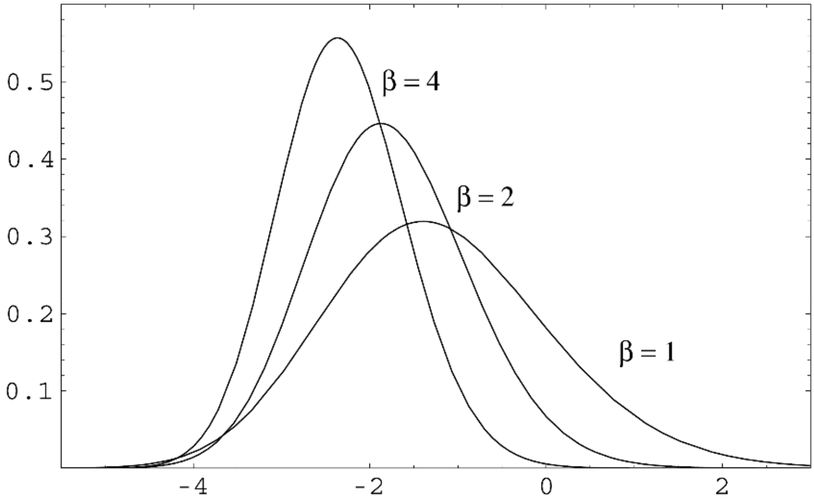

Fig. 5.7 The density function of F_β for $\beta = 1, 2, 4$.

5.3.3 TW Law for a Sample Covariance Matrix

It is interesting that the normalized largest eigenvalue of the standard Wishart matrix tends to the same TW law under the assumption of normality. The following result was established by Johnstone [168].

We first consider the real case.

Theorem 5.21. *Let $\lambda_{\max}$ denote the largest eigenvalue of the real Wishart matrix $W(n, I_p)$. Define*

$$\mu_{n,p} = (\sqrt{n-1} + \sqrt{p})^2,$$
$$\sigma_{n,p} = (\sqrt{n-1} + \sqrt{p}) \left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}} \right)^{1/3}.$$

Then

$$\frac{\lambda_{\max} - \mu_{n,p}}{\sigma_{n,p}} \xrightarrow{\mathscr{D}} W_1 \sim F_1,$$

where F_1 is the TW distribution with $\beta = 1$.

The complex Wishart case is due to Johansson [164].

Theorem 5.22. *Let $\lambda_{\max}$ denote the largest eigenvalue of a complex Wishart matrix $W(n, I_p)$. Define*

$$\mu_{n,p} = (\sqrt{n} + \sqrt{p})^2,$$
$$\sigma_{n,p} = (\sqrt{n} + \sqrt{p}) \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p}} \right)^{1/3}.$$

Then

$$\frac{\lambda_{\max} - \mu_{n,p}}{\sigma_{n,p}} \xrightarrow{\mathscr{D}} W_2 \sim F_2,$$

where F_2 is the TW distribution with $\beta = 2$.

Chapter 6
Spectrum Separation

6.1 What Is Spectrum Separation?

The results in this chapter are based on Bai and Silverstein [32, 31]. We consider the matrix $\mathbf{B}_n = \frac{1}{n}\mathbf{T}^{1/2}\mathbf{X}_n\mathbf{X}_n^*\mathbf{T}_n^{1/2}$, where $\mathbf{T}_n^{1/2}$ is a Hermitian square root of the Hermitian nonnegative definite $p \times p$ matrix $\mathbf{T}_n$, with $\mathbf{X}_n$ and $\mathbf{T}_n$ satisfying the (a.s.) assumptions of Theorem 4.1. We will investigate the spectral properties of $\mathbf{B}_n$ in relation to the eigenvalues of $\mathbf{T}_n$. A relationship is expected to exist since, for nonrandom $\mathbf{T}_n$, $\mathbf{B}_n$ can be viewed as the sample covariance matrix of n samples of the random vector $\mathbf{T}_n^{1/2}\mathbf{x}_1$, which has $\mathbf{T}_n$ for its population matrix. When n is significantly larger than p, the law of large numbers tells us that $\mathbf{B}_n$ will be close to $\mathbf{T}_n$ with high probability. Consider then an interval $J \subset \mathbb{R}^+$ that does not contain any eigenvalues of $\mathbf{T}_n$ for all large n. For small y (to which p/n converges), it is reasonable to expect an interval $[a, b]$ close to J which contains no eigenvalues of $\mathbf{B}_n$. Moreover, the number of eigenvalues of $\mathbf{B}_n$ on one side of $[a, b]$ should match up with those of $\mathbf{T}_n$ on the same side of J. Under the assumptions on the entries of $\mathbf{X}_n$ given in Theorem 5.11 with $\sigma^2 = 1$, this can be proven using the Fan Ky inequality (see Theorem A.10).

Extending the notation introduced in Theorem A.10 to eigenvalues and, for notational convenience, defining $\lambda_0^{\mathbf{A}} = \infty$, suppose $\lambda_{i_n}^{\mathbf{T}_n}$ and $\lambda_{i_n+1}^{\mathbf{T}_n}$ lie, respectively, to the right and left of J. From Theorem A.10, we have (using the fact that the spectra of $\mathbf{B}_n$ and $(1/n)\mathbf{X}_n\mathbf{X}_n^*\mathbf{T}_n$ are identical)

$$\lambda_{i_n+1}^{\mathbf{B}_n} \le \lambda_1^{(1/n)\mathbf{X}_n\mathbf{X}_n^*}\lambda_{i_n+1}^{\mathbf{T}_n} \qquad \text{and} \qquad \lambda_{i_n}^{\mathbf{B}_n} \ge \lambda_p^{(1/n)\mathbf{X}_n\mathbf{X}_n^*}\lambda_{i_n}^{\mathbf{T}_n}. \qquad (6.1.1)$$

From Theorem 5.11, we can, with probability 1, ensure that $\lambda_1^{(1/n)\mathbf{X}_n\mathbf{X}_n^*}$ and $\lambda_p^{(1/n)\mathbf{X}_n\mathbf{X}_n^*}$ are as close as we please to one by making y suitably small. Thus, an interval $[a, b]$ does indeed exist that separates the eigenvalues of $\mathbf{B}_n$ in exactly the same way the eigenvalues of $\mathbf{T}_n$ are split by J. Moreover, a, b can be made arbitrarily close to the endpoints of J.

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_6,

Even though the splitting of the support of F, the a.s. LSD of $F^{\mathbf{B}_n}$ (guaranteed by Theorem 4.1), is a function of y (more details will be given later), splitting may occur regardless of whether y is small or not. Our goal is to extend the result above on exact separation beginning with any interval $[a, b]$ of $\mathbb{R}^+$ outside the support of F. We present an example of its importance that was the motivating force behind the pursuit of this topic. It arises from the detection problem in array signal processing. An unknown number q of sources emit signals onto an array of p sensors in a noise-filled environment ($q < p$). If the population covariance matrix $\mathbf{R}$ of the vector of random values recorded from the sensors is known, then the value q can be determined from it due to the fact that the multiplicity of the smallest eigenvalue of $\mathbf{R}$, attributed to the noise, is $p - q$. The matrix $\mathbf{R}$ is approximated by a sample covariance matrix $\widehat{\mathbf{R}}$, which, with a sufficiently large sample, will have with high probability $p - q$ *noise* eigenvalues clustering near each other and to the left of the other eigenvalues. The problem is that for p and/or q sizable, the number of samples needed for $\widehat{\mathbf{R}}$ to adequately approximate $\mathbf{R}$ would be prohibitively large. However, if for p large the number n of samples were to be merely of the same order of magnitude as p, then, under certain conditions on the signals and noise propagation, it is shown in Silverstein and Combettes [268] that $F^{\widehat{\mathbf{R}}}$ would, with high probability, be close to the nonrandom LSD F. Moreover, it can be shown that, for y sufficiently small, the support of F will split into two parts, with mass $(p - q)/p$ on the left and q/p on the right. In Silverstein and Combettes [268], extensive computer simulations were performed to demonstrate that, at the least, the proportion of sources to sensors can be reliably estimated. It came as a surprise to find that not only were there no eigenvalues outside the support of F (except those near the boundary of the support) but the *exact* number of eigenvalues appeared on intervals slightly larger than those within the support of F. Thus, the simulations demonstrate that, in order to detect the *number* of sources in the large dimensional case, it is not necessary for $\widehat{\mathbf{R}}$ to be close to $\mathbf{R}$; the number of samples only needs to be large enough so that the support of F splits.

It is of course crucial to be able to recognize and characterize intervals outside the support of F and to establish a correspondence with intervals outside the support of H, the LSD of $F^{\mathbf{T}_n}$. This is achieved through the Stieltjes transforms, $s_F(z)$ and $\underline{s}(z) \equiv s_{\underline{F}}(z)$, of, respectively, F and $\underline{F}$, where the latter denotes the LSD of $\underline{\mathbf{B}}_n \equiv (1/n)\mathbf{X}_n^*\mathbf{T}_n\mathbf{X}_n$. From Theorem 4.3, it is conceivable and will be proven that for each $z \in \mathbb{C}^+$, $s = s_F(z)$ is a solution to the equation

$$s = \int \frac{1}{t(1 - y - yzs) - z} dH(t), \tag{6.1.2}$$

which is unique in the set $\{s \in \mathbb{C} : -(1 - y)/z + ys \in \mathbb{C}^+\}$. Since the spectra of $\mathbf{B}_n$ and $\underline{\mathbf{B}}_n$ differ by $|p - n|$ zero eigenvalues, it follows that

$$F^{\underline{\mathbf{B}}_n} = (1 - (p/n))I_{[0,\infty)} + (p/n)F^{\mathbf{B}_n},$$

from which we get

$$s_{F^{\underline{\mathbf{B}}_n}}(z) = -\frac{(1-p/n)}{z} + (p/n)s_{F^{\mathbf{B}_n}}(z), \quad z \in \mathbb{C}^+, \tag{6.1.3}$$

$$\underline{F} = (1-y)I_{[0,\infty)} + yF,$$

and

$$s_{\underline{F}}(z) = -\frac{(1-y)}{z} + ys_F(z), \quad z \in \mathbb{C}^+.$$

It follows that

$$s_F = -z^{-1}\int \frac{1}{1+ts_{\underline{F}}}dH(t), \tag{6.1.4}$$

for each $z \in \mathbb{C}^+$, $\underline{s} = s_{\underline{F}}(z)$, is the unique solution in $\mathbb{C}^+$ to the equation

$$\underline{s} = -\left(z - y\int \frac{t\,dH(t)}{1+t\underline{s}}\right)^{-1}, \tag{6.1.5}$$

and $s_{\underline{F}}(z)$ has an inverse, explicitly given by

$$z(s) = z_{y,H}(s) \equiv -\frac{1}{s} + y\int \frac{t\,dH(t)}{1+ts}. \tag{6.1.6}$$

Note that this could have been derived from (4.1.2) by setting $s_A = -z^{-1}$. The unique solution to (6.1.2) follows from (4.5.8).

Let $F^{y,H}$ denote $\underline{F}$ in order to express the dependence of the LSD of $F^{\underline{\mathbf{B}}_n}$ on the limiting dimension to sample size ratio y and LSD H of the population matrix. Then $s = s_{F^{y,H}}(z)$ has inverse $z = z_{y,H}(s)$.

From (6.1.6), much of the analytic behavior of F can be derived (see Silverstein and Choi [267]). This includes the continuous dependence of F on y and H, the fact that F has a continuous density on $\mathbb{R}^+$, and, most importantly for our present needs, a way of understanding the support of F. On any closed interval outside the support of $F^{y,H}$, $s_{F^{y,H}}$ exists and is increasing. Therefore, on the range of this interval, its inverse exists and is also increasing. In Silverstein and Choi [267], the converse is shown to be true along with some other results. We summarize the relevant facts in the following lemma.

Lemma 6.1. *(Silverstein and Choi [267]). For any c.d.f. G, let S_G denote its support and S_G^c, the complement of its support. If $u \in S_{F^{y,H}}^c$, then $s = s_{F^{y,H}}(u)$ satisfies:*

(1) $s \in \mathbb{R}\backslash\{0\}$,

(2) $-s^{-1} \in S_H^c$,

and

(3) $\frac{d}{ds}z_{y,H}(s) > 0$.

Conversely, if s satisfies (1)–(3), then $u = z_{y,H}(s) \in S_{F^{y,H}}^c$.

Thus, by plotting $z_{y,H}(s)$ for $s \in \mathbb{R}$, the range of values where it is increasing yields $S^c_{F^{y,H}}$ (see Fig. 6.1).

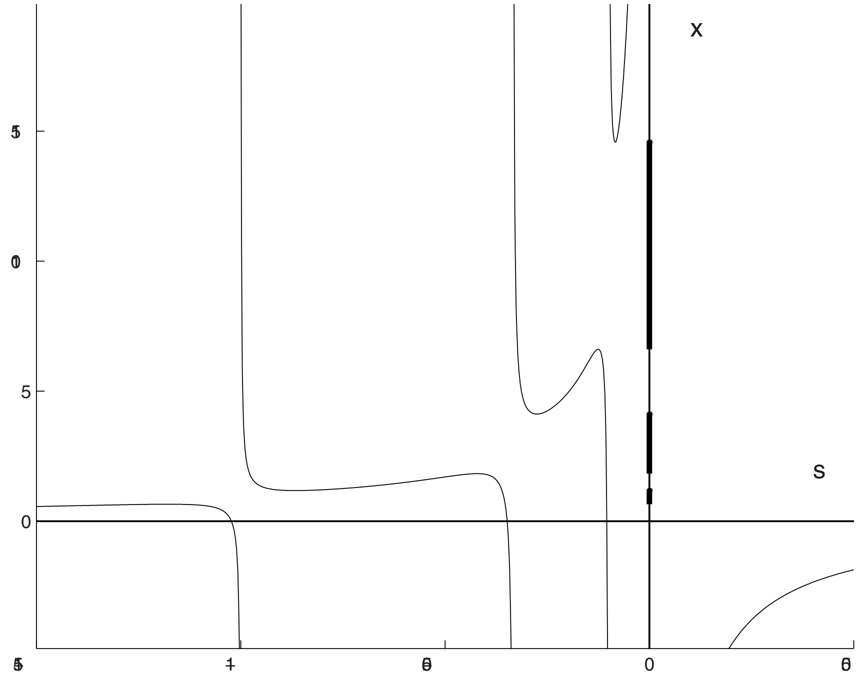

Fig. 6.1 The function $z_{0.1,H}(s)$ for a three-mass point H placing masses 0.2, 0.4, and 0.4 at three points $1, 3, 10$. The intervals of bold lines on the vertical axes are the support of $F^{0.1,H}$.

Of course, the supports of F and $F^{y,H}$ are identical on $\mathbb{R}^+$. The density function of $F^{0.1,H}$ is given in Fig. 6.2.

As for whether F places any mass at 0, it is shown in Silverstein and Choi [267] that

$$F^{y,H}(0) = \max(0, 1 - y[1 - H(0)]),$$

which implies

$$F(0) = \begin{cases} H(0), & y[1 - H(0)] \leq 1, \\ 1 - y^{-1}, & y[1 - H(0)] > 1. \end{cases} \tag{6.1.7}$$

It is appropriate at this time to state a lemma that lists all the ways intervals in $S^c_{F^{y,H}}$ can arise with respect to the graph of $z_{y,H}(s)$, $s \in \mathbb{R}$. It also states the dependence of these intervals on y. The proof will be given in later sections.

Lemma 6.2. (*a*) *If* (t_1, t_2) *is contained in* S^c_H *with* $t_1, t_2 \in \partial S_H$ *and* $t_1 > 0$, *then there is a* $y_0 > 0$ *for which* $y < y_0 \Rightarrow$ *there are two values* $s^1_y < s^2_y$ *in*

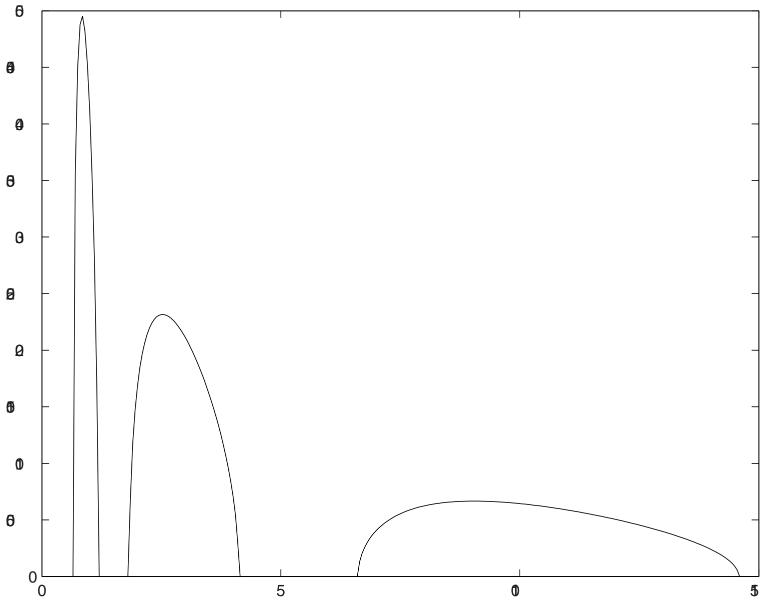

Fig. 6.2 The density function of $F^{0.1,H}$ with H defined in Fig. 6.1

$[-t_1^{-1}, -t_2^{-1}]$ *for which* $(z_{y,H}(s_y^1), z_{y,H}(s_y^2)) \subset S^c_{F^{y,H}}$, *with endpoints lying in* $\partial S_{F^{y,H}}$ *and* $z_{y,H}(s_y^1) > 0$. *Moreover,*

$$z_{y,H}(s_y^i) \to t_i, \quad \text{as } y \to 0, \tag{6.1.8}$$

for $i = 1, 2$. *The endpoints vary continuously with* y, *shrinking down to a point as* $y \uparrow y_0$, *while* $z_{y,H}(s_y^2) - z_{y,H}(s_y^1)$ *is monotone in* y.

(In the graph of $z_{y,H}(s)$, s_y^1 and s_y^2 are the local minimizer and maximizer in the interval $(-1/t_1, -1/t_2)$, and $z_{y,H}(s_y^1)$, and $z_{y,H}(s_y^2)$ are the local minimum and maximum values. As an example, notice the minimizers and maximizers of the two curves in the middle of Fig. 6.1.)

(*b*) *If* $(t_3, \infty) \subset S_H^c$ *with* $0 < t_3 \in \partial S_H$, *then there exists* $s_y^3 \in [-1/t_3, 0)$ *such that* $z_{y,H}(s_y^3)$ *is the largest number in* $S_{F^{y,H}}$. *As* y *decreases from* ∞ *to* 0, *(6.1.8) holds for* $i = 3$ *with convergence monotone from* ∞ *to* t_3.

(The value s_y^3 is the rightmost minimizer of the graph $z_{y,H}(s), s < 0$, and $z_{y,H}(s_y^3)$ is the largest local minimum value. See the curve immediately to the left of the vertical axis in Fig. 6.1.)

(*c*) *If* $y[1 - H(0)] < 1$ *and* $(0, t_4) \subset S_H^c$ *with* $t_4 \in \partial S_H$, *then there exists* $s_y^4 \in (-\infty, -1/t_4]$ *such that* $z_{y,H}(s_y^4)$ *is the smallest positive number in* $S_{F^{y,H}}$, *and (6.1.8) holds with* $i = 4$, *the convergence being monotone from* 0 *as* y *decreases from* $[1 - H(0)]^{-1}$.

(The value s_y^4 is the leftmost local maximizer, and $z_{y,H}(s_y^4)$ is the smallest local maximum value; i.e., the smallest point of the support of $F^{y,H}$. See the leftmost curve in Fig. 6.1.)

(*d*) *If* $y[1-H(0)] > 1$, *then, regardless of the existence of* $(0, t_4) \subset S_H^c$, *there exists* $s_y > 0$ *such that* $z_{y,H}(s_y) > 0$ *and is the smallest number in* $S_{F^{y,H}}$. *It decreases from* ∞ *to* 0 *as* y *decreases from* ∞ *to* $[1-H(0)]^{-1}$.

(In this case, the curve in Fig. 6.1 should have a different shape. It will increase from $-\infty$ to the positive value $z_{y,H} > 0$ at s_y and then decrease to 0 as s increases from 0 to ∞.)

(*e*) *If* $H = I_{[0,\infty)}$ *(that is,* H *places all mass at* 0*), then* $F = F^{y,I_{[0,\infty)}} = I_{[0,\infty)}$.

All intervals in $S^c_{F^{y,H}} \cap [0,\infty)$ *arise from one of the above. Moreover, disjoint intervals in* S_H^c *yield disjoint intervals in* $S^c_{F^{y,H}}$.

Thus, for interval $[a, b] \subset S^c_{F^{y,H}} \cap \mathbb{R}^+$, it is possible for $s_{F^{y,H}}(a)$ to be positive. This will occur only in case (d) of Lemma 6.2 when $b < z_{y,H}(s_y)$. For any other location of $[a, b]$ in $\mathbb{R}^+$, it follows from Lemma 6.1 that $s_{F^{y,H}}$ is negative and

$$[-1/s_{F^{y,H}}(a), -1/s_{F^{y,H}}(b)] \tag{6.1.9}$$

is contained in S_H^c. This interval is the proper choice of J.

The main result can now be stated.

Theorem 6.3. *Assume the following.*

(*a*) *Assumptions in Theorem 5.10 hold:* x_{ij}, $i, j = 1, 2, ...$ *are iid random variables in* $\mathbb{C}$ *with* $\mathrm{E}x_{11} = 0$, $\mathrm{E}|x_{11}|^2 = 1$, *and* $\mathrm{E}|x_{11}|^4 < \infty$.

(*b*) $p = p(n)$ *with* $y_n = p/n \to y > 0$ *as* $p \to \infty$.

(*c*) *For each* n, $\mathbf{T} = \mathbf{T}_n$ *is a nonrandom* $p \times p$ *Hermitian nonnegative definite matrix satisfying* $H_n \equiv F^{\mathbf{T}_n} \xrightarrow{\mathscr{D}} H$, *a c.d.f.*

(*d*) $\|\mathbf{T}_n\|$, *the spectral norm of* $\mathbf{T}_n$ *is bounded in* n.

(*e*) $\mathbf{B}_n = (1/n)\mathbf{T}_n^{1/2}\mathbf{X}_n\mathbf{X}_n^*\mathbf{T}_n^{1/2}$, $\mathbf{T}_n^{1/2}$ *is any Hermitian square root of* $\mathbf{T}_n$, *and* $\underline{\mathbf{B}}_n = (1/n)\mathbf{X}_n^*\mathbf{T}_n\mathbf{X}_n$, *where* $\mathbf{X}_n = (x_{ij})$, $i = 1, 2, \cdots, p$, $j = 1, 2, \cdots, n$.

(*f*) *Interval* $[a, b]$ *with* $a > 0$ *lies in an open interval outside the support of* F^{y_n,H_n} *for all large* n.

Then:

(1) $\mathrm{P}($*no eigenvalues of* $\mathbf{B}_n$ *appear in* $[a, b]$ *for all large* $n) = 1$.

(2) *If* $y[1-H(0)] > 1$, *then* x_0, *the smallest value in the support of* $F^{y,H}$, *is positive, and with probability one* $\lambda_n^{\mathbf{B}_n} \to x_0$ *as* $n \to \infty$. *The number* x_0 *is the maximum value of the function* $z_{y,H}(s)$ *for* $s \in \mathbb{R}^+$.

(3) *If* $y[1-H(0)] \le 1$, *or* $y[1-H(0)] > 1$ *but* $[a, b]$ *is not contained in* $(0, x_0)$, *then by assumption* (*f*) *and Lemma 6.1, the interval (6.1.9) is contained in* $S^c_{H_n} \cap \mathbb{R}^+$ *for all large* n. *For these* n, *let* $i_n \ge 0$ *be such that*

$$\lambda_{i_n}^{\mathbf{T}_n} > -1/s_{F^{y,H}}(b) \qquad \text{and} \qquad \lambda_{i_n+1}^{\mathbf{T}_n} < -1/s_{F^{y,H}}(a). \tag{6.1.10}$$

Then

$$P\left(\lambda_{i_n}^{\mathbf{B}_n} > b \quad and \quad \lambda_{i_n+1}^{\mathbf{B}_n} < a \quad for\ all\ large\ n\right) = 1.$$

Remark 6.4. Conclusion (2) occurs when $n < p$ for large n, in which case $\lambda_{n+1}^{\mathbf{B}_n} = 0$. Therefore exact separation should not be expected to occur for $[a, b] \subset [0, x_0]$. Regardless of their values, the $p - n$ smallest eigenvalues of $\mathbf{T}_n$ are essentially being converted to zero by $\mathbf{B}_n$. It is worth noting that when $y[1 - H(0)] > 1$ and F and (consequently) H each have at least two nonconnected members in their support in $\mathbb{R}^+$, the number of eigenvalues of $\mathbf{B}_n$ and $\mathbf{T}_n$ will match up in each respective member **except** the leftmost member. Thus the conversion to zero is affecting only this member.

Remark 6.5. The assumption of nonrandomness of $\mathbf{T}_n$ is made only for convenience. Using Fubini's theorem, Theorem 6.3 can easily be extended to random $\mathbf{T}_n$ (independent of x_{ij}) as long as the limit H is nonrandom and assumption (f) is true almost surely. At present, it is unknown whether the boundedness of $\|\mathbf{T}_n\|$ can be relaxed.

Conclusion (1) and the results on the extreme eigenvalues of $(1/n)\mathbf{X}\mathbf{X}^*$ yield properties on the extreme eigenvalues of $\mathbf{B}_n$. Notice that the interval $[a, b]$ can also be unbounded; that is, $\limsup_n \|\mathbf{B}_n\|$ stays a.s. bounded (nonrandom bound). Also, when $p < n$ and $\lambda_p^{\mathbf{T}_n}$ is bounded away from 0 for all n, we can use

$$\lambda_p^{\mathbf{B}_n} \geq \lambda_p^{(1/n)\mathbf{X}\mathbf{X}^*} \lambda_p^{\mathbf{T}_n}$$

to conclude that a nonrandom $b > 0$ exists for which a.s. $\lambda_p^{\mathbf{B}_n} > b$. Therefore we have the following corollary.

Corollary 6.6. *If $\|\mathbf{T}_n\|$ converges to the largest number in the support of H, then $\|\mathbf{B}_n\|$ converges a.s. to the largest number in the support of F. If the smallest eigenvalue of $\mathbf{T}_n$ converges to the smallest number in the support of H and $y < 1$, then the smallest eigenvalue of $\mathbf{B}_n$ converges to the smallest number in the support of F.*

The proof of Theorem 6.3 will begin in Section 6.2 with the proof of (1). It is achieved by showing the convergence of Stieltjes transforms at an appropriate rate, uniform with respect to the real part of z over certain intervals, while the imaginary part of z converges to zero. Besides relying on standard results on matrices, the proof requires Lemmas 2.12 and 2.13 (bounds on moments of martingale difference sequences) as well as Lemma B.26 (an extension of Lemma 2.13 to random quadratic forms). Additional results are given in the next section. Subsection 6.2.2 establishes a rate of convergence of $F^{\underline{\mathbf{B}}_n}$, needed in proving the convergence of the Stieltjes transforms. The latter will be broken down into two parts (Subsections 6.2.3 and 6.2.4), while Subsection 6.2.5 completes the proof of (1).

Conclusion (2) is proven in Section 6.3 and conclusion (3) in Section 6.4 . Both rely on (1) and on the properties of the extreme eigenvalues of $(1/n)\mathbf{X}_n\mathbf{X}_n^*$. The proof of (3) involves systematically increasing the number of columns of $\mathbf{X}_n$, while keeping track of the movements of the eigenvalues of the new matrices, until the limiting y is sufficiently small that the result obtained at the beginning of the introduction can be applied.

Along the way to proving (2) and (3), Lemma 6.2 will be proven (in Section 6.3 and Subsection 6.4.3).

6.1.1 Mathematical Tools

We list in this section additional results needed to prove Theorem 6.3. Throughout the rest of this chapter, constants appearing in inequalities are represented by K and occasionally subscripted with variables they depend on. They are nonrandom and may take on different values from one appearance to the next.

The first lemma can be found in most probability textbooks, see, e.g., Chung [79].

Lemma 6.7. (Kolmogorov's inequality for submartingales). *If $X_1, \ldots, X_m$ is a submartingale, then, for any $\alpha > 0$,*

$$P\left(\max_{k\le m} X_k \ge \alpha\right) \le \frac{1}{\alpha}\mathrm{E}(|X_m|).$$

Lemma 6.8. *If, for all $t > 0$, $\mathrm{P}(|X| > t)t^p \le K$ for some positive p, then for any positive $q < p$,*

$$\mathrm{E}|X|^q \le K^{q/p}\left(\frac{p}{p-q}\right).$$

Proof. For any $a > 0$, we have

$$\mathrm{E}|X|^q = \int_0^\infty \mathrm{P}(|X|^q > t)dt \le a + K\int_a^\infty t^{-p/q}dt = a + K\frac{q}{p-q}a^{1-p/q}.$$

By differentiating the last expression with respect to a and setting to zero, we find its minimum occurs when $a = K^{q/p}$. Plugging this value into the last expression gives us the result.

Lemma 6.9. *Let $z \in \mathbb{C}^+$ with $v = \Im z$, $\mathbf{A}$ and $\mathbf{B}$ $n \times n$ with $\mathbf{B}$ Hermitian, and $\mathbf{r} \in \mathbb{C}^n$. Then*

$$\left|\mathrm{tr}\left((\mathbf{B}-z\mathbf{I})^{-1}-(\mathbf{B}+\mathbf{r}\mathbf{r}^*-z\mathbf{I})^{-1}\right)\mathbf{A}\right| = \left|\frac{\mathbf{r}^*(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{A}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}}{1+\mathbf{r}^*(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}}\right| \le \frac{\|\mathbf{A}\|}{v}.$$

Proof. Since $(\mathbf{B}-z\mathbf{I})^{-1}-(\mathbf{B}+\mathbf{r}\mathbf{r}^*-z\mathbf{I})^{-1} = (\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}\mathbf{r}^*(\mathbf{B}+\mathbf{r}\mathbf{r}^*-z\mathbf{I})^{-1}$, similar to (4.5.11), we use the identity

$$\mathbf{r}^*(\mathbf{C}+\mathbf{r}\mathbf{r}^*)^{-1} = \frac{1}{1+\mathbf{r}^*\mathbf{C}^{-1}\mathbf{r}}\mathbf{r}^*\mathbf{C}^{-1}, \tag{6.1.11}$$

valid for any square $\mathbf{C}$ for which $\mathbf{C}$ and $\mathbf{C}+\mathbf{r}\mathbf{r}^*$ are invertible, to get

$$\left|\mathrm{tr}\big((\mathbf{B}-z\mathbf{I})^{-1}-(\mathbf{B}+\mathbf{r}\mathbf{r}^*-z\mathbf{I})^{-1}\big)\mathbf{A}\right| = \left|\frac{\mathrm{tr}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}\mathbf{r}^*(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{A}}{1+\mathbf{r}^*(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}}\right|$$
$$= \left|\frac{\mathbf{r}^*(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{A}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}}{1+\mathbf{r}^*(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}}\right| \le \|\mathbf{A}\|\,\frac{\|(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}\|^2}{|1+\mathbf{r}^*(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}|}.$$

Write $\mathbf{B}=\sum\lambda_i^{\mathbf{B}}\mathbf{e}_i\mathbf{e}_i^*$, where the $\mathbf{e}_i$'s are the orthonormal eigenvectors of $\mathbf{B}$. Then

$$\|(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{q}\|^2 = \sum\frac{|\mathbf{e}_i^*\mathbf{q}|^2}{|\lambda_i^{\mathbf{B}}-z|^2},$$

and

$$|1+\mathbf{r}^*(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}| \ge \Im\,\mathbf{r}^*(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r} = v\sum\frac{|\mathbf{e}_i^*\mathbf{q}|^2}{|\lambda_i^{\mathbf{B}}-z|^2}.$$

The result follows.

Lemma 6.10. *For $z=u+iv\in\mathbb{C}^+$, let $s_1(z)$, $s_2(z)$ be Stieltjes transforms of any two c.d.f.s, $\mathbf{A}$ and $\mathbf{B}$ $n\times n$ with $\mathbf{A}$ Hermitian nonnegative definite, and $\mathbf{r}\in\mathbb{C}^n$. Then*

(a) $\|(s_1(z)\mathbf{A}+\mathbf{I})^{-1}\| \le \max(4\|\mathbf{A}\|/v, 2)$

(b) $|\mathrm{tr}\mathbf{B}((s_1(z)\mathbf{A}+\mathbf{I})^{-1}-(s_2(z)\mathbf{A}+\mathbf{I})^{-1}|$
$\le |s_2(z)-s_1(z)|n\|\mathbf{B}\|\,\|\mathbf{A}\|(\max(4\|\mathbf{A}\|/v,2))^2$

(c) $|\mathbf{r}^*\mathbf{B}(s_1(z)\mathbf{A}+\mathbf{I})^{-1}\mathbf{r}-\mathbf{r}^*\mathbf{B}(s_2(z)\mathbf{A}+\mathbf{I})^{-1}\mathbf{r}|$
$\le |s_2(z)-s_1(z)|\|\mathbf{r}\|^2\|\mathbf{B}\|\|\mathbf{A}\|(\max(4\|\mathbf{A}\|/v,2))^2$

($\|\mathbf{r}\|$ *denoting the Euclidean norm on* $\mathbf{r}$).

Proof. Notice (b) and (c) follow easily from (a) using basic matrix properties. Using the Cauchy-Schwarz inequality, it is easy to show $|\Re\, s_1(z)| \le (\Im\, s_1(z)/v)^{1/2}$. Then, for any positive x,

$$\begin{aligned}|s_1(z)x+1|^2 &= (\Re\, s_1(z)x+1)^2+(\Im\, s_1(z))^2x^2\\ &\ge (\Re\, s_1(z)x+1)^2+(\Re\, s_1(z))^4v^2x^2\\ &\ge \min\left(\frac{1}{4},\ \frac{v^2}{16x^2}\right),\end{aligned}$$

where the last inequality follows by considering the two cases where $|\Re s_1(z)x| < \frac{1}{2}$ and otherwise. From this inequality, conclusion (a) follows.

Lemma 6.11. *Let $\{\mathcal{F}_n\}$ be an increasing sequence of σ-fields and $\{X_n\}$ a sequence of random variables. Write $\mathrm{E}_k=\mathrm{E}(\cdot|\mathcal{F}_k)$, $\mathrm{E}_\infty=\mathrm{E}(\cdot|\mathcal{F}_\infty)$, $\mathcal{F}_\infty\equiv\bigvee_j\mathcal{F}_j$. If $X_n\to 0$ a.s. and $\sup_n|X_n|$ is integrable, then*

$$\lim_{n\to\infty}\max_{k\le n} \mathrm{E}_k X_n = 0, \quad a.s.$$

Proof. Write, for integer $m \geq 1$, $Y_m = \sup_{p\geq m}|X_p|$. We have Y_m integrable for all m, bounded in absolute value by $\sup_n |X_n|$. We will use the fact that for integrable Y

$$\lim_{n\to\infty} \mathrm{E}_n Y = \mathrm{E}_\infty Y, \quad \text{a.s.}$$

(Theorem 9.4.8 of Chung [79]). Then, by the dominated convergence theorem, for any m and positive integer K

$$Z \equiv \limsup_n \max_{k\le n} \mathrm{E}_k|X_n| \le \limsup_n \left[\sum_{k\le K} \mathrm{E}_k|X_n| + \max_{K\le k\le n} \mathrm{E}_k|X_n|\right]$$

$$\le \limsup_n \max_{K\le k\le n} \mathrm{E}_k Y_m = \sup_{K\le k<\infty} \mathrm{E}_k Y_m \quad \text{a.s.}$$

Therefore

$$Z \le \limsup_K \mathrm{E}_K Y_m = \lim_{K\to\infty} \mathrm{E}_K Y_m = \mathrm{E}_\infty Y_m \quad \text{a.s.}$$

Since $Y_m \to 0$ a.s. as $m \to \infty$, we get from the dominated convergence theorem $\mathrm{E}_\infty Y_m \to 0$ a.s. Therefore we must have $Z = 0$ a.s. The result follows.

Basic properties on matrices will be used throughout this chapter, the two most common being $\mathrm{tr}\mathbf{AB} \le \|\mathbf{A}\|\mathrm{tr}\mathbf{B}$ for Hermitian nonnegative definite $\mathbf{A}$ and $\mathbf{B}$ and (6.1.11).

6.2 Proof of (1)

6.2.1 Truncation and Some Simple Facts

We begin by simplifying our assumptions. Because of assumption (d) in the theorem we can assume $\|\mathbf{T}_n\| \le 1$.

For $C > 0$, let $y_{ij} = x_{ij}I_{[|x_{ij}|\le C]} - \mathrm{E}x_{ij}I_{[|x_{ij}|\le C]}$, $\mathbf{Y} = (y_{ij})$ and $\tilde{\mathbf{B}}_n = (1/n)\mathbf{T}_n^{1/2}\mathbf{Y}_n\mathbf{Y}_n^*\mathbf{T}_n^{1/2}$. Denote the eigenvalues of $\mathbf{B}_n$ and $\tilde{\mathbf{B}}_n$ by λ_k and $\tilde{\lambda}_k$ (in decreasing order). Since these are the squares of the k-th largest singular values of $(1/\sqrt{n})\mathbf{T}_n\mathbf{X}_n$ and $(1/\sqrt{n})\mathbf{T}_n\mathbf{Y}_n$ (respectively), we find using Theorem A.46 that

$$\max_{k\le n} |\lambda_k^{1/2} - \tilde{\lambda}_k^{1/2}| \le (1/\sqrt{n})\|\mathbf{X}_n - \mathbf{Y}_n\|.$$

Since $x_{ij} - y_{ij} = x_{ij} I_{[|x_{ij}|>C]} - \mathrm{E} x_{ij} I_{[|x_{ij}|>C]}$, from Theorem 5.8 we have with probability 1 that

$$\limsup_{n\to\infty} \max_{k\le n} |\lambda_k^{1/2} - \tilde{\lambda}_k^{1/2}| \le (1+\sqrt{y})\mathrm{E}^{1/2}|x_{11}|^2 I_{[|x_{11}|>C]}.$$

Because of assumption (a), we can make the bound above arbitrarily small by choosing C sufficiently large. Thus, in proving Theorem 6.3, it is enough to consider the case where the underlying variables are uniformly bounded.

In this case, the conditions in Theorem 5.9 are met. It follows then that $\lambda_{\max}$, the largest eigenvalue of $\mathbf{B}_n$, satisfies

$$P(\lambda_{\max} \ge K) = o(n^{-t}) \tag{6.2.1}$$

for any $K > (1+\sqrt{y})^2$ and any positive t.

Also, since, for square $\mathbf{A}$, $\mathrm{tr}(\mathbf{A}\mathbf{A}^*)^\ell \le (\mathrm{tr}\mathbf{A}\mathbf{A}^*)^\ell$, we get from Lemma B.26 for any $\ell \ge 1$ when x_{11} is bounded

$$\mathrm{E}|\mathbf{x}_1^*\mathbf{A}\mathbf{x}_1 - \mathrm{tr}\mathbf{A}|^{2\ell} \le K_\ell(\mathrm{tr}\mathbf{A}\mathbf{A}^*)^\ell \tag{6.2.2}$$

($\mathbf{x}_j$ denoting the j-th column of $\mathbf{X}$), where K_ℓ also depends on the bound of x_{11}. From (6.2.2), we easily get

$$\mathrm{E}|\mathbf{x}_1^*\mathbf{A}\mathbf{x}_1|^{2\ell} \le K_\ell((\mathrm{tr}\mathbf{A}\mathbf{A}^*)^\ell + |\mathrm{tr}\mathbf{A}|^{2\ell}). \tag{6.2.3}$$

6.2.2 A Preliminary Convergence Rate

After truncation, no assumptions need to be made on the relationship between the $\mathbf{X}_n$'s for different n (that is, the entries of $\mathbf{X}_n$ need not come from the same doubly infinite array). Also, variable $z = u+iv$ will be the argument of any Stieltjes transform.

Let $s_n = s_n(z) = s_{F^{\mathbf{B}_n}}$ and $\underline{s}_n = \underline{s}_n(z) = s_{F^{\underline{\mathbf{B}}_n}}$. For $j = 1, 2, \cdots, n$, let $\mathbf{q}_j = (1/\sqrt{p})\mathbf{x}_j$, $\mathbf{r}_j = (1/\sqrt{n})\mathbf{T}_n^{1/2}\mathbf{x}_j$, and $\mathbf{B}_{(j)} = \mathbf{B}_{(j)}^n = \mathbf{B}_n - \mathbf{r}_j\mathbf{r}_j^*$. Let $y_n = p/n$.

Write

$$\mathbf{B}_n - z\mathbf{I} + z\mathbf{I} = \sum_{j=1}^n \mathbf{r}_j\mathbf{r}_j^*.$$

Taking the inverse of $\mathbf{B}_n - z\mathbf{I}$ on the right on both sides and using (6.1.11), we find

$$\mathbf{I} + z(\mathbf{B}_n - z\mathbf{I})^{-1} = \sum_{j=1}^n \frac{1}{1+\mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j}\mathbf{r}_j\mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}.$$

Taking the trace on both sides and dividing by n, we have

$$y_n + zy_n s_n = \frac{1}{n}\sum_{j=1}^{n} \frac{\mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j}{1 + \mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j} = 1 - \frac{1}{n}\sum_{j=1}^{n} \frac{1}{1 + \mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j}.$$

From (6.1.3), we see that

$$\underline{s}_n = -\frac{1}{n}\sum_{j=1}^{n} \frac{1}{z(1 + \mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j)}. \tag{6.2.4}$$

For each j, we have

$$\begin{aligned}\Im\, \mathbf{r}_j^*((1/z)\mathbf{B}_{(j)} - \mathbf{I})^{-1}\mathbf{r}_j &= \frac{1}{2i}\mathbf{r}_j^*\big(((1/z)\mathbf{B}_{(j)} - \mathbf{I})^{-1} - ((1/\overline{z})\mathbf{B}_{(j)} - \mathbf{I})^{-1}\big)\mathbf{r}_j \\ &= \frac{v}{|z|^2}\mathbf{r}_j^*((1/z)\mathbf{B}_{(j)} - \mathbf{I})^{-1}\mathbf{B}_{(j)}((1/\overline{z})\mathbf{B}_{(j)} - \mathbf{I})^{-1}\mathbf{r}_j \\ &\geq 0.\end{aligned}$$

Therefore

$$\frac{1}{|z(1 + \mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j)|} \leq \frac{1}{v}. \tag{6.2.5}$$

Write $\mathbf{B}_n - z\mathbf{I} - \big(-z\underline{s}_n\mathbf{T}_n - z\mathbf{I}\big) = \sum_{j=1}^n \mathbf{r}_j\mathbf{r}_j^* - (-z\underline{s}_n)\mathbf{T}_n$. Taking inverses and using (6.1.11) and (6.2.4), we have

$$\begin{aligned}&(-z\underline{s}_n\mathbf{T}_n - z\mathbf{I})^{-1} - (\mathbf{B}_n - z\mathbf{I})^{-1} \\ &= (-z\underline{s}_n\mathbf{T}_n - z\mathbf{I})^{-1}\left[\sum_{j=1}^{n}\mathbf{r}_j\mathbf{r}_j^* - (-z\underline{s}_n)\mathbf{T}_n\right](\mathbf{B}_n - z\mathbf{I})^{-1} \\ &= \sum_{j=1}^{n}\frac{-1}{z(1+\mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j)}\Big[(\underline{s}_n\mathbf{T}_n + \mathbf{I})^{-1}\mathbf{r}_j\mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1} \\ &\qquad - \frac{1}{n}(\underline{s}_n\mathbf{T}_n + \mathbf{I})^{-1}\mathbf{T}_n(\mathbf{B}_n - z\mathbf{I})^{-1}\Big].\end{aligned} \tag{6.2.6}$$

Taking the trace and dividing by p, we find

$$\begin{aligned}w_n = w_n(z) &= \frac{1}{p}\mathrm{tr}(-z\underline{s}_n\mathbf{T}_n - z\mathbf{I})^{-1} - s_n \\ &= \frac{1}{n}\sum_{j=1}^{n}\frac{-1}{z(1 + \mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j)}d_j,\end{aligned} \tag{6.2.7}$$

where

$$d_j = \mathbf{q}_j^* \mathbf{T}_n^{1/2} (\mathbf{B}_{(j)} - z\mathbf{I})^{-1} (\underline{s}_n \mathbf{T}_n + \mathbf{I})^{-1} \mathbf{T}_n^{1/2} \mathbf{q}_j \\ -(1/p)\mathrm{tr}(\underline{s}_n \mathbf{T}_n + \mathbf{I})^{-1} \mathbf{T}_n (\mathbf{B}_n - z\mathbf{I})^{-1}.$$

Since it has been shown in the proof of Theorem 4.1 that the LSD of $\mathbf{B}_n$ depends only on y and H, (6.1.4) and hence (6.1.5) will follow if one proves $w_n \to 0$. In the next step, we will do more than this. We will prove, for $v = v_n \geq n^{-1/17}$ and for any n, $z = u + iv_n$-values whose real parts are collected as the set $S_n \subset (-\infty, \infty)$,

$$\max_{u \in S_n} \frac{|w_n|}{v_n^5} \to 0, \text{ a.s.} \tag{6.2.8}$$

Write, for each $j \leq n$, $d_j = d_j^1 + d_j^2 + d_j^3 + d_j^4$, where

$$\begin{aligned} d_j^1 &= \mathbf{q}_j^* \mathbf{T}_n^{1/2} (\mathbf{B}_{(j)} - z\mathbf{I})^{-1} (\underline{s}_n \mathbf{T}_n + \mathbf{I})^{-1} \mathbf{T}_n^{1/2} \mathbf{q}_j \\ &\quad - \mathbf{q}_j^* \mathbf{T}_n^{1/2} (\mathbf{B}_{(j)} - z\mathbf{I})^{-1} (\underline{s}_{(j)} \mathbf{T}_n + \mathbf{I})^{-1} \mathbf{T}_n^{1/2} \mathbf{q}_j, \\ d_j^2 &= \mathbf{q}_j^* \mathbf{T}_n^{1/2} (\mathbf{B}_{(j)} - z\mathbf{I})^{-1} (\underline{s}_{(j)} \mathbf{T}_n + \mathbf{I})^{-1} \mathbf{T}_n^{1/2} \mathbf{q}_j \\ &\quad - (1/p)\mathrm{tr}(\underline{s}_{(j)} \mathbf{T}_n + \mathbf{I})^{-1} \mathbf{T}_n (B_{(j)} - z\mathbf{I})^{-1}, \\ d_j^3 &= (1/p)\mathrm{tr}(\underline{s}_{(j)} \mathbf{T}_n + \mathbf{I})^{-1} \mathbf{T}_n (\mathbf{B}_{(j)} - z\mathbf{I})^{-1} \\ &\quad - (1/p)\mathrm{tr}(\underline{s}_{(j)} \mathbf{T}_n + \mathbf{I})^{-1} \mathbf{T}_n (\mathbf{B}_n - z\mathbf{I})^{-1}, \\ d_j^4 &= (1/p)\mathrm{tr}(\underline{s}_{(j)} \mathbf{T}_n + \mathbf{I})^{-1} \mathbf{T}_n (\mathbf{B}_n - z\mathbf{I})^{-1} \\ &\quad - (1/p)\mathrm{tr}(\underline{s}_n \mathbf{T}_n + \mathbf{I})^{-1} \mathbf{T}_n (\mathbf{B}_n - z\mathbf{I})^{-1}, \end{aligned}$$

and let $\underline{s}_{(j)} = -\frac{(1-y_n)}{z} + y_n s_{F^{\mathbf{B}_{(j)}}}(z)$. From Lemma 6.9, we have

$$\max_{j \leq n} |\underline{s}_n - \underline{s}_{(j)}| \leq \frac{1}{nv}. \tag{6.2.9}$$

Moreover, it is easy to verify that $\underline{s}_{(j)}$ is the Stieltjes transform of a c.d.f., so that $|\underline{s}_{(j)}| \leq v_n^{-1}$.

In view of (6.2.5), to prove (6.2.8), it is sufficient to show the a.s. convergence of

$$\max_{j \leq n,\ u \in S_n} \frac{|d_j^i|}{v_n^6} \tag{6.2.10}$$

to zero for $i = 1, 2, 3, 4$.

Using $\|(\mathbf{A} - z\mathbf{I})^{-1}\| \leq 1/v_n$ for any Hermitian matrix $\mathbf{A}$, we get from Lemma 6.10 (c) and (6.2.9)

$$|d_j^1| \leq 16 \frac{\|\mathbf{x}_j\|^2}{p} \frac{1}{nv_n^4}.$$

Using (6.2.2), it follows that, for any $t \geq 1$, we have for all n sufficiently large

$$\mathrm{P}\left(\max_{j\leq n,\ u\in S_n} \frac{|d_j^1|}{v_n^6} > v_n\right) \leq p\mathrm{P}\left(\max_{j\leq n} \tfrac{1}{p}\|\mathbf{x}_j\|^2 > \tfrac{1}{16}nv_n^{11}\right)$$
$$\leq K_t \frac{pn}{(nv_n^{11})^t}.$$

The last bound is summable when $t > 17/2$, so we have (6.2.10)$\xrightarrow{\text{a.s.}} 0$ when $i = 1$.

Using Lemmas 6.9 and 6.10 (a), we find

$$v_n^{-6}|d_j^3| \leq \frac{4}{pv_n^8},$$

so that (6.2.10) $\to 0$ for $i = 3$.

We get from Lemma 6.10 (b) and (6.2.9)

$$v_n^{-6}|d_j^4| \leq 16\frac{1}{nv_n^{10}},$$

so that (6.2.10) $\to 0$ for $i = 4$.

Using (6.2.2), we find, for any $t \geq 1$,

$$\begin{aligned}
\mathrm{E}|v_n^{-6}d_j^2|^{2t} &\leq \frac{K_t}{v_n^{12t}p^{2t}}(\mathrm{tr}\mathbf{T}_n^{1/2}(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}(\underline{s}_{(j)}\mathbf{T}_n + \mathbf{I})^{-1}\mathbf{T}_n \\
&\quad (\overline{\underline{s}}_{(j)}\mathbf{T}_n + \mathbf{I})^{-1}(\mathbf{B}_{(j)} - \overline{z}\mathbf{I})^{-1}\mathbf{T}_n^{1/2})^t \\
&= \frac{K_t}{v_n^{12t}p^{2t}}(\mathrm{tr}(\underline{s}_{(j)}\mathbf{T}_n + \mathbf{I})^{-1}\mathbf{T}_n(\overline{\underline{s}}_{(j)}\mathbf{T}_n + \mathbf{I})^{-1} \\
&\quad (\mathbf{B}_{(j)} - \overline{z}\mathbf{I})^{-1}\mathbf{T}_n(\mathbf{B}_{(j)} - z\mathbf{I})^{-1})^t \\
&\quad \text{(using Lemma 6.10 a))} \\
&\leq \frac{K_t}{v_n^{12t}p^{2t}}\frac{1}{v_n^{2t}}(\mathrm{tr}(\mathbf{B}_{(j)} - \overline{z}\mathbf{I})^{-1}\mathbf{T}_n(\mathbf{B}_{(j)} - z\mathbf{I})^{-1})^t \\
&= \frac{K_t}{(pv_n^7)^{2t}}(\mathrm{tr}\mathbf{T}_n(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}(\mathbf{B}_{(j)} - \overline{z}\mathbf{I})^{-1})^t \\
&\leq \frac{K_t}{(pv_n^7)^{2t}}(p/v_n^2)^t \\
&= \frac{K_t}{(pv_n^{16})^t}.
\end{aligned}$$

We have then, for any $\varepsilon > 0$ and $t \geq 1$,

$$\mathrm{P}\left(\max_{j\leq n,\ u\in S_n} |v_n^{-6}d_j^2| > \varepsilon\right) \leq K_t\frac{1}{\varepsilon^{2t}}\frac{pn}{(pv_n^{16})^t},$$

which implies that (6.2.10) $\xrightarrow{\text{a.s.}} 0$ for $i = 2$ by taking $t > 51$. Obviously, the estimation above remains if d_j^2 is replaced by d_j^i for all $i = 1, 3, 4$.

Thus we have shown, when $v_n \geq n^{-1/17}$, for any positive t and all $\varepsilon > 0$,

$$\mathrm{P}\left(\max_{u \in S_n} |w_n(z)| v_n^{-5} > \varepsilon \right) \leq K_t \varepsilon^{-2t} n^{2-t/17}. \tag{6.2.11}$$

Therefore, $\max_{u \in S_n} |w_n(z)| v_n^{-5} \xrightarrow{\text{a.s.}} 0$ by choosing $t > 51$.

Moreover, for any $\varepsilon > 0$, replacing ε in (6.2.11) by ε/μ_n, we obtain

$$\mathrm{P}\left(\mu_n \max_{u \in S_n} |w_n(z)| v_n^{-5} > \varepsilon \right) \leq K_t \varepsilon^{-2t} n^{2-t/34}, \tag{6.2.12}$$

where $\mu_n = n^{1/68}$ and $v = v_n = n^{-\delta}$ with $\delta \leq 1/17$.

We now rewrite w_n totally in terms of $\underline{s}_n$. Using identity (6.1.3), we have

$$\begin{aligned} w_n &= \frac{1}{y_n}\left(-\frac{y_n}{z} \int \frac{dH_n(t)}{1 + t\underline{s}_n} - \underline{s}_n - \frac{(1 - y_n)}{z} \right) \\ &= \frac{\underline{s}_n}{y_n z}\left(-\frac{y_n}{\underline{s}_n} \int \frac{dH_n(t)}{1 + t\underline{s}_n} - z - \frac{(1 - y_n)}{\underline{s}_n} \right) \\ &= \frac{\underline{s}_n}{y_n z}\left(-z - \frac{1}{\underline{s}_n} + y_n \int \frac{t\,dH_n(t)}{1 + t\underline{s}_n} \right). \end{aligned}$$

Let

$$R_n = -z - \frac{1}{\underline{s}_n} + y_n \int \frac{t\,dH_n(t)}{1 + t\underline{s}_n}. \tag{6.2.13}$$

Then $R_n = w_n z y_n / \underline{s}_n$.

Returning now to F^{y_n, H_n} and $F^{y,H}$, let $\underline{s}_n^0 = s_{F^{y_n,H_n}}$ and $\underline{s}^0 = s_{F^{y,H}}$. Then $\underline{s}^0$ solves (6.1.5), its inverse is given by (6.1.6),

$$\underline{s}_n^0 = \frac{1}{-z + y_n \int \frac{t\,dH_n(t)}{1 + t\underline{s}_n^0}}, \tag{6.2.14}$$

and the inverse of $\underline{s}_n^0$, denoted z_n^0, is given by

$$z_n^0(\underline{s}_n) = -\frac{1}{\underline{s}_n} + y_n \int \frac{t\,dH_n(t)}{1 + t\underline{s}_n}. \tag{6.2.15}$$

From (6.2.15) and the inversion formula for Stieltjes transforms, it is obvious that $F^{y_n,H_n} \xrightarrow{\mathscr{D}} F^{y,H}$ as $n \to \infty$. Therefore, from assumption (f), an $\underline{\varepsilon} > 0$ exists for which $[a - 2\underline{\varepsilon}, b + 2\underline{\varepsilon}]$ also satisfies (f). This interval will stay uniformly bounded away from the boundary of the support of F^{y_n,H_n} for all large n, so that for these n both $\sup_{u \in [a-2\underline{\varepsilon}, b+2\underline{\varepsilon}]} \frac{d}{du} \underline{s}_n^0(u)$ is bounded and $-1/\underline{s}_n^0(u)$ for $u \in [a - 2\underline{\varepsilon}, b + 2\underline{\varepsilon}]$ stays uniformly away from the support of H_n. Therefore, for all n sufficiently large,

$$\sup_{u\in[a-2\underline{\varepsilon},b+2\underline{\varepsilon}]}\left(\frac{d}{du}\underline{s}_n^0(u)\right)\int\frac{t^2\,dH_n(t)}{(1+t\underline{s}_n^0(u))^2}\le K. \tag{6.2.16}$$

Let $a' = a - \underline{\varepsilon}$, $b' = b + \underline{\varepsilon}$. On either $(-\infty, a']$ or $[b', \infty)$, each collection of functions in λ, $\{(\lambda - u)^{-1} : u \in [a, b]\}$, $\{(\lambda - u)^{-2} : u \in [a, b]\}$, forms a uniformly bounded, equicontinuous family. It is straightforward then to show

$$\lim_{n\to\infty}\sup_{u\in[a,b]}|\underline{s}_n^0(u)-\underline{s}^0(u)|=0 \tag{6.2.17}$$

and

$$\lim_{n\to\infty}\sup_{u\in[a,b]}\left|\frac{d}{du}\underline{s}_n^0(u)-\frac{d}{du}\underline{s}^0(u)\right|=0 \tag{6.2.18}$$

(see, e.g., Billingsley [57], problem 8, p. 17). Since, for all $u \in [a, b]$, $\lambda \in [a', b']^c$, and positive v,

$$\left|\frac{1}{\lambda-(u+iv)}-\frac{1}{\lambda-u}\right|<\frac{v}{\underline{\varepsilon}^2},$$

we have, for any sequence of positive v_n converging to 0,

$$\lim_{n\to\infty}\sup_{u\in[a,b]}|\underline{s}_n^0(u+iv_n)-\underline{s}_n^0(u)|=0. \tag{6.2.19}$$

Similarly

$$\lim_{n\to\infty}\sup_{u\in[a,b]}\left|\frac{\Im\,\underline{s}_n^0(u+iv_n)}{v_n}-\frac{d}{du}\underline{s}_n^0(u)\right|=0. \tag{6.2.20}$$

Expressions (6.2.16), (6.2.17), (6.2.19), and (6.2.20) will be needed in the latter part of Subsection 6.2.4.

Let $\underline{s}_2^0 = \Im\,\underline{s}_n^0$. From (6.2.14), we have

$$\underline{s}_2^0=\frac{v_n+\underline{s}_2^0 y_n\int\frac{t^2\,dH_n(t)}{|1+t\underline{s}_n^0|^2}}{\left|-z+y_n\int\frac{t\,dH_n(t)}{1+t\underline{s}_n^0}\right|^2}. \tag{6.2.21}$$

For any real u, by Lemma 6.10 a),

$$\underline{s}_2^0 y_n\int\frac{t^2\,dH_n(t)}{|1+t\underline{s}_n^0|^2}=y_n\Im\left(\int\frac{t\,dH_n(t)}{1+t\underline{s}_n^0}\right)$$

$$\le y_n||\mathbf{T}_n(\mathbf{I}+\mathbf{T}_n\underline{s}_n^0)^{-1}||\le 4y_n/v_n.$$

Applying $\sqrt{1-a}\le 1-\frac{1}{2}a$ for $a \le 1$, it follows that

$$\left(\frac{\underline{s}_2^0 y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n^0|^2}}{v_n + \underline{s}_2^0 y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n^0|^2}} \right)^{1/2} < 1 - K v_n^2 \tag{6.2.22}$$

for some positive constant K.

Let $\underline{s}_n = \underline{s}_{n1} + i\underline{s}_{n2}$, where $\underline{s}_{n1} = \Re\, \underline{s}_n$, $\underline{s}_{n2} = \Im\, \underline{s}_n$. We have $\underline{s}_n$ satisfying

$$\underline{s}_n = \frac{1}{-z + y_n \int \frac{t\, dH_n(t)}{1+t\underline{s}_n} - R_n} \tag{6.2.23}$$

and

$$\underline{s}_{n2} = \frac{v_n + \underline{s}_{n2} y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n|^2} + \Im\, R_n}{\left| -z + y_n \int \frac{t\, dH_n(t)}{1+t\underline{s}_n} - R_n \right|^2}. \tag{6.2.24}$$

From (6.2.14) and (6.2.23), we get

$$\underline{s}_n - \underline{s}_n^0 = \frac{(\underline{s}_n - \underline{s}_n^0) y_n \int \frac{t^2\, dH_n(t)}{(1+t\underline{s}_n)(1+t\underline{s}_n^0)}}{\left(-z + y_n \int \frac{t\, dH_n(t)}{1+t\underline{s}_n} - R_n \right)\left(-z + y_n \int \frac{t\, dH_n(t)}{1+t\underline{s}_n^0} \right)} + \underline{s}_n \underline{s}_n^0 R_n. \tag{6.2.25}$$

When $|\Im\, R_n| < v_n$, by the Cauchy-Schwarz inequality, (6.2.21), (6.2.22), and (6.2.24), we get

$$\begin{aligned}
&\left| \frac{y_n \int \frac{t^2\, dH_n(t)}{(1+t\underline{s}_n)(1+t\underline{s}_n^0)}}{\left(-z + y_n \int \frac{t\, dH_n(t)}{1+t\underline{s}_n} - R_n \right)\left(-z + y_n \int \frac{t\, dH_n(t)}{1+t\underline{s}_n^0} \right)} \right| \\
&\le \left(\frac{y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n|^2}}{\left| -z + y_n \int \frac{t\, dH_n(t)}{1+t\underline{s}_n} - R_n \right|^2} \right)^{1/2} \left(\frac{y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n^0|^2}}{\left| -z + y_n \int \frac{t\, dH_n(t)}{1+t\underline{s}_n^0} \right|^2} \right)^{1/2} \\
&= \left(\frac{\underline{s}_2 y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n|^2}}{v_n + \underline{s}_2 y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n|^2} + \Im\, R_n} \right)^{1/2} \left(\frac{\underline{s}_2^0 y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n^0|^2}}{v_n + \underline{s}_2^0 y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n^0|^2}} \right)^{1/2} \\
&\le \left(\frac{\underline{s}_2^0 y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n^0|^2}}{v_n + \underline{s}_2^0 y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n^0|^2}} \right)^{1/2} \\
&\le 1 - K v_n^2.
\end{aligned} \tag{6.2.26}$$

We claim that on the set $\{\lambda_{\max} \le K_1\}$, where $K_1 > (1+\sqrt{y})^2$, for all n sufficiently large, $|\underline{s}_n| \ge \frac{1}{2}\mu_n^{-1} v_n$ whenever $|u| \le \mu_n v_n^{-1}$. Indeed, when $u \le -v_n$ or $u \ge \lambda_{\max} + v_n$,

$$|\underline{s}_n| \geq |\Re \underline{s}_n| \geq \frac{K_1 + \mu_n v_n^{-1}}{(K_1 + \mu_n v_n^{-1})^2 + v_n^2} \geq \frac{1}{2\mu_n v_n^{-1}}$$

for large n. When $-v_n < u < \lambda_{\max} + v_n$,

$$|\underline{s}_n| \geq |\Im \underline{s}_n| \geq \frac{v_n}{(K_1 + v_n)^2 + v_n^2} \geq \mu_n^{-1} v_n$$

for large n. Thus the claim is proven.

Therefore, when $|u| \leq \mu_n v_n^{-1}$, $|w_n| \leq \mu_n^{-1} v_n^5$, and $\lambda_{\max} \leq K_1$, we have, for large n, $|z| \leq 2\mu_n v_n^{-1}$ and

$$|R_n| = |y_n z w_n / \underline{s}_n| \leq K \mu_n^2 v_n^{-2} |w_n| < v_n.$$

Consequently, by (6.2.25), (6.2.26), and the fact that $|z\underline{s}_n^0| \leq 1 + K/v_n$, for all large n, we have

$$\begin{aligned} |\underline{s}_n - \underline{s}_n^0| &\leq K v_n^{-2} |\underline{s}_n \underline{s}_n^0 R_n| \\ &= K v_n^{-2} |y_n z \underline{s}_n^0 w_n| \leq K' v_n^{-3} |w_n| \leq 3\mu_n^{-1} v_n. \end{aligned}$$

Furthermore, when $z = u + iv_n$ with $|u| \geq \mu_n v_n^{-1}$ and $\lambda_{\max} \leq K_1$, it is easy to verify that, for all large n, we still have

$$|\underline{s}_n - \underline{s}_n^0| \leq 3\mu_n^{-1} v_n.$$

Therefore, for large n, we have

$$\max_{u \in S_n} v_n^{-1} |\underline{s}_n - \underline{s}_n^0| \leq 3\mu_n^{-1} + 2v_n^{-2} \max_{u \in S_n} \left(I(|w_n| > \mu_n^{-1} v_n^5) + I(\lambda_{\max} > K_1) \right).$$

Thus, for these n and for any positive ε and t, from (6.2.1) and (6.2.12) we obtain

$$\begin{aligned} &\mathrm{P}(v_n^{-1} \max_{u \in S_n} |\underline{s}_n - \underline{s}_n^0| > \varepsilon) \\ &\leq K_t \varepsilon^{-t} \left(\mu_n^{-t} + v_n^{-2t} \left[\sum_{u \in S_n} \mathrm{P}(\mu_n v_n^{-5} |w_n| > 1) + \mathrm{P}(\lambda_{\max} > K_1) \right] \right) \\ &\leq K_t \varepsilon^{-t} n^{-t/68}, \end{aligned} \tag{6.2.27}$$

where the last step follows by replacing t with $5t + 102$ in (6.2.12) and t with $t/68$ in (6.2.1).

We now assume the n elements of S_n to be equally spaced between $-\sqrt{n}$ and $\sqrt{n}$. Since, for $|u_1 - u_2| \leq 2n^{-1/2}$,

$$|\underline{s}_n(u_1 + iv_n) - \underline{s}_n(u_2 + iv_n)| \leq 2n^{-1/2} v_n^{-2},$$

$$|\underline{s}_n^0(u_1 + iv_n) - \underline{s}_n^0(u_2 + iv_n)| \leq 2n^{-1/2} v_n^{-2},$$

and when $|u| \geq \sqrt{n}$, for large n,

$$|\underline{s}_n(u+iv_n)| \leq 2n^{-1/2} + v_n^{-1} I(\lambda_{\max} > K_1),$$

$$|\underline{s}_n^0(u+iv_n)| \leq 2n^{-1/2},$$

we conclude from (6.2.27) and (6.2.1) that for these n and any positive ε and t,

$$\mathrm{P}\left(v_n^{-1} \sup_{u\in\mathbb{R}} |\underline{s}_n(u+iv_n) - \underline{s}_n^0(u+iv_n)| > \varepsilon\right) \leq K_t \varepsilon^{-t} n^{-t/68}. \tag{6.2.28}$$

Let $\mathrm{E}_0(\cdot)$ denote expectation and $\mathrm{E}_k(\cdot)$ denote conditional expectation with respect to the σ-field generated by $\mathbf{r}_1, \cdots, \mathbf{r}_k$. Since, for any $r > 0$,

$$\mathrm{E}_k\left(v_n^{-r} \sup_{u\in\mathbb{R}} |\underline{s}_n(u+iv_n) - \underline{s}_n^0(u+iv_n)|^r\right)$$

for $k = 0, \ldots, n$ forms a martingale, it follows (from Jensen's inequality) that, for any $t \geq 1$, $(\mathrm{E}_k(v_n^{-r} \sup_{u\in\mathbb{R}} |\underline{s}_n(u+iv_n) - \underline{s}_n^0(u+iv_n)|^r))^t$, $k = 0, \ldots, n$, forms a submartingale. Therefore, for any $\varepsilon > 0$, $t \geq 1$, and $r > 0$, from Lemmas 6.7 and 6.8 and (6.2.28) with t replaced by $2rt$, we have

$$\begin{aligned}
&\mathrm{P}\left(\max_{k\leq n} \mathrm{E}_k\left(v_n^{-r} \sup_{u\in\mathbb{R}} |\underline{s}_n(u+iv_n) - \underline{s}_n^0(u+iv_n)|^r\right) > \varepsilon\right) \\
&\leq \varepsilon^{-t} \mathrm{E}\left(v_n^{-rt} \sup_{u\in\mathbb{R}} |\underline{s}_n(u+iv_n) - \underline{s}_n^0(u+iv_n)|^{rt}\right) \\
&\leq 2\varepsilon^{-t} K_{rt}^{1/2} n^{-rt/68}
\end{aligned} \tag{6.2.29}$$

whenever $\delta \leq 1/17$. From this, it follows that with probability 1,

$$\max_{k\leq n} \mathrm{E}_k\left(v_n^{-r} \sup_{u\in\mathbb{R}} |\underline{s}_n(u+iv_n) - \underline{s}_n^0(u+iv_n)|^r\right) \to 0. \tag{6.2.30}$$

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of $\underline{\mathbf{B}}_n$, and write

$$\underline{s}_{nj} = \underline{s}_{nj}^{out} + \underline{s}_{nj}^{in}, \quad j = 1, 2,$$

where

$$\underline{s}_{n2}^{out}(u+iv_n) = \frac{1}{n} \sum_{\lambda_j \in [a',b']} \frac{v_n}{(u-\lambda_j)^2 + v_n^2}.$$

Similarly, define

$$\underline{s}_{02}^{out}(u+iv_n) = \int_{x\in[a',b']} \frac{v_n}{(u-x)^2 + v_n^2} d\underline{F}^{y_n,H_n}(x) = 0.$$

By (6.2.30),

$$\max_{k\le n} \mathrm{E}_k \left(v_n^{-r} \sup_{u\in\mathbb{R}} |\underline{s}_{n2}(u+iv_n) - \underline{s}_2^0(u+iv_n)|^r \right) \to 0 \quad \text{a.s.} \tag{6.2.31}$$

Noting that on either $(\infty, a']$ or $[b', \infty)$ the collection of functions in λ $\{((\lambda - u)^2 + v_n^2)^{-1} : u \in [a,b]\}$ forms a uniformly bounded, equicontinuous family, we get as in (6.2.18)

$$\sup_{u\in[a,b]} v_n^{-r} |\underline{s}_{n2}^{in}(u+iv_n) - \underline{s}_{02}^{in}(u+iv_n)|$$

$$= \sup_{u\in[a,b]} \left| \int_{x\in[a',b']^c} \frac{1}{(x-u)^2+v_n^2} d(F^{\underline{\mathbf{B}}_n}(x) - \underline{F}^{y_n,H_n}(x)) \right| \to 0 \quad \text{a.s.}$$

Therefore, from Lemma 6.11,

$$\max_{k\le n} \mathrm{E}_k v_n^{-r} \sup_{u\in[a,b]} |\underline{s}_{n2}^{in}(u+iv_n) - \underline{s}_{02}^{in}(u+iv_n)|^r \to 0 \quad \text{a.s.}$$

This, together with (6.2.31), implies that

$$\max_{k\le n} v_n^{-r} \sup_{u\in[a,b]} \mathrm{E}_k(\underline{s}_{n2}^{out}(u+iv_n))^r \to 0 \quad \text{a.s.} \tag{6.2.32}$$

For any $u \in [a,b]$, we have

$$\begin{aligned}
&v_n^{-1} \underline{s}_{n2}^{out}(u+iv_n) \\
&\ge \int_{[a,b]} \frac{1}{(x-u)^2+v_n^2} dF^{\underline{\mathbf{B}}_n}(x) \\
&\ge \int_{[a,b]\cap[u-v_n,\, u+v_n]} \frac{1}{(x-u)^2+v_n^2} dF^{\underline{\mathbf{B}}_n}(x) \\
&\ge \frac{1}{2v_n^2} F^{\underline{\mathbf{B}}_n}([a,b]\cap[u-v_n,\ u+v_n]).
\end{aligned}$$

Therefore, by selecting $u_j \in [a,b]$ such that $v_n < u_j - u_{j-1}$ and $\cup[u_j - v_n,\ u_j + v_n] \supset [a,b]$, it follows that

$$\begin{aligned}
&v_n^{-r} \mathrm{E}_k(F^{\underline{\mathbf{B}}_n}([a,b]))^r \\
&\le v_n^{-r} \mathrm{E}_k \left(\sum_j F^{\underline{\mathbf{B}}_n}([a,b]\cap[u_j - v_n,\ u_j + v_n]) \right)^r \\
&\le v_n^{-r} \mathrm{E}_k \left(2 \sum_j (u_j - u_{j-1}) \sup_{u\in[a,b]} (\underline{s}_{n2}^{out}(u+iv_n)) \right)^r
\end{aligned}$$

$$\leq 2^r(b-a)^r v_n^{-r} \max_{k\leq n} \mathrm{E}_k \sup_{u\in[a,b]} (\underline{s}_{n2}^{out}(u+iv_n))^r \to 0, \text{ a.s.}$$

This shows that

$$\max_{k\leq n} \mathrm{E}_k (F^{\underline{\mathbf{B}}_n}\{[a,b]\})^r = o_{\text{a.s.}}(v_n^r) = o_{\text{a.s.}}(n^{-r/17}).$$

By replacing $[a,b]$ with the interval $[a',b']$, we get

$$\max_{k\leq n} \mathrm{E}_k (F^{\underline{\mathbf{B}}_n}\{[a',b']\})^r = o_{\text{a.s.}}(v_n^r) = o_{\text{a.s.}}(n^{-r/17}). \tag{6.2.33}$$

6.2.3 Convergence of $s_n - \mathrm{E}s_n$

We now restrict $\delta = 1/68$, that is, $v = v_n = n^{-1/68}$. The reader should note that the v_n defined in this subsection is different from what was defined in the last subsection, where $v_n \geq n^{-1/17}$.

Our goal is to show that

$$\sup_{u\in[a,b]} nv_n|s_n - \mathrm{E}s_n| \to 0 \qquad \text{a.s.} \qquad n\to\infty. \tag{6.2.34}$$

Write $\mathbf{D} = \mathbf{B}_n - z\mathbf{I}$, $\mathbf{D}_j = \mathbf{D} - \mathbf{r}_j\mathbf{r}_j^*$, and $\mathbf{D}_{j\underline{j}} = \mathbf{D} - (\mathbf{r}_j\mathbf{r}_j^* + \mathbf{r}_{\underline{j}}\mathbf{r}_{\underline{j}}^*)$, $j \neq \underline{j}$. Then $s_n = \frac{1}{p}\mathrm{tr}(\mathbf{D}^{-1})$. Let us also denote

$$\alpha_j = \mathbf{r}_j^*\mathbf{D}_j^{-2}\mathbf{r}_j - n^{-1}\mathrm{tr}(\mathbf{D}_j^{-2}\mathbf{T}_n), \quad a_j = n^{-1}\mathrm{tr}(\mathbf{D}_j^{-2}\mathbf{T}_n),$$

$$\beta_j = \frac{1}{1+\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j}, \quad \bar{\beta}_k = \frac{1}{1+n^{-1}\mathrm{tr}(\mathbf{T}_n\mathbf{D}_k^{-1})}, \quad b_n = \frac{1}{1+n^{-1}\mathrm{E}\mathrm{tr}(\mathbf{T}_n\mathbf{D}_1^{-1})},$$

$$\gamma_j = \mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j - n^{-1}\mathrm{E}(\mathrm{tr}(\mathbf{D}_j^{-1}\mathbf{T}_n)), \quad \hat{\gamma}_j = \mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j - n^{-1}\mathrm{tr}(\mathbf{D}_j^{-1}\mathbf{T}_n).$$

We first derive bounds on moments of γ_j and $\hat{\gamma}_j$. Using (6.2.2), we find for all $\ell \geq 1$,

$$\mathrm{E}|\hat{\gamma}_j|^{2\ell} \leq K_\ell n^{-2\ell}\mathrm{E}(\mathrm{tr}\mathbf{T}_n^{1/2}\mathbf{D}_j^{-1}\mathbf{T}_n\bar{\mathbf{D}}_j^{-1}\mathbf{T}_n^{1/2})^\ell \leq K_\ell n^{-\ell}v_n^{-2\ell}. \tag{6.2.35}$$

Using Lemma 2.12 and Lemma 6.9, we have, for $\ell \geq 1$,

$$\mathrm{E}|\gamma_j - \hat{\gamma}_j|^{2\ell} = \mathrm{E}|\gamma_1 - \hat{\gamma}_1|^{2\ell} = \mathrm{E}\left|\frac{1}{n}\sum_{j=2}^n \mathrm{E}_j\mathrm{tr}\mathbf{T}_n\mathbf{D}_1^{-1} - \mathrm{E}_{j-1}\mathrm{tr}\mathbf{T}_n\mathbf{D}_1^{-1}\right|^{2\ell}$$

$$= \mathrm{E}\left|\frac{1}{n}\sum_{j=2}^n \mathrm{E}_j\mathrm{tr}\mathbf{T}_n(\mathbf{D}_1^{-1} - \mathbf{D}_{1j}^{-1}) - \mathrm{E}_{j-1}\mathrm{tr}\mathbf{T}_n(\mathbf{D}_1^{-1} - \mathbf{D}_{1j}^{-1})\right|^{2\ell}$$

$$= \mathrm{E}\left|\frac{1}{n}\sum_{j=2}^{n}(\mathrm{E}_j - \mathrm{E}_{j-1})\frac{\mathbf{r}_j^*\mathbf{D}_{1j}^{-1}\mathbf{T}_n\mathbf{D}_{1j}^{-1}\mathbf{r}_j}{1+\mathbf{r}_j^*\mathbf{D}_{1j}^{-1}\mathbf{r}_j}\right|^{2\ell}$$
$$\le K_\ell \frac{1}{n^{2\ell}}\mathrm{E}\left(\sum_{j=2}^{n}\left|(\mathrm{E}_j - \mathrm{E}_{j-1})\frac{\mathbf{r}_j^*\mathbf{D}_{1j}^{-1}\mathbf{T}_n\mathbf{D}_{1j}^{-1}\mathbf{r}_j}{1+\mathbf{r}_j^*\mathbf{D}_{1j}^{-1}\mathbf{r}_j}\right|^2\right)^\ell$$
$$\le K_\ell n^{-\ell} v_n^{-2\ell}.$$

Therefore

$$\mathrm{E}|\gamma_j|^{2\ell} \le K_\ell n^{-\ell} v_n^{-2\ell}. \tag{6.2.36}$$

We next prove that b_n is bounded for all n. We have b_n, β_k, and $\bar{\beta}_k$ all bounded in absolute value by $|z|/v_n$ (see (6.2.5)). From (6.2.4), we see that $\mathrm{E}\beta_1 = -z\mathrm{E}\underline{s}_n$. Using (6.2.30), we get

$$\sup_{u\in[a,b]} |\mathrm{E}(\underline{s}_n(z)) - \underline{s}_n^0(z)| = o(v_n).$$

Since $\underline{s}_n^0$ is bounded for all n, $u \in [a,b]$ and v, we have $\sup_{u\in[a,b]} |\mathrm{E}\beta_1| \le K$.

Since $b_n = \beta_1 + \beta_1 b_n \gamma_1$, we get

$$\sup_{u\in[a,b]} |b_n| = \sup_{u\in[a,b]} |\mathrm{E}\beta_1 + \mathrm{E}\beta_1 b_n\gamma_1| \le K + K_2^{1/2} v_n^{-3} n^{-1/2} \le K.$$

Since $|s_n(u_1+iv_n) - s_n(u_2+iv_n)| \le |u_1-u_2|v_n^{-2}$, we see that (6.2.34) will follow from

$$\max_{u\in S_n} nv_n|s_n - \mathrm{E}s_n| \to 0 \qquad \text{a.s.},$$

where S_n now contains n^2 elements, equally spaced in $[a,b]$.

We write

$$\begin{aligned}
\mathrm{E}s_n - s_n &= -\frac{1}{p}\sum_{k=1}^{n}(\mathrm{E}_k\mathrm{tr}\mathbf{D}^{-1} - \mathrm{E}_{k-1}\mathrm{tr}\mathbf{D}^{-1}) \\
&= \frac{1}{p}\sum_{k=1}^{n}(\mathrm{E}_k - \mathrm{E}_{k-1})\left(\frac{\mathbf{r}_k^*\mathbf{D}_k^{-2}\mathbf{r}_k}{1+\mathbf{r}_k^*\mathbf{D}_k^{-1}\mathbf{r}_k}\right) \\
&= \frac{1}{p}\sum_{k=1}^{n}(\mathrm{E}_k - \mathrm{E}_{k-1})\frac{\mathbf{r}_k^*\mathbf{D}_k^{-2}\mathbf{r}_k - n^{-1}\mathrm{tr}(\mathbf{D}_k^{-2}\mathbf{T}_n)}{1+\mathbf{r}_k^*\mathbf{D}_k^{-1}\mathbf{r}_k} \\
&\quad + \frac{1}{p}\sum_{k=1}^{n}(\mathrm{E}_k - \mathrm{E}_{k-1})\frac{n^{-1}\mathrm{tr}(\mathbf{D}_k^{-2}\mathbf{T}_n)(n^{-1}\mathrm{tr}\mathbf{T}_n\mathbf{D}_k^{-1} - \mathbf{r}_k^*\mathbf{D}_k^{-1}\mathbf{r}_k)}{(1+n^{-1}\mathrm{tr}\mathbf{T}_n\mathbf{D}_k^{-1})(1+\mathbf{r}_k^*\mathbf{D}_k^{-1}\mathbf{r}_k)} \\
&= \frac{1}{p}\sum_{k=1}^{n}(\mathrm{E}_k - \mathrm{E}_{k-1})\alpha_k\beta_k - \frac{1}{p}\sum_{k=1}^{n}(\mathrm{E}_k - \mathrm{E}_{k-1})a_k\hat{\gamma}_k\bar{\beta}_k\beta_k \\
&\equiv W_1 - W_2.
\end{aligned}$$

Let F_{nj} be the ESD of the matrix $\sum_{k\neq j} \mathbf{r}_k \mathbf{r}_k^*$. From Lemma A.43 and (6.2.33), for any r, we have

$$\max_k \mathrm{E}_k (F_{nk}([a', b']))^r = o(n^{-r/17}) = o(v_n^{4r}) \quad \text{a.s.} \tag{6.2.37}$$

Define

$$\begin{aligned} B_k &= I(\mathrm{E}_{k-1} F_{nk}([a', b']) \leq v_n^4) \cap (\mathrm{E}_{k-1}(F_{nk}([a', b']))^2 \leq v_n^8) \\ &= I(\mathrm{E}_k F_{nk}([a', b']) \leq v_n^4) \cap (\mathrm{E}_k(F_{nk}([a', b']))^2 \leq v_n^8). \end{aligned}$$

By (6.2.37), we have

$$\mathrm{P}\left(\bigcup_{k=1}^n [B_k = 0], \text{i.o.}\right) = 0.$$

Therefore, we have, for any $\varepsilon > 0$,

$$\begin{aligned} &P\left(\max_{u\in S_n} |n v_n W_1| > \varepsilon, \text{i.o.}\right) \\ &\leq \mathrm{P}\left(\left(\left[\max_{u\in S_n}\left| v_n \sum_{k=1}^n (\mathrm{E}_k - \mathrm{E}_{k-1})(\alpha_k \beta_k)\right| > \tilde{\varepsilon}\right] \bigcap_{k=1}^n [B_k = 1]\right), \text{i.o.}\right) \\ &= \mathrm{P}\left(\left(\left[\max_{u\in S_n}\left| v_n \sum_{k=1}^n (\mathrm{E}_k - \mathrm{E}_{k-1})(\alpha_k \beta_k) B_k\right| > \tilde{\varepsilon}\right] \bigcap_{k=1}^n [B_k = 1]\right), \text{i.o.}\right) \\ &\leq \mathrm{P}\left(\max_{u\in S_n}\left| v_n \sum_{k=1}^n (\mathrm{E}_k - \mathrm{E}_{k-1})(\alpha_k \beta_k) B_k\right| > \tilde{\varepsilon}, \text{i.o.}\right), \end{aligned}$$

where $\tilde{\varepsilon} = \inf_n p\varepsilon/n > 0$. Note that, for each $u \in \mathbb{R}$, $\{(\mathrm{E}_k - \mathrm{E}_{k-1})(\alpha_k \beta_k) B_k\}$ forms a martingale difference sequence.

By Lemma 2.13, we have for each $u \in [a, b]$ and $\ell \geq 1$,

$$\begin{aligned} &\mathrm{E}\left| v_n \sum_{k=1}^n (\mathrm{E}_k - \mathrm{E}_{k-1})(\alpha_k \beta_k) B_k\right|^{2\ell} \\ &\leq K_\ell \left(\mathrm{E}\left(\sum_{k=1}^n \mathrm{E}_{k-1} |v_n (\alpha_k \beta_k) B_k|^2\right)^\ell + \sum_{k=1}^n \mathrm{E}|v_n (\alpha_k \beta_k) B_k|^{2\ell}\right) \\ &\leq K_\ell \left(\mathrm{E}\left(\sum_{k=1}^n \mathrm{E}_{k-1} |v_n (\alpha_k \beta_k) B_k|^2\right)^\ell + \sum_{k=1}^n \mathrm{E}|z|^{2\ell} |\alpha_k|^{2\ell} B_k\right) \\ &\leq K_\ell \left(\mathrm{E}\left(\sum_{k=1}^n \mathrm{E}_{k-1} |v_n (\alpha_k \beta_k) B_k|^2\right)^\ell + n^{1-\ell} v_n^{-4\ell}\right). \end{aligned} \tag{6.2.38}$$

Note that when $|b_n| \le K_0$, by $|\alpha_k\beta_k| \le v_n^{-1} + (p/n)|z|v_n^{-3}$,

$$\begin{aligned}|\alpha_k\beta_k|^2 &\le 4K_0^2|\alpha_k|^2 + Kv_n^{-6}I(|\beta_k| \ge 2K_0)\\ &\le 4K_0^2|\alpha_k|^2 + Kv_n^{-6}I(|\gamma_k| \ge 1/(2K_0)).\end{aligned}$$

On the other hand, by (6.2.2),

$$\begin{aligned}&\mathrm{E}_{k-1}|\alpha_k B_k|^2\\ &\le K\mathrm{E}_{k-1}n^{-2}B_k\mathrm{tr}(\mathbf{D}_k^{-2}\mathbf{T}_n\overline{\mathbf{D}}_k^{-2}\mathbf{T}_n)\\ &\le Kn^{-2}B_k\mathrm{E}_{k-1}\mathrm{tr}(\mathbf{D}_k^{-2}\overline{\mathbf{D}}_k^{-2}).\end{aligned}$$

Let λ_{kj} denote the j-th largest eigenvalue of $\sum_{j\neq k}\mathbf{r}_j\mathbf{r}_j^*$. By (6.2.37), we have

$$\begin{aligned}&\sum_{k=1}^n B_k\mathrm{E}_{k-1}\mathrm{tr}\mathbf{D}_k^{-2}\overline{\mathbf{D}}_k^{-2}\\ &= \sum_{k=1}^n B_k\mathrm{E}_{k-1}\left[\sum_{\lambda_{kj}\notin[a'b']}\frac{1}{((\lambda_{kj}-u)^2+v_n^2)^2} + \sum_{\lambda_{kj}\in[a'b']}\frac{1}{((\lambda_{kj}-u)^2+v_n^2)^2}\right]\\ &\le \sum_{k=1}^n(p\underline{\varepsilon}^{-4} + B_k v_n^{-4}\mathrm{E}_{k-1}pF_{nk}([a',b'])) \le Kn^2.\end{aligned}$$

Substituting the two estimates above into (6.2.38) and applying (6.2.36), for any $\ell > 2$ and $t > 2\ell$, we have

$$\begin{aligned}&\mathrm{P}\left(\max_{u\in S_n}\left|v_n\sum_{k=1}^n(\mathrm{E}_k - \mathrm{E}_{k-1})(\alpha_k\beta_k)B_k\right| > \tilde{\varepsilon}\right)\\ &\le Kn^2\left[\mathrm{E}\left(v_n^2 + v_n^{-4}\sum_{k=1}^n\mathrm{E}_{k-1}I(|\gamma_k| \ge 1/(2K_0))\right)^\ell + n^{1-\ell}v_n^{1-4\ell}\right]\\ &\le Kn^2\left[v_n^{2\ell} + v_n^{-4\ell}n^{\ell-1}\sum_{k=1}^n\mathrm{P}(|\gamma_k| \ge 1/(2K_0))\right]\\ &\le Kn^2\left[v_n^{2\ell} + v_n^{-4\ell}n^{\ell-1}\sum_{k=1}^n\mathrm{E}|\gamma_k|^{2t}\right]\\ &\le K_{\ell,\tilde{\varepsilon}}n^{2-\ell/34},\end{aligned}$$

which is summable when $\ell > 102$. Therefore,

$$\max_{u\in S_n}|W_1| = o(1/(nv_n)) \quad \text{a.s.} \tag{6.2.39}$$

We can proceed the same way for the proof of

$$\max_{u \in S_n} |W_2| = o(1/(nv_n)) \quad \text{a.s.} \tag{6.2.40}$$

It is straightforward to show that $|a_k \bar{\beta}_k| \le v_n^{-1}$. Again, with K_0 a bound on b_n, we have

$$\begin{aligned} |a_k \hat{\gamma}_k \bar{\beta}_k \beta_k|^2 &\le (2K_0)^4 |a_k \hat{\gamma}_k|^2 + K v_n^{-4} |\hat{\gamma}_k|^2 I(|\bar{\beta}_k \beta_k| \ge (2K_0)^2) \\ &\le (2K_0)^4 |a_k \hat{\gamma}_k|^2 + K v_n^{-4} |\hat{\gamma}_k|^2 I(|\gamma_k| \text{ or } |\hat{\gamma}_k| \ge 1/(4K_0)). \end{aligned}$$

We have, by (6.2.2),

$$\begin{aligned} &\sum_{k=1}^n \mathrm{E}_{k-1} v_n^2 |a_k \hat{\gamma}_k|^2 B_k \\ &\le K v_n^2 n^{-2} \sum_{k=1}^n \mathrm{E}_{k-1} B_k |a_k|^2 \mathrm{tr} \mathbf{D}_k^{-1} \overline{\mathbf{D}}_k^{-1} \\ &\le K v_n^2 n^{-4} \sum_{k=1}^n \mathrm{E}_{k-1} B_k p \sum_j \frac{1}{((\lambda_{kj} - u)^2 + v_n^2)^2} \sum_k \frac{1}{(\lambda_{kj} - u)^2 + v_n^2} \\ &\le K n^{-3} v_n^2 \sum_{k=1}^n \mathrm{E}_{k-1} B_k (p \underline{\varepsilon}^{-4} + v_n^{-4} p F_{nk}([a', b']))(p \underline{\varepsilon}^{-2} + v_n^{-2} p F_{nk}([a', b'])) \\ &\le K v_n^2. \end{aligned}$$

By noting $|a_k| \le (p/n) v_n^{-2}$ and using (6.2.36), we have, for $\ell \ge 2$,

$$\sum_{k=1}^n \mathrm{E}|v_n (a_k \gamma_k \bar{\beta}_k \beta_k) B_k|^{2\ell} \le K v_n^{-6\ell} \sum_{k=1}^n \mathrm{E}|\gamma_k|^{2\ell} \le K v_n^{-8\ell} n^{1-\ell} \le K v_n^{2\ell}.$$

Therefore, by Lemma 2.13, (6.2.35), and (6.2.36), we have, for all $\ell \ge 2$,

$$\begin{aligned} &n^2 \mathrm{E} \left| v_n \sum_{k=1}^n (\mathrm{E}_k - \mathrm{E}_{k-1})(a_k \hat{\gamma}_k \bar{\beta}_k \beta_k) B_k \right|^{2\ell} \\ &\le K n^2 \left(\mathrm{E} \left(\sum_{k=1}^n \mathrm{E}_{k-1} |v_n (a_k \hat{\gamma}_k \bar{\beta}_k \beta_k) B_k|^2 \right)^\ell + \sum_{k=1}^n \mathrm{E}|v_n (a_k \hat{\gamma}_k \bar{\beta}_k \beta_k) B_k|^{2\ell} \right) \\ &\le K n^2 \left(\mathrm{E} \left(v_n^2 + v_n^{-4} \sum_{k=1}^n \mathrm{E}_{k-1} |\hat{\gamma}_k|^2 I(|\gamma_k| \text{ or } |\bar{\gamma}_k| \ge 1/(4K_0)) \right)^\ell + v_n^{2\ell} \right) \\ &\le K n^2 \left(v_n^{2\ell} + v_n^{-4\ell} n^{\ell-1} \sum_{k=1}^n \mathrm{E}|\hat{\gamma}_k|^{2\ell} I(|\gamma_k| \text{ or } |\hat{\gamma}_k| \ge 1/(4K_0)) \right) \\ &\le K n^2 \left(v_n^{2\ell} + v_n^{-4\ell} n^{\ell-1} \sum_{k=1}^n \mathrm{E}|\hat{\gamma}_k|^{2\ell} [|\gamma_k|^{2\ell} + |\hat{\gamma}_k|^{2\ell}] \right) \end{aligned}$$

$$\leq Kn^2 \left(v_n^{2\ell} + v_n^{-4\ell} n^{\ell-1} \sum_{k=1}^{n} (\mathrm{E}|\hat{\gamma}_k|^{2\ell}|\gamma_k|^{2\ell} + \mathrm{E}|\hat{\gamma}_k|^{4\ell}) \right)$$
$$\leq Kn^2 \left(v_n^{2\ell} + v_n^{-8\ell} n^{-\ell} \right)$$
$$\leq Kn^2 v_n^{2\ell} = Kn^{2-\ell/34}.$$

Then, (6.2.40) follows if we choose $\ell > 102$.

Combining (6.2.39) and (6.2.40), then (6.2.34) follows.

6.2.4 Convergence of the Expected Value

Our next goal is to show that, for $v = n^{-1/68}$,

$$\sup_{u\in[a,b]} |\mathrm{E}\underline{s}_n - \underline{s}_n^0| = O(1/n). \tag{6.2.41}$$

We begin by deriving an identity similar to (6.2.7). Write

$$\mathbf{D} - (-z\mathrm{E}\underline{s}_n(z)\mathbf{T}_n - z\mathbf{I}) = \sum_{j=1}^{n} \mathbf{r}_j\mathbf{r}_j^* - (-z\mathrm{E}\underline{s}_n(z))\mathbf{T}_n.$$

Taking first inverses and then the expected value, we get

$$(-z\mathrm{E}\underline{s}_n\mathbf{T}_n - z\mathbf{I})^{-1} - \mathrm{E}\mathbf{D}^{-1}$$
$$= (-z\mathrm{E}\underline{s}_n\mathbf{T}_n - z\mathbf{I})^{-1}\mathrm{E}\left[\left(\sum_{j=1}^{n}\mathbf{r}_j\mathbf{r}_j^* - (-z\mathrm{E}\underline{s}_n)\mathbf{T}_n\right)\mathbf{D}^{-1}\right]$$
$$= -z^{-1}\sum_{j=1}^{n}\mathrm{E}\beta_j\left[(\mathrm{E}\underline{s}_n\mathbf{T}_n + \mathbf{I})^{-1}\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1} - \frac{1}{n}(\mathrm{E}\underline{s}_n\mathbf{T}_n + \mathbf{I})^{-1}\mathbf{T}_n\mathrm{E}\mathbf{D}^{-1}\right]$$
$$= -z^{-1}n\mathrm{E}\beta_1\left[(\mathrm{E}\underline{s}_n\mathbf{T}_n + \mathbf{I})^{-1}\mathbf{r}_1\mathbf{r}_1^*\mathbf{D}_1^{-1} - \frac{1}{n}(\mathrm{E}\underline{s}_n\mathbf{T}_n + \mathbf{I})^{-1}\mathbf{T}_n\mathrm{E}\mathbf{D}^{-1}\right].$$

Taking the trace on both sides and dividing by $-n/z$, we get

$$y_n\int\frac{dH_n(t)}{1+t\mathrm{E}\underline{s}_n} + zy_n\mathrm{E}(s_n(z))$$
$$= \mathrm{E}\beta_1\left[\mathbf{r}_1^*\mathbf{D}_1^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}_n + \mathbf{I})^{-1}\mathbf{r}_1 - \frac{1}{n}\mathrm{Etr}(\mathrm{E}\underline{s}_n\mathbf{T}_n + \mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}^{-1}\right]. \tag{6.2.42}$$

We first show

$$\sup_{u\in[a,b]}\left|\mathrm{Etr}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}^{-1}-\mathrm{Etr}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}_1^{-1}\right|\le K. \quad (6.2.43)$$

From (6.2.37), we get

$$\sup_{u\in[a,b]}\mathrm{E}(\mathrm{tr}\mathbf{D}_1^{-1}\overline{\mathbf{D}}_1^{-1})^2\le \mathrm{E}(p\underline{\varepsilon}^{-2}+v_n^{-2}pF_{n1}([a',b']))^2\le Kn^2 \qquad (6.2.44)$$

and

$$\sup_{u\in[a,b]}\mathrm{Etr}\mathbf{D}_1^{-2}\overline{\mathbf{D}}_1^{-2}\le \mathrm{E}(p\underline{\varepsilon}^{-4}+v_n^{-4}pF_{n1}([a',b']))\le Kn. \qquad (6.2.45)$$

Also, because of (6.2.29) and the fact that $-1/\underline{s}_n^0(z)$ stays uniformly away from the eigenvalues of $\mathbf{T}_n$ for all $u\in[a,b]$, we must have

$$\sup_{u\in[a,b]}\|(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\|\le K. \qquad (6.2.46)$$

Therefore, from (6.2.3), (6.2.36), (6.2.44)–(6.2.46), and the fact that $\sup_{u\in[a,b]}|b_n|$ is bounded, we get

$$\text{left-hand side of (6.2.43)} = \sup_{u\in[a,b]}|\mathrm{E}\beta_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}_1^{-1}\mathbf{r}_1|$$
$$\le \sup_{u\in[a,b]}(|b_n|\cdot|\mathrm{E}\mathbf{r}_1^*\mathbf{D}_1^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}_1^{-1}\mathbf{r}_1|$$
$$+\mathrm{E}|\beta_1 b_n\gamma_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}_1^{-1}\mathbf{r}_1|)$$
$$\le K\sup_{u\in[a,b]}(n^{-1}|\mathrm{Etr}\mathbf{T}_n^{1/2}\mathbf{D}_1^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}_1^{-1}\mathbf{T}_n^{1/2}|$$
$$+v_n^{-1}(\mathrm{E}|\gamma_1|^2)^{1/2}(E|\mathbf{r}_1^*\mathbf{D}_1^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}_1^{-1}\mathbf{r}_1|^2)^{1/2})$$
$$\le K\sup_{u\in[a,b]}(n^{-1}\mathrm{Etr}\mathbf{D}_1^{-1}\overline{\mathbf{D}}_1^{-1}+v_n^{-2}n^{-3/2}(\mathrm{Etr}\mathbf{D}_1^{-2}\overline{\mathbf{D}}_1^{-2}+\mathrm{E}(\mathrm{tr}\mathbf{D}_1^{-1}\overline{\mathbf{D}}_1^{-1})^2)^{1/2})$$
$$\le K.$$

Thus (6.2.43) holds.

From (6.2.2), (6.2.44), and (6.2.46), we get

$$\sup_{u\in[a,b]}\mathrm{E}|\mathbf{r}_1^*\mathbf{D}_1^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{r}_1-n^{-1}\mathrm{tr}\mathbf{D}_1^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n|^2$$
$$\le Kn^{-2}\sup_{u\in[a,b]}\mathrm{Etr}\mathbf{D}_1^{-1}\overline{\mathbf{D}}_1^{-1}\le Kn^{-1}. \qquad (6.2.47)$$

Next, we show

$$\sup_{u\in[a,b]}\mathrm{E}|\mathrm{tr}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}_1^{-1}-\mathrm{Etr}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}_1^{-1}|^2\le Kn. \qquad (6.2.48)$$

Let

$$\beta_{1j} = \frac{1}{1 + \mathbf{r}_j^* \mathbf{D}_{1j}^{-1} \mathbf{r}_j},$$
$$b_{1n} = \frac{1}{1 + n^{-1}\mathrm{Etr}(\mathbf{T}_n \mathbf{D}_{12}^{-1})},$$
$$\gamma_{1j} = \mathbf{r}_j^* \mathbf{D}_{1j}^{-1} \mathbf{r}_j - n^{-1}\mathrm{E}(\mathrm{tr}(\mathbf{D}_{1j}^{-1}\mathbf{T}_n)).$$

It is easy to see that these three quantities are the same as their counterparts in the previous section with n replaced by $n-1$ and z replaced by $(n/(n-1))z$. Thus, by deriving the bounds on the quantities in the previous section with an interval slightly larger than $[a, b]$ (still satisfying assumption (f)), we see that γ_{1j} satisfies the same bound as in (6.2.36) and that $\sup_{u\in[a,b]} |\mathrm{E}\beta_{1j}|$ and $\sup_{u\in[a,b]} |b_{1n}|$ are both bounded.

It is also clear that the bounds in (6.2.37), (6.2.44), and (6.2.45) hold when two columns of $\mathbf{X}$ are removed. Moreover, with F_{n12} denoting the ESD of $\sum_{j\neq 1,2} \mathbf{r}_j \mathbf{r}_j^*$, we get

$$\sup_{u\in[a,b]} \mathrm{E}(\mathrm{tr}\mathbf{D}_{12}^{-1}\overline{\mathbf{D}}_{12}^{-1})^4 \le \mathrm{E}(p\underline{\varepsilon}^{-2} + v_n^{-2} p F_{n12}([a', b']))^4$$
$$\le K n^4(\underline{\varepsilon}^{-8} + v_n^{-8}\mathrm{E}(F_{n12}([a', b']))^2) \le K n^4$$

and

$$\sup_{u\in[a,b]} \mathrm{E}(\mathrm{tr}\mathbf{D}_{12}^{-2}\overline{\mathbf{D}}_{12}^{-2})^2 \le \mathrm{E}(p\underline{\varepsilon}^{-4} + v_n^{-4} p F_{n12}([a', b']))^2 \le K n^2.$$

With these facts and (6.2.3), for any nonrandom $p \times p$ matrix $\mathbf{A}$ with bounded norm, we have

$$\sup_{u\in[a,b]} \mathrm{E}|\mathrm{tr}\mathbf{A}\mathbf{D}_1^{-1} - \mathrm{Etr}\mathbf{A}\mathbf{D}_1^{-1}|^2 = \sup_{u\in[a,b]} \sum_{j=2}^{n} \mathrm{E}|(\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{A}\mathbf{D}_1^{-1}|^2$$
$$\le 2 \sup_{u\in[a,b]} \sum_{j=2}^{n} \mathrm{E}|\beta_{1j}\mathbf{r}_j^*\mathbf{D}_{1j}^{-1}\mathbf{A}\mathbf{D}_{1j}^{-1}\mathbf{r}_j|^2$$
$$= 2(n-1) \sup_{u\in[a,b]} \mathrm{E}|(b_{1n} - \beta_{12}b_{1n}\gamma_{12})\mathbf{r}_2^*\mathbf{D}_{12}^{-1}\mathbf{A}\mathbf{D}_{12}^{-1}\mathbf{r}_2|^2$$
$$\le Kn \sup_{u\in[a,b]} \left(\mathrm{E}|\mathbf{r}_2^*\mathbf{D}_{12}^{-1}\mathbf{A}\mathbf{D}_{12}^{-1}\mathbf{r}_2|^2 + v_n^{-2}(\mathrm{E}|\gamma_{12}|^4\mathrm{E}|\mathbf{r}_2^*\mathbf{D}_{12}^{-1}\mathbf{A}\mathbf{D}_{12}^{-1}\mathbf{r}_2|^4)^{1/2}\right)$$
$$\le Kn^{-1} \sup_{u\in[a,b]} \Big[\mathrm{E}(\mathrm{tr}\mathbf{D}_{12}^{-2}\overline{\mathbf{D}}_{12}^{-2}) + \mathrm{E}(\mathrm{tr}\mathbf{D}_{12}^{-1}\overline{\mathbf{D}}_{12}^{-1})^2$$
$$+ n^{-1}v_n^{-4}(\mathrm{Etr}(\mathbf{D}_{12}^{-2}\overline{\mathbf{D}}_{12}^{-2})^2 + \mathrm{E}(\mathrm{tr}\mathbf{D}_{12}^{-1}\overline{\mathbf{D}}_{12}^{-1})^4)^{1/2}\Big]$$
$$\le Kn^{-1}(n^2 + nv_n^{-4}) \le Kn.$$

Thus, using (6.2.46), when $\mathbf{A} = (\mathrm{E}\underline{s}_n\mathbf{T}_n + \mathbf{I})^{-1}\mathbf{T}_n$, we get (6.2.48). Moreover, when $\mathbf{A} = \mathbf{I}$, we have just shown

$$\sup_{u\in[a,b]} \mathrm{E}|\gamma_1 - \hat{\gamma}_1|^2 \leq Kn^{-1}.$$

Also, from (6.2.2) and (6.2.44), when $\ell = 1$,

$$\sup_{u\in[a,b]} \mathrm{E}|\hat{\gamma}_1|^2 \leq \sup_{u\in[a,b]} Kn^{-2}\mathrm{Etr}\mathbf{D}_1^{-1}\overline{\mathbf{D}}_1^{-1} \leq Kn^{-1}.$$

Therefore

$$\sup_{u\in[a,b]} \mathrm{E}|\gamma_1|^2 \leq Kn^{-1}. \tag{6.2.49}$$

From (6.2.36), (6.2.42), (6.2.43), and (6.2.47)–(6.2.49), we get

$$\begin{aligned}
&\sup_{u\in[a,b]}\left|y_n\int\frac{dH_n(t)}{1+t\mathrm{E}\underline{s}_n} + zy_n\mathrm{E}(s_n)\right| \\
&\leq Kn^{-1} + \sup_{u\in[a,b]}\left|\mathrm{E}\beta_1\left[\mathbf{r}_1^*\mathbf{D}_1^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{r}_1\right.\right. \\
&\qquad \left.\left.-(1/n)\mathrm{Etr}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}_1^{-1}\right]\right| \\
&= Kn^{-1} + \sup_{u\in[a,b]}|b_n|^2\left|\mathrm{E}(\gamma_1-\beta_1\gamma_1^2)\left[\mathbf{r}_1^*\mathbf{D}_1^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{r}_1\right.\right. \\
&\qquad \left.\left.-(1/n)\mathrm{Etr}(\mathrm{E}\underline{s}_n\mathbf{T}_n+\mathbf{I})^{-1}\mathbf{T}_n\mathbf{D}_1^{-1}\right]\right| \\
&\leq K\left(n^{-1} + \sup_{u\in[a,b]}(\mathrm{E}|\gamma_1|^2 + v_n^{-2}\mathrm{E}|\gamma_1|^4)^{1/2}n^{-1/2}\right) \\
&\leq K(n^{-1} + (n^{-1}+v_n^{-2}n^{-2}v_n^{-4})^{1/2}n^{-1/2}) \leq Kn^{-1}.
\end{aligned}$$

As in Subsection 6.2.2, we let

$$w_n = -\frac{1}{z}\int\frac{dH_n(t)}{1+t\mathrm{E}\underline{s}_n(z)} - \mathrm{E}(s_n(z))$$

and

$$R_n = -z - \frac{1}{\mathrm{E}\underline{s}_n} + y_n\int\frac{tdH_n(t)}{1+t\mathrm{E}\underline{s}_n}.$$

Then

$$\sup_{u\in[a,b]}|w_n| \leq Kn^{-1},$$

$R_n = w_n z y_n/\mathrm{E}\underline{s}_n$, and equation (6.2.25) together with the steps leading to (6.2.26) hold with $\underline{s}_n$ replaced with its expected value. From (6.2.15) it is clear that $\underline{s}_n^0$ must be uniformly bounded away from 0 for all $u \in [a,b]$ and all n. From (6.2.30), we see that $\mathrm{E}\underline{s}_n$ must also satisfy this same property. Therefore

$$\sup_{u\in[a,b]} |R_n| \le Kn^{-1}.$$

Using (6.2.16), (6.2.17), (6.2.19), and (6.2.20), it follows that $\sup_{u\in[a,b]} |v_n^{-1}\underline{s}_{n2}^0|$ is bounded in n and hence

$$\sup_{u\in[a,b]} \frac{\underline{s}_{n2}^0 y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n^0|^2}}{v_n + \underline{s}_{n2}^0 y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n^0|^2}}$$

is bounded away from 1 for all n. Therefore, we get for all n sufficiently large,

$$\sup_{u\in[a,b]} |\mathrm{E}\underline{s}_n - \underline{s}_n^0| \le Ky_n|z\underline{s}_n^0 w_n| \le Kn^{-1},$$

which is (6.2.41).

6.2.5 Completing the Proof

From the last two sections, we get

$$\sup_{u\in[a,b]} |\underline{s}_n(z) - \underline{s}_n^0(z)| = o(1/(nv_n)) \quad \text{a.s.} \tag{6.2.50}$$

when $v_n = n^{-1/68}$. It is clear from the arguments used in Subsections 6.2.2–6.2.4 that (6.2.50) is true when the imaginary part of z is replaced by a constant multiple of v_n. In fact, we have

$$\max_{k\in\{1,2,\cdots,34\}} \sup_{u\in[a,b]} |\underline{s}_n(u+i\sqrt{k}v_n) - \underline{s}_n^0(u+i\sqrt{k}v_n)| = o(1/(nv_n)) = o(v_n^{67}) \text{ a.s.}$$

We take the imaginary part and get

$$\max_{k\in\{1,2,\cdots,34\}} \sup_{u\in[a,b]} \left| \int \frac{d(F^{\underline{\mathbf{B}}_n}(\lambda) - F^{y_n,H_n}(\lambda))}{(u-\lambda)^2 + kv_n^2} \right| = o(v_n^{66}) \text{ a.s.}$$

Upon taking differences, we find

$$\max_{k_1\ne k_2} \sup_{u\in[a,b]} \left| \int \frac{v_n^2\, d(F^{\underline{\mathbf{B}}_n}(\lambda) - F^{y_n,H_n}(\lambda))}{((u-\lambda)^2 + k_1v_n^2)((u-\lambda)^2 + k_2v_n^2)} \right| = o(v_n^{66}) \quad \text{a.s.}$$

$$\vdots$$

$$\sup_{u\in[a,b]}\left|\int \frac{(v_n^2)^{33}\, d(F^{\underline{\mathbf{B}}_n}(\lambda) - F^{y_n,H_n}(\lambda))}{((u-\lambda)^2+v_n^2)((u-\lambda)^2+2v_n^2)\cdots((u-\lambda)^2+34v_n^2)}\right| = o(v_n^{66}),\ \text{a.s.}$$

Thus

$$\sup_{u\in[a,b]}\left|\int \frac{d(F^{\underline{\mathbf{B}}_n}(\lambda) - F^{y_n,H_n}(\lambda))}{((u-\lambda)^2+v_n^2)((u-\lambda)^2+2v_n^2)\cdots((u-\lambda)^2+34v_n^2)}\right| = o(1) \quad \text{a.s.}$$

We split up the integral and get

$$\sup_{u\in[a,b]}\left|\int \frac{I_{[a',b']^c}\, d(F^{\underline{\mathbf{B}}_n}(\lambda) - F^{y_n,H_n}(\lambda))}{((u-\lambda)^2+v_n^2)((u-\lambda)^2+2v_n^2)\cdots((u-\lambda)^2+34v_n^2)}\right. \tag{6.2.51}$$

$$\left. + \sum_{\lambda_j\in[a',b']} \frac{v_n^{68}}{((u-\lambda_j)^2+v_n^2)((u-\lambda_j)^2+2v_n^2)\cdots((u-\lambda_j)^2+34v_n^2)}\right| = o(1),\ \text{a.s.}$$

Now if, for each term in a subsequence satisfying (6.2.51), there is at least one eigenvalue contained in $[a,b]$, then the sum in (6.2.51), with u evaluated at these eigenvalues, will be uniformly bounded away from 0. Thus, at these same u values, the integral in (6.2.51) must also stay uniformly bounded away from 0. But the integral converges to zero a.s. since the integrand is bounded and, with probability 1, both $F^{\underline{\mathbf{B}}_n}$ and F^{y_n,H_n} converge weakly to the same limit having no mass on $\{a',b'\}$. Thus, with probability 1, no eigenvalues of $\mathbf{B}_n$ will appear in $[a,b]$ for all n sufficiently large. This completes the proof of (1).

6.3 Proof of (2)

Throughout the remainder of this chapter, there will be frequent referrals to Theorem 5.11 whenever the limiting properties of the extreme eigenvalues of $(1/n)\mathbf{X}_n\mathbf{X}_n^*$ are needed, even though the assumptions of the theorem are not necessarily met. However, it can be seen from the proof of Theorem 5.10 that the results are true for our $\mathbf{X}_n$, namely, the $\mathbf{X}_n$'s need not come from one doubly infinite array of random variables.

We now begin the proof of (2). We see first off that x_0 must coincide with the boundary point in (d) of Lemma 6.2. Most of (d) will be proven in the following lemma.

Lemma 6.12. *If $y[1-H(0)]>1$, then the smallest value in the support of F^{y_n,H_n} is positive for all large n, and it converges to the smallest value, also positive, in the support of $F^{y,H}$ as $n\to\infty$.*

Proof. Assume $y[1-H(0)]>1$. Write

$$z_{y,H}(s) = \frac{1}{s}\left(-1 + y\int \frac{ts}{1+ts}dH(t)\right),$$

$$z'_{y,H}(s) = \frac{1}{s^2}\left(1 - y\int\left(\frac{ts}{1+ts}\right)^2 dH(t)\right).$$

As s increases in $\mathbb{R}^+$, the two integrals increase from 0 to $1 - H(0)$, which implies $z_{y,H}(s)$ increases from $-\infty$ to a maximum value and decreases to zero. Let $\hat{s}$ denote the number where the maximum occurs. Then, by Lemma 6.1, $x_0 \equiv z_{y,H}(\hat{s})$ is the smallest value in the support of $F^{y,H}$. We see that $\hat{s}$ is s_y in (d) of Lemma 6.2.

We have

$$y\int\left(\frac{t\hat{s}}{1+t\hat{s}}\right)^2 dH(t) = 1.$$

From this it is easy to verify

$$z_{y,H}(\hat{s}) = y\int\frac{t}{(1+t\hat{s})^2}dH(t).$$

Therefore $z_{y,H}(\hat{s}) > 0$.

Since $\limsup_n H_n(0) \le H(0)$, we have $y_n(1 - H_n(0)) > 1$ for all large n. We consider now only these n and we let $\hat{s}_n$ denote the value where the maximum of $z_{y_n,H_n}(s)$ occurs in $\mathbb{R}^+$. We see that $z_{y_n,H_n}(\hat{s}_n)$ is the smallest positive value in the support of F^{y_n,H_n}. It is clear that, for all positive s, $z_{y_n,H_n}(s) \to z_{y,H}(s)$ and $z'_{y_n,H_n}(s) \to z'_{y,H}(s)$ as $n \to \infty$ uniformly on any closed subset of $\mathbb{R}^+$. Thus, for any positive s_1, s_2 such that $s_1 < \hat{s} < s_2$, we have, for all large n,

$$z'_{y_n,H_n}(s_1) > 0 > z'_{y_n,H_n}(s_2),$$

which implies $s_1 < \hat{s}_n < s_2$. Therefore, $\hat{s}_n \to \hat{s}$ and, in turn, $z_{y_n,H_n}(\hat{s}_n) \to x_0$ as $n \to \infty$. This completes the proof of the lemma.

We now prove that when $y[1 - H(0)] > 1$,

$$\lambda_n^{\mathbf{B}_n} \xrightarrow{\text{a.s.}} x_0 \quad \text{as } n \to \infty. \tag{6.3.1}$$

Assume first that $\mathbf{T}_n$ is nonsingular with $\lambda_n^{\mathbf{T}_n}$ uniformly bounded away from 0. Using Theorem A.10, we find

$$\lambda_n^{(1/n)\mathbf{X}_n\mathbf{X}_n^*} \le \lambda_n^{\mathbf{B}_n}\lambda_1^{\mathbf{T}_n^{-1}} = \lambda_n^{\mathbf{B}_n}\left(\lambda_n^{\mathbf{T}_n}\right)^{-1}.$$

Since by Theorem 5.11 $\lambda_n^{(1/n)\mathbf{X}_n\mathbf{X}_n^*} \xrightarrow{\text{a.s.}} (1-\sqrt{y})^2$ as $n \to \infty$, we conclude that $\liminf_n \lambda_n^{\mathbf{B}_n} > 0$ a.s. Since, by Lemma 6.12, the interval $[a, b]$ in (1) can be made arbitrarily close to $(0, x_0)$, we get

$$\liminf_n \lambda_n^{\mathbf{B}_n} \ge x_0 \quad \text{a.s.}$$

But since $F^{\mathbf{B}_n} \xrightarrow{\mathscr{D}} F$ a.s., we must have

$$\limsup_n \lambda_n^{\mathbf{B}_n} \le x_0 \quad \text{a.s.}$$

Thus we get (6.3.1).

For general $\mathbf{T}_n$, let, for $\varepsilon > 0$ suitably small, $\mathbf{T}_n^\varepsilon$ denote the matrix by replacing all eigenvalues of $\mathbf{T}_n$ less than ε with ε. Let $H_n^\varepsilon = F^{\mathbf{T}_n^\varepsilon} = I_{[\varepsilon,\infty)} H_n$. Then $H_n^\varepsilon \xrightarrow{\mathscr{D}} H^\varepsilon \equiv I_{[\varepsilon,\infty)} H$. Let $\mathbf{B}_n^\varepsilon$ denote the sample covariance matrix corresponding to $\mathbf{T}_n^\varepsilon$.

Let $\hat{s}^\varepsilon$ denote the value where the maximum of $z_{y,H^\varepsilon}(s)$ occurs on $\mathbb{R}^+$. Then

$$\lambda_n^{\mathbf{B}_n^\varepsilon} \xrightarrow{\text{a.s.}} z_{y,H^\varepsilon}(\hat{s}^\varepsilon) \quad \text{as } n \to \infty. \tag{6.3.2}$$

Using Theorem A.46, we have

$$\begin{aligned} |\lambda_n^{\mathbf{B}_n^\varepsilon} - \lambda_n^{\mathbf{B}_n}| &= \left| \lambda_n^{(1/n)\mathbf{X}_n^* \mathbf{T}_n^\varepsilon \mathbf{X}_n} - \lambda_n^{(1/n)\mathbf{X}_n^* \mathbf{T}_n \mathbf{X}_n} \right| \\ &\le \|(1/n)\mathbf{X}_n^*(\mathbf{T}_n^\varepsilon - \mathbf{T}_n)\mathbf{X}_n\| \le \|(1/n)\mathbf{X}_n\mathbf{X}_n^*\|\varepsilon. \end{aligned} \tag{6.3.3}$$

Since $H^\varepsilon \xrightarrow{\mathscr{D}} H$ as $\varepsilon \to 0$, we get from Lemma 6.12

$$z_{y,H^\varepsilon}(\hat{s}^\varepsilon) \to z_{y,H}(\hat{s}) = x_0 \quad \text{as } \varepsilon \to 0. \tag{6.3.4}$$

Therefore, for ε sufficiently small, we see from (6.3.2)–(6.3.4) and the a.s. convergence of $\lambda_1^{(1/n)\mathbf{X}_n\mathbf{X}_n^*}$ (Theorem 5.11) that $\liminf_n \lambda_n^{\mathbf{B}_n} > 0$ a.s., which, as above, implies (6.3.1).

6.4 Proof of (3)

6.4.1 Convergence of a Random Quadratic Form

The goal of this section is to prove a limiting result on a random quadratic form involving the resolvent of $\mathbf{B}_n$.

Lemma 6.13. *Let u be any point in $[a,b]$ and $s = s_{F^{y,H}}(u)$. Let $\widetilde{\mathbf{x}} \in \mathbb{C}^p$ be distributed the same as $\mathbf{x}_1$ and independent of $\mathbf{X}_n$. Set $\mathbf{r} = (1/\sqrt{n})\mathbf{T}_n^{1/2}\widetilde{\mathbf{x}}$. Then*

$$\mathbf{r}^*(u\mathbf{I} - \mathbf{B}_n)^{-1}\mathbf{r} \xrightarrow{\text{a.s.}} 1 + 1/(us) \quad \textit{as } n \to \infty. \tag{6.4.1}$$

Proof. Let $\mathbf{B}_n^{n+1}$ denote $(1/n)\mathbf{T}_n^{1/2}\mathbf{X}_n^{n+1}\mathbf{X}_n^{n+1*}\mathbf{T}_n^{1/2}$, where $\mathbf{X}_n^{n+1} \equiv [\mathbf{X}_n, \widetilde{\mathbf{x}}]$, and $\underline{\mathbf{B}}_n^{n+1} = (1/n)\mathbf{X}_n^{n+1*}\mathbf{T}_n\mathbf{X}_n^{n+1}$. Let $z = u + iv_n$, $v_n > 0$. For Hermitian $\mathbf{A}$, let $s_\mathbf{A}$ denote the Stieltjes transform of the ESD of $\mathbf{A}$. Using Lemma 6.9, we have

$$|s_{\mathbf{B}_n}(z) - s_{\mathbf{B}_n^{n+1}}(z)| \le \frac{1}{nv_n}.$$

From (6.1.3) and its equivalent

$$s_{\underline{\mathbf{B}}_n^{n+1}}(z) = -\frac{1-p/(n+1)}{z} + (p/(n+1))s_{\mathbf{B}_n^{n+1}}(z)$$

for $\mathbf{B}_n^{n+1}$ and $\underline{\mathbf{B}}_n^{n+1}$, we conclude that

$$|s_{\underline{\mathbf{B}}_n}(z) - s_{\underline{\mathbf{B}}_n^{n+1}}(z)| \le \frac{(2y_n+1)}{v(n+1)}. \tag{6.4.2}$$

For $j = 1, 2, \cdots, n+1$, let $\mathbf{r}_j = (1/\sqrt{n})\mathbf{T}_n^{1/2}\mathbf{x}_j$ ($\mathbf{x}_j$ denoting the j-th column of $\mathbf{X}_n^{n+1}$) and $\mathbf{B}_{(j)} = \mathbf{B}_n^{n+1} - \mathbf{r}_j\mathbf{r}_j^*$. Notice $\mathbf{B}_{(n+1)} = \mathbf{B}_n$.

For $\underline{\mathbf{B}}_n^{n+1}$, (6.2.4) becomes

$$s_{\underline{\mathbf{B}}_n^{n+1}}(z) = -\frac{1}{n+1}\sum_{j=1}^{n+1}\frac{1}{z(1+\mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j)}. \tag{6.4.3}$$

Let

$$\mu_n(z) = -\frac{1}{z(1+\mathbf{r}^*(\mathbf{B}_n - z\mathbf{I})^{-1}\mathbf{r})},$$

where $\mathbf{r} = \mathbf{r}_{n+1} = (1/\sqrt{n})\mathbf{T}_n^{1/2}\widetilde{\mathbf{x}}$.

Our present goal is to show that for any $i \le n+1$, $\varepsilon > 0$, $z = z_n = u + iv_n$ with $v_n = n^{-\delta}$, $\delta \in [0, 1/3)$, and $\ell > 1$, we have for all n sufficiently large

$$\mathrm{P}(|s_{\underline{\mathbf{B}}_n(z)} - \mu_n(z)| > \varepsilon) \le K|z|^{2\ell}\varepsilon^{-2\ell}v_n^{-6\ell}n^{-\ell+1}. \tag{6.4.4}$$

We have from (6.4.3)

$$\begin{aligned}
&s_{\underline{\mathbf{B}}_n^{n+1}}(z) - \mu_n(z)\\
&= -\frac{1}{(n+1)z}\sum_{j=1}^{n}\left(\frac{1}{1+\mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j} - \frac{1}{1+\mathbf{r}^*(\mathbf{B}_n - z\mathbf{I})^{-1}\mathbf{r}}\right)\\
&= -\frac{1}{(n+1)z}\sum_{j=1}^{n}\frac{\mathbf{r}^*(\mathbf{B}_n - z\mathbf{I})^{-1}\mathbf{r} - \mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j}{(1+\mathbf{r}^*(\mathbf{B}_n - z\mathbf{I})^{-1}\mathbf{r})(1+\mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j)}.
\end{aligned}$$

Using (6.2.5), we find

$$|s_{\underline{\mathbf{B}}_n^{n+1}}(z) - \mu_n(z)| \le \frac{|z|}{v_n^2}\max_{j\le n}|\mathbf{r}^*(\mathbf{B}_n - z\mathbf{I})^{-1}\mathbf{r} - \mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j|. \tag{6.4.5}$$

Write

$$\mathbf{r}^*(\mathbf{B}_n - z\mathbf{I})^{-1}\mathbf{r} - \mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j$$

$$= \mathbf{r}^*(\mathbf{B}_n - z\mathbf{I})^{-1}\mathbf{r} - (1/n)\mathrm{tr}(\mathbf{B}_n - z\mathbf{I})^{-1}\mathbf{T}_n$$
$$-\big(\mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j - (1/n)\mathrm{tr}(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{T}_n\big)$$
$$+(1/n)\mathrm{tr}((\mathbf{B}_n - z\mathbf{I})^{-1} - (\mathbf{B}_{(j)} - z\mathbf{I})^{-1})\mathbf{T}_n.$$

Using Lemma 6.9, we find

$$(1/n)|\mathrm{tr}((\mathbf{B}_n - z\mathbf{I})^{-1} - (\mathbf{B}_{(j)} - z\mathbf{I})^{-1})\mathbf{T}_n| \le 2/(nv_n). \tag{6.4.6}$$

Using (6.2.2), we have, for any $j \le n+1$ and $\ell \ge 1$,

$$\mathrm{E}|\mathbf{r}_j^*(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{r}_j - (1/n)\mathrm{tr}\mathbf{T}_n^{1/2}(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{T}_n^{1/2}|^{2\ell}$$
$$\le K n^{-\ell} v_n^{-2\ell}. \tag{6.4.7}$$

Therefore, from (6.4.2) and (6.4.5)–(6.4.7), we get (6.4.4).

Setting $v_n = n^{-1/17}$, from (6.2.30) we have

$$s_{\underline{\mathbf{B}}_n}(u + iv_n) - s_{F^{y_n,H_n}}(u + iv_n) \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \to \infty.$$

Since $s_{F^{y_n,H_n}}(u + iv_n) \to s$ as $n \to \infty$, we have

$$s_{\underline{\mathbf{B}}_n}(u + iv_n) \xrightarrow{\text{a.s.}} s \quad \text{as } n \to \infty.$$

When $\ell > 34/11$, the bound in (6.4.4) is summable and we conclude that

$$|\mu_n(z_n) - s| \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \to \infty.$$

Therefore

$$|\mathbf{r}^*(z\mathbf{I} - \mathbf{B}_n)^{-1}\mathbf{r} - (1 + (1/us))| \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \to \infty. \tag{6.4.8}$$

Let d_n denote the distance between u and the nearest eigenvalue of $\mathbf{B}_n$. Then, because of (1), there exists a nonrandom $d > 0$ such that, almost surely, $\liminf_n d_n \ge d$.

When $d_n > 0$,

$$|\mathbf{r}^*(z\mathbf{I} - \mathbf{B}_n)^{-1}\mathbf{r} - \mathbf{r}^*(u\mathbf{I} - \mathbf{B}_n)^{-1}\mathbf{r}| \le \frac{v_n}{d_n^2}\frac{\widetilde{\mathbf{x}}^*\widetilde{\mathbf{x}}}{n}. \tag{6.4.9}$$

Using (6.2.2), we have for any $\varepsilon > 0$ and $\ell = 3$,

$$\mathrm{P}(|(1/p)\widetilde{\mathbf{x}}^*\widetilde{\mathbf{x}} - 1| > \varepsilon) \le K\frac{1}{\varepsilon^3}p^{-3/2},$$

which gives us

$$|(1/p)\widetilde{\mathbf{x}}^*\widetilde{\mathbf{x}} - 1| \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \to \infty. \tag{6.4.10}$$

Therefore, from (6.4.8)–(6.4.10), we get (6.4.1).

6.4.2 spread of eigenvaluesSpread of Eigenvalues

In this subsection, we assume the sequence $\{\mathbf{S}_n\}$ of Hermitian matrices is arbitrary except that their eigenvalues lie in the fixed interval $[d,e]$. To simplify the notation, we arrange the eigenvalues of $\mathbf{S}_n$ in nondecreasing order, denoting them as $s_1 \le \cdots \le s_p$. Our goal is to prove the following lemma.

Lemma 6.14. *For any $\varepsilon > 0$, we have for all M sufficiently large*

$$\limsup_{n\to\infty}\left(\lambda_1^{(1/n)\mathbf{Y}_n^*\mathbf{S}_n\mathbf{Y}_n} - \lambda_{[n/M]}^{(1/n)\mathbf{Y}_n^*\mathbf{S}_n\mathbf{Y}_n}\right) < \varepsilon \quad a.s., \tag{6.4.11}$$

where $\mathbf{Y}_n$ is $p\times[n/M]$ containing iid elements distributed the same as x_{11} ($[\,\cdot\,]$ denotes the greatest integer function). Moreover, the size of M depends only on ε and the endpoints d,e.

Proof. We first verify a basic inequality.

Lemma 6.15. *Suppose $\mathbf{A}$ and $\mathbf{B}$ are $p\times p$ Hermitian matrices. Then*

$$\lambda_1^{\mathbf{A}+\mathbf{B}} - \lambda_p^{\mathbf{A}+\mathbf{B}} \le \lambda_1^{\mathbf{A}} - \lambda_p^{\mathbf{A}} + \lambda_1^{\mathbf{B}} - \lambda_p^{\mathbf{B}}.$$

Proof. Let unit vectors $\mathbf{x},\mathbf{y}\in\mathbb{C}^p$ be such that $\mathbf{x}^*(\mathbf{A}+\mathbf{B})\mathbf{x} = \lambda_1^{\mathbf{A}+\mathbf{B}}$ and $\mathbf{y}^*(\mathbf{A}+\mathbf{B})\mathbf{y} = \lambda_p^{\mathbf{A}+\mathbf{B}}$. Then

$$\lambda_1^{\mathbf{A}+\mathbf{B}} - \lambda_p^{\mathbf{A}+\mathbf{B}} = \mathbf{x}^*\mathbf{A}\mathbf{x} + \mathbf{x}^*\mathbf{B}x - (\mathbf{y}^*\mathbf{A}\mathbf{y} + \mathbf{y}^*\mathbf{B}\mathbf{y}) \le \lambda_1^{\mathbf{A}} + \lambda_1^{\mathbf{B}} - \lambda_p^{\mathbf{A}} - \lambda_p^{\mathbf{B}}.$$

We continue now with the proof of Lemma 6.14. Since each $\mathbf{S}_n$ can be written as the difference between two nonnegative Hermitian matrices, because of Lemma 6.15 we may as well assume $d \ge 0$. Choose any positive α so that

$$\frac{e(e-d)}{\alpha} < \frac{\varepsilon}{24y}. \tag{6.4.12}$$

Choose any positive integer L_1 satisfying

$$\frac{\alpha}{L_1}(1+\sqrt{y})^2 < \frac{\varepsilon}{3}. \tag{6.4.13}$$

Choose any $M > 1$ so that

$$\frac{My}{L_1} > 1 \quad \text{and} \quad 4\sqrt{\frac{yL_1}{M}}e < \frac{\varepsilon}{3}. \tag{6.4.14}$$

Let

$$L_2 = \left[\frac{My}{L_1}\right] + 1. \tag{6.4.15}$$

Assume $p \ge L_1L_2$. For $k = 1,2,\cdots,L_1$, let

$$\ell_k = \{s_{[(k-1)p/L_1]+1}, \cdots, s_{[kp/L_1]}\},$$
$$\mathcal{L}_1 = \{\ell_k \,:\, s_{[kp/L_1]} - s_{[(k-1)p/L_1]+1} \le \alpha/L_1\}.$$

For any $\ell_k \notin \mathcal{L}_1$, define for $j = 1, 2, \cdots, L_2$,

$$\ell_{k\,j} = \{s_{[(k-1)p/L_1+(j-1)p/(L_1L_2)]+1}, \cdots, s_{[(k-1)p/L_1+jp/(L_1L_2)]}\},$$

and let $\mathcal{L}_2$ be the collection of all the latter sets. Notice that the number of elements in $\mathcal{L}_2$ is bounded by $L_1L_2(e-d)/\alpha$.

For $\ell \in \mathcal{L}_1 \cup \mathcal{L}_2$, write

$$\mathbf{S}_{n,\ell} = \sum_{s_i \in \ell} s_i e_i e_i^* \quad (e_i \text{ the unit eigenvector of } \mathbf{S}_n \text{ corresponding to } s_i),$$

$$\mathbf{A}_{n,\ell} = \sum_{s_i \in \ell} e_i e_i^*, \quad \overline{s}_\ell = \max_i\{s_i \in \ell\}, \quad \text{and} \quad \underline{s}_\ell = \min_i\{s_i \in \ell\}.$$

We have

$$\underline{s}_\ell \mathbf{Y}^*\mathbf{A}_{n,\ell}\mathbf{Y} \le \mathbf{Y}^*\mathbf{S}_{n,\ell}\mathbf{Y} \le \overline{s}_\ell \mathbf{Y}^*\mathbf{A}_{n,\ell}\mathbf{Y}, \tag{6.4.16}$$

where "$\le$" denotes partial ordering on Hermitian matrices (that is, $\mathbf{A} \le \mathbf{B}$ $\Longleftrightarrow$ $\mathbf{B} - \mathbf{A}$ is nonnegative definite).

Using Lemma 6.15 and (6.4.16), we have

$$\begin{aligned}
&\lambda_1^{(1/n)\mathbf{Y}_n^*\mathbf{S}_n\mathbf{Y}_n} - \lambda_{[n/M]}^{(1/n)\mathbf{Y}_n^*\mathbf{S}_n\mathbf{Y}_n} \\
\le &\sum_\ell \left[\lambda_1^{(1/n)\mathbf{Y}_n^*\mathbf{S}_{n,\ell}\mathbf{Y}_n} - \lambda_{[n/M]}^{(1/n)\mathbf{Y}_n^*\mathbf{S}_{n,\ell}\mathbf{Y}_n}\right] \\
\le &\sum_\ell \left[\overline{s}_\ell \lambda_1^{(1/n)\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n} - \underline{s}_\ell \lambda_{[n/M]}^{(1/n)\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n}\right] \\
= &\sum_\ell \overline{s}_\ell \left(\lambda_1^{(1/n)\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n} - \lambda_{[n/M]}^{(1/n)\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n}\right) + \sum_\ell (\overline{s}_\ell - \underline{s}_\ell)\lambda_{[n/M]}^{(1/n)\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n}.
\end{aligned}$$

From (6.4.15), we have

$$\lim_{n\to\infty} \frac{\left[\frac{p}{L_1L_2}\right]}{\left[\frac{n}{M}\right]} = \frac{My}{L_1L_2} < 1. \tag{6.4.17}$$

Therefore, for $\ell \in \mathcal{L}_2$, we have for all n sufficiently large

$$\operatorname{rank} \mathbf{A}_{n,\ell} \le \left[\frac{p}{L_1L_2}\right] + 1 < \left[\frac{n}{M}\right],$$

where we have used the fact that, for $a, r > 0$, $[a + r] - [a] = [r]$ or $[r] + 1$. This implies $\lambda_{[n/M]}^{(1/n)\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n} = 0$ for all large n. Thus, for these n,

$$\lambda_1^{(1/n)\mathbf{Y}_n^*\mathbf{S}_n\mathbf{Y}_n} - \lambda_{[n/M]}^{(1/n)\mathbf{Y}_n^*\mathbf{S}_n\mathbf{Y}_n} \le eL_1 \max_{\ell\in\mathcal{L}_1}\left(\lambda_1^{(1/n)\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n} - \lambda_{[n/M]}^{(1/n)\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n}\right)$$

$$+\frac{e(e-d)L_1L_2}{\alpha}\max_{\ell\in\mathcal{L}_2}\lambda_1^{(1/n)\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n} + \frac{\alpha}{L_1}\lambda_{[n/M]}^{(1/n)\mathbf{Y}_n^*\mathbf{Y}_n},$$

where for the last term we use the fact that, for Hermitian $\mathbf{C}_i$, $\sum\lambda_{\min}^{\mathbf{C}_i} \le \lambda_{\min}^{\sum\mathbf{C}_i}$.

We have with probability 1

$$\lambda_{[n/M]}^{(1/[n/M])\mathbf{Y}_n^*\mathbf{Y}_n} \longrightarrow (1-\sqrt{My})^2.$$

Therefore, from (6.4.13) we have almost surely

$$\lim_{n\to\infty}\frac{\alpha}{L_1}\lambda_{[n/M]}^{(1/n)\mathbf{Y}_n^*\mathbf{Y}_n} < \frac{\varepsilon}{3}.$$

We have

$$F^{\mathbf{A}_{n,\ell}} = \left(1-\frac{|\ell|}{p}\right)I_{[0,\infty)} + \frac{|\ell|}{p}I_{[1,\infty)},$$

where $|\ell|$ is the size of ℓ, and from the expression for the inverse of the Stieltjes transform of the limiting distribution it is a simple matter to show

$$F^{p/[n/M],F^{\mathbf{A}_{n,\ell}}} = F^{|\ell|/[n/M],I_{[1,\infty)}}.$$

For $\ell\in\mathcal{L}_1$, we have

$$F^{\mathbf{A}_{n,\ell}} \xrightarrow{\mathscr{D}} \left(1-\frac{1}{L_1}\right)I_{[0,\infty)} + \frac{1}{L_1}I_{[1,\infty)} \equiv G.$$

From Corollary 6.6, the first inequality in (6.4.14), and conclusion (2), we have the extreme eigenvalues of $(1/[n/M])\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n$ converging a.s. to the extreme values in the support of $F^{My,G} = F^{(My)/L_1,I_{[1,\infty)}}$. Therefore, from Theorem 5.11 we have with probability 1

$$\lambda_1^{(1/[n/M])\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n} - \lambda_{[n/M]}^{(1/[n/M])\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n} \longrightarrow 4\sqrt{\frac{My}{L_1}},$$

and from the second inequality in (6.4.14) we have almost surely

$$\lim_{n\to\infty} eL_1\max_{\ell\in\mathcal{L}_1}\left(\lambda_1^{(1/n)\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n} - \lambda_{[n/M]}^{(1/n)\mathbf{Y}_n^*\mathbf{A}_{n,\ell}\mathbf{Y}_n}\right) < \frac{\varepsilon}{3}.$$

Finally, from (6.4.17) we see that, for $\ell\in\mathcal{L}_2$, $\lim_{n\to\infty}|\ell|/[n/M] < 1$, so that from (6.4.12), the first inequality in (6.4.14), and Corollary 6.6 we have with probability 1

$$\lim_{n\to\infty} \frac{e(e-d)L_1L_2}{\alpha} \max_{\ell\in\mathcal{L}_2} \lambda_1^{(1/n)Y_n^*\mathbf{A}_{n,\ell}Y_n} < \frac{e(e-d)}{\alpha} L_1L_2\frac{4}{M} < \frac{\varepsilon}{3}.$$

This completes the proof of Lemma 6.14.

6.4.3 Dependence on y

We now finish the proof of Lemma 6.2. The following relies on Lemma 6.1 and (6.1.6), the explicit form of $z_{y,H}$.

For (a), we have $(t_1, t_2) \subset S'_H$ with $t_1, t_2 \in \partial S_H$ and $t_1 > 0$. On $(-t_1^{-1}, -t_2^{-1})$, $z_{y,H}(s)$ is well defined, and its derivative is positive if and only if

$$g(s) \equiv \int \left(\frac{ts}{1+ts}\right)^2 dH(t) < \frac{1}{y}.$$

It is easy to verify that $g''(s) > 0$ for all $s \in (-t_1^{-1}, -t_2^{-1})$. Let $\hat{s}$ be the value in $[-t_1^{-1}, -t_2^{-1}]$ where the minimum of $g(s)$ occurs, the two endpoints being included in case $g(s)$ has a finite limit at either value. Write $y_0 = 1/g(\hat{s})$. Then, for any $y < y_0$, the equation $yg(s) = 1$ has two solutions in the interval $[-t_1^{-1}, -t_2^{-1}]$, denoted by $s_y^1 < s_y^2$. Then, $s \in (s_y^1, s_y^2) \Leftrightarrow yg(s) < 1 \Leftrightarrow z'_{y,H}(s) > 0$. By Lemma 6.1, this is further equivalent to $(z_{y,H}(s_y^1), z_{y,H}(s_y^2)) \subset S'_{F^{y,H}}$, with endpoints lying in the boundary of $S_{F^{y,H}}$. From the identity (6.1.6), we see that, for $i = 1, 2$,

$$\begin{aligned} z_{y,H}(s_y^i) &= \frac{1}{s}(yg(s_y^i) - 1) + y\int \frac{t}{(1+ts_y^i)^2} dH(t) \\ &= y\int \frac{t}{(1+ts_y^i)^2} dH(t) > 0. \end{aligned} \tag{6.4.18}$$

As y decreases to zero, we have $s_y^1 \downarrow -t_1^{-1}$, $s_y^2 \uparrow -t_2^{-1}$, which also includes the possibility that either endpoint will reach its limit for positive y (when $g(s)$ has a limit at an endpoint). We now show (6.1.8) for $i = 1, 2$. If eventually $s_y^i = -t_i^{-1}$, then clearly (6.1.8) holds. Otherwise we must have $yg(s_y^i) = 1$, and so by the Cauchy-Schwarz inequality,

$$y\left|\int \left(\frac{ts_y^i}{1+ts_y^i}\right) dH(t)\right| \le y\left(\int \left(\frac{ts_y^i}{1+ts_y^i}\right)^2 dH(t)\right)^{1/2} = y^{1/2},$$

and so again (6.1.8) holds.

It is straightforward to show

$$\frac{dz_{y,H}(s_y^i)}{dy} = \int \frac{t}{1+ts_y^i} dH(t). \tag{6.4.19}$$

Since $(1+ts)(1+ts') > 0$ for $t \in S_H$ and $s, s' \in (-t_1^{-1}, -t_2^{-1})$, we get from (6.4.19)

$$\frac{d(z_{y,H}(s_y^2) - z_{y,H}(s_y^1))}{dy} = (s_y^1 - s_y^2) \int \frac{t^2}{(1+ts_y^2)(1+ts_y^1)} dH(t) < 0.$$

Therefore

$$z_{y,H}(s_y^2) - z_{y,H}(s_y^1) \uparrow t_2 - t_1 \quad \text{as } y \downarrow 0.$$

As $y \downarrow y_0 = g(\hat{s})$, the minimum of $g(s)$, we see that s_y^1 and s_y^2 approach $\hat{s}$ and so the interval $(z_{y,H}(s_y^1), z_{y,H}(s_y^2))$ shrinks to a point. This establishes (a).

We have a similar argument for (b), where now $s_y^3 \in [-1/t_3, 0)$ such that $z'_{y,H}(s) > 0$ for $s \in (-1/t_3, 0) \iff s \in (s_y^3, 0)$. Since $z_{y,H}(s) \to \infty$ as $s \uparrow 0$, we have $(z_{y,H}(s_y^3), \infty) \subset S'_{F^{y,H}}$ with $z_{y,H}(s_y^3) \in \partial S_{F^{y,H}}$. Equation (6.4.19) holds also in this case, and from it and the fact that $(1+ts) > 0$ for $t \in S_H$, $s \in (-1/t_3, 0)$, we see that boundary point $z_{y,H}(s_y^3) \downarrow t_3$ as $y \to 0$. On the other hand, $s_y^3 \uparrow 0$ and, consequently, $z_{y,H}(s_y^3) \uparrow \infty$ as $y \uparrow \infty$. Thus we get (b).

When $y[1-H(0)] < 1$, as s increases from $-\infty$ to $-1/t_4$, $yg(s)$ increases from $y[1-H(0)] < 1$ to ∞. Thus, we can find a unique $s_y^4 \in (-\infty, -1/t_4]$ such that $yg(s_y^4) = 1$. So, on the interval $(-\infty, s_y^4)$, $z'_{y,H}(s) > 0$. Since $z_{y,H}(s) \downarrow 0$ as $s \downarrow -\infty$, we have $(0, z_{y,H}(s_y^4)) \in S^c_{F^{y,H}}$ with $z_{y,H}(s_y^4) \in \partial S_{F^{y,H}}$. From (6.4.19), we have $z_{y,H}(s_y^4) \uparrow t_4$ as $y \downarrow 0$. Since $g(s)$ is increasing on $(-\infty, -1/t_4)$, we have $s_y^4 \downarrow -\infty$, and consequently $z_{y,H}(s_y^4) \downarrow 0$ as $y \uparrow [1-H(0)]^{-1}$. Therefore we get (c).

When $y[1-H(0)] > 1$, $g(s)$ increases from 0 to $y[1-H(0)]$ as s increases from 0 to ∞. Thus, there is a unique s_y such that $yg(s_y) = 1$. When $s \in (0, s_y)$, $g(s) < 1$, and hence $z_{y,H}(s)$ is strictly increasing from $-\infty$ to $x_0 := z_{y,H}(s_y)$. And then $z_{y,H}(s)$ strictly decreases from x_0 to 0 as s increases from s_y to ∞. Thus, $x_0 > 0$, which is the smallest value of the support of $F^{y,H}$. It can be verified that (6.4.19) is also true for s_y. Since its right-hand side is always positive, x_0 is strictly increasing as y increases. Subsequently, from (6.4.18), $x_0 = z_{y,H}(s_y)$ ranges from 0 to ∞ as y increases from 0 to ∞, which completes (d).

(e) is obvious since $z_{y,I_{[0,\infty)}} = -1/s$ for all $s \neq 0$ and so $s_{F^{y,I_{[0,\infty)}}}(z) = -1/z$, the Stieltjes transform of $I_{[0,\infty)}$.

From Lemma 6.1, we can only get intervals in $S^c_{F^{y,H}}$ from intervals arising from (a)–(e). By (6.1.6), for any two solutions $s_y > s'_y$ (note that they may not be in the same interval of S^c_H) of the equation $yg(s) = 1$,

$$z_{y,H}(s_y) - z_{y,H}(s'_y) = \frac{s_y - s'_y}{s_y s'_y}\left[1 - y\int \frac{t^2 s_y s'_y}{(1+ts_y)(1+ts'_y)} dH(t)\right] \geq 0, \tag{6.4.20}$$

where the last step follows by the Cauchy-Schwarz inequality and the fact that both s_y and s'_y are solutions of the equation $yg(s) = 1$. If the above is not strict, then for all $t \in S_H$,

$$\frac{ts_y}{1+ts_y} = a\frac{ts'_y}{1+ts'_y}$$

for some constant a. If S_H contains at least two points, then the identity above implies that

$$a = 1 \quad \text{and} \quad s_y = s'_y,$$

which contradicts the assumption $s_y > s'_y$. If S_H contains only one point, say t_0, then the same identity as well as the definition of s_y imply that

$$1 = \frac{t_0^2 s_y s'_y}{(1+t_0 s_y)(1+t_0 s'_y)} = \frac{t_0^2 s_y^2}{(1+t_0 s_y)^2}.$$

This also implies the contradiction $s_y = s'_y$. Thus, inequality (6.4.20) is strict and hence the last statement in Lemma 6.2 follows. The proof of Lemma 6.2 is complete.

We finish this section with a lemma important to the final steps in the proof of (3).

Lemma 6.16. *If the interval $[a, b]$ satisfies condition (f) of Theorem 6.3 for $y_n \to y$, then for any $\hat{y} < y$ and sequence $\{\hat{y}_n\}$ converging to $\hat{y}$, the interval $[z_{\hat{y},H}(s_{F^{y,H}}(a)), z_{\hat{y},H}(s_{F^{y,H}}(b))]$ satisfies assumption (f) of Theorem 6.3 for $\hat{y}_n \to \hat{y}$. Moreover, its length increases from $b - a$ as $\hat{y}$ decreases from y.*

Proof. According to (f), there exists an $\varepsilon > 0$ such that $[a-\varepsilon, b+\varepsilon] \subset S^c_{F^{y_n,H_n}}$ for all large n. From Lemma 6.1, we have for these n

$$\begin{aligned}&[s_{F^{y,H}}(a-\varepsilon), s_{F^{y,H}}(b+\varepsilon)] \subset \mathbf{A}_{y_n,H_n}\\ &\equiv \{s \in \mathbb{R} : s \neq 0, -s^{-1} \in S^c_{H_n}, z'_{y_n,H_n}(s) > 0\}.\end{aligned}$$

Since $z'_{y,H}(s)$ increases as y decreases, $[s_{F^{y,H}}(a-\varepsilon), s_{F^{y,H}}(b+\varepsilon)]$ is also contained in $\mathbf{A}_{\hat{y}_n,H_n}$. Therefore, by Lemma 6.1,

$$(z_{\hat{y},H}(s_{F^{y,H}}(a-\varepsilon)), z_{\hat{y},H}(s_{F^{y,H}}(b+\varepsilon))) \subset S^c_{F^{\hat{y}_n,H_n}}.$$

Since $z_{\hat{y},H}$ and $s_{F^{y,H}}$ are monotonic on, respectively, $(s_{F^{y,H}}(a-\varepsilon), s_{F^{y,H}}(b+\varepsilon))$ and $(a-\varepsilon, b+\varepsilon)$, we have

$$[z_{\hat{y},H}(s_{F^{y,H}}(a)), z_{\hat{y},H}(s_{F^{y,H}}(b))] \subset (z_{\hat{y},H}(s_{F^{y,H}}(a-\varepsilon)), z_{\hat{y},H}(s_{F^{y,H}}(b+\varepsilon))),$$

so assumption (f) is satisfied.

Since $z'_{\hat{y}',H}(s) > z'_{\hat{y},H}(s) > z'_{y,H}(s)$ for $\hat{y}' < \hat{y}$, we have

$$z_{\hat{y}',H}(s_{F^{y,H}}(b)) - z_{\hat{y}',H}(s_{F^{y,H}}(a)) > z_{\hat{y},H}(s_{F^{y,H}}(b)) - z_{\hat{y},H}(s_{F^{y,H}}(a))$$

$$> z_{y,H}(s_{F^{y,H}}(b)) - z_{y,H}(s_{F^{y,H}}(a)) = b - a.$$

6.4.4 Completing the Proof of (3)

We begin with some basic lemmas. For the following, $\mathbf{A}$ is assumed to be a $p \times p$ Hermitian matrix, $\lambda \in \mathbb{R}$ is not an eigenvalue of $\mathbf{A}$, and $\mathbf{Y}$ is any matrix with p rows.

Lemma 6.17. *λ is an eigenvalue of $\mathbf{A} + \mathbf{Y}\mathbf{Y}^*$ $\iff$ $\mathbf{Y}^*(\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}$ has eigenvalue* 1.

Proof. Suppose $\mathbf{x} \in \mathbb{C}^p \backslash \{0\}$ is such that $(\mathbf{A} + \mathbf{Y}\mathbf{Y}^*)\mathbf{x} = \lambda\mathbf{x}$. It follows that $\mathbf{Y}^*\mathbf{x} \neq 0$ and

$$\mathbf{Y}^*(\lambda I - \mathbf{A})^{-1}\mathbf{Y}\mathbf{Y}^*\mathbf{x} = \mathbf{Y}^*\mathbf{x}$$

so that $\mathbf{Y}^*(\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}$ has eigenvalue 1 (with eigenvector $\mathbf{Y}^*\mathbf{x}$).

Suppose $\mathbf{Y}^*(\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}$ has eigenvalue 1 with eigenvector $\mathbf{z}$. Then $(\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}\mathbf{z} \neq 0$ and

$$(\mathbf{A} + \mathbf{Y}\mathbf{Y}^*)(\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}\mathbf{z} = -\mathbf{Y}z + \lambda(\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}\mathbf{z} + \mathbf{Y}\mathbf{z} = \lambda(\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}\mathbf{z}.$$

Thus $\mathbf{A} + \mathbf{Y}\mathbf{Y}^*$ has eigenvalue λ (with eigenvector $(\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}\mathbf{z}$).

Lemma 6.18. *Suppose $\lambda_j^{\mathbf{A}} < \lambda$. If $\lambda_1^{\mathbf{Y}^*(\lambda\mathbf{I}-\mathbf{A})^{-1}\mathbf{Y}} < 1$, then $\lambda_j^{\mathbf{A}+\mathbf{Y}\mathbf{Y}^*} < \lambda$.*

Proof. Suppose $\lambda_j^{\mathbf{A}+\mathbf{Y}\mathbf{Y}^*} \geq \lambda$. Then, since $\lambda_j^{\mathbf{A}+\alpha\mathbf{Y}\mathbf{Y}^*}$ is continuously increasing in $\alpha \in \mathbb{R}^+$ (Corollary 4.3.3 of Horn and Johnson [154]), there is an $\alpha \in (0, 1]$ such that $\lambda_j^{\mathbf{A}+\alpha\mathbf{Y}\mathbf{Y}^*} = \lambda$. Therefore, from Lemma 6.17, $\alpha\mathbf{Y}^*(\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}$ has eigenvalue 1, which means $\mathbf{Y}^*(\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}$ has an eigenvalue ≥ 1.

Lemma 6.19. *For any $i \in \{1, 2, \cdots, p\}$, $\lambda_1^{\mathbf{A}} \leq \lambda_1^{\mathbf{A}} - \lambda_p^{\mathbf{A}} + \mathbf{A}_{ii}$.*

Proof. Simply use the fact that $\mathbf{A}_{ii} \geq \lambda_p^{\mathbf{A}}$.

We now complete the proof of (3). Because of the conditions of (3) and Lemma 6.2, we may assume $s_{F^{y,H}}(b) < 0$. For $M > 0$ (its size to be determined later), let $y^j = y/(1 + j/M)$ for $j = 0, 1, 2, \cdots$, and define the intervals

$$[a^j, b^j] = [z_{y^j,H}(s_{F^{y,H}}(a)), z_{y^j,H}(s_{F^{y,H}}(b))].$$

By Lemma 6.16, these intervals increase in length as j increases, and, for each j, the interval together with y^j satisfies assumption (f) for any sequence y_n^j converging to y^j. Here we take

$$y_n^j = \frac{p}{n + j[n/M]}.$$

Let $s_a = s_{F^{y,H}}(a)$. We have

$$a^j - a = z_{y^j,H}(s_a) - z_{y,H}(s_a) = (y^j - y) \int \frac{t}{1 + ts_a} dH(t).$$

Therefore, for each j,

$$a^j \le \hat{a} \equiv a + y \left| \int \frac{t}{1 + ts_a} dH(t) \right|.$$

We also have

$$a^{j+1} - a^j = z_{y^{j+1},H}(s_a) - z_{y^j,H}(s_a) = (y^{j+1} - y^j) \int \frac{t}{1 + ts_a} dH(t).$$

Thus, we can find an $M_1 > 0$ so that, for any $M \ge M_1$ and any j,

$$|a^{j+1} - a^j| < \frac{b - a}{4}. \tag{6.4.21}$$

Let $M_2 \ge M_1$ be such that, for all $M \ge M_2$,

$$\frac{1}{1 + 1/M} > \frac{3}{4} + \frac{1}{4} \frac{\hat{a}}{b - a + \hat{a}}.$$

This will ensure that, for all n, $j \ge 0$ and $M \ge M_2$,

$$\frac{n + j[n/M]}{n + (j+1)[n/M]} b^j > b^j - \frac{(b^j - a^j)}{4}. \tag{6.4.22}$$

From Lemma 6.14, we can find an $M_3 \ge M_2$ such that, for all $M \ge M_3$, (6.4.11) is true for any sequence of $\mathbf{S}_n$ with

$$d = -\frac{4}{3(b - a)}, \quad e = \frac{4}{b - a}, \quad \text{and } \varepsilon = \frac{1}{\hat{a}|s_a|}.$$

We now fix $M \ge M_3$.

For each j, let

$$\mathbf{B}_n^j = \frac{1}{n + j[n/M]} \mathbf{T}_n^{1/2} \mathbf{X}_n^{n+j[n/M]} \mathbf{X}_n^{n+j[n/M]*} \mathbf{T}_n^{1/2},$$

where $\mathbf{X}_n^{n+j[n/M]} \equiv (x_{i\,k}^{n,j})$, $i = 1, 2, \cdots, p$, $k = 1, 2, \ldots, n + j[n/M]$, are defined on a common probability space, entries iid distributed as $x_{1\,1}$ (no relation assumed for different n, j).

Since a^j and b^j can be made arbitrarily close to $-1/s_{F^{y,H}(a)}$ and $-1/s_{F^{y,H}(b)}$, respectively, by making j sufficiently large, we can find a K_1 such that, for all $K \ge K_1$,

$$\lambda_{i_n+1}^{\mathbf{T}_n} < a^K \quad \text{and} \quad b^K < \lambda_{i_n}^{\mathbf{T}_n} \quad \text{for all large } n.$$

Therefore, using (6.1.1) and Theorem 5.11, we can find a $K \geq K_1$ such that with probability 1

$$\limsup_{n\to\infty} \lambda_{i_n+1}^{\mathbf{B}_n^K} < a^K \quad \text{and} \quad b^K < \liminf_{n\to\infty} \lambda_{i_n}^{\mathbf{B}_n^K}. \tag{6.4.23}$$

We fix this K.

Let

$$E_j = \{ \text{ no eigenvalue of } \mathbf{B}_n^j \text{ appears in } [a^j, b^j] \text{ for all large } n \ \}.$$

Let

$$\ell_n^j = \begin{cases} k, & \text{if } \lambda_k^{\mathbf{B}_n^j} > b^j,\ \lambda_{k+1}^{\mathbf{B}_n^j} < a^j, \\ -1, & \text{if there is an eigenvalue of } \mathbf{B}_n^j \text{ in } [a^j, b^j]. \end{cases}$$

For notational convenience, let $\lambda_{-1}^{\mathbf{A}} = \infty$ for Hermitian $\mathbf{A}$.

Define

$$\hat{a}^j = a^j + \frac{1}{4}(b^j - a^j),$$
$$\hat{b}^j = b^j - \frac{1}{4}(b^j - a^j).$$

Fix $j \in \{0, 1, \cdots, K-1\}$. On the same probability space, we define for each $n \geq M$, $\mathbf{Y}_n = (Y_{i\,k})$, $i = 1, 2, \cdots, p$, $k = 1, \cdots, [n/M]$, entries iid distributed the same as x_{11}, with $\{\mathbf{B}_n^j\}_n$ and $\{\mathbf{Y}_n\}_n$ independent (no restriction on $\mathbf{Y}_n$ for different n). Let $\mathbf{R}_n = \mathbf{T}_n^{1/2}\mathbf{Y}_n$.

Whenever $\hat{a}^j$ is not an eigenvalue of $\mathbf{B}_n^j$, we have by Lemma 6.19

$$\begin{aligned} &\lambda_1^{\frac{1}{n+j[n/M]}\mathbf{R}_n^*(\hat{a}^j\mathbf{I}-\mathbf{B}_n^j)^{-1}\mathbf{R}_n} \\ &\leq \lambda_1^{\frac{1}{n+j[n/M]}\mathbf{R}_n^*(\hat{a}^j\mathbf{I}-\mathbf{B}_n^j)^{-1}\mathbf{R}_n} - \lambda_{[n/M]}^{\frac{1}{n+j[n/M]}\mathbf{R}_n^*(\hat{a}^j\mathbf{I}-B_n^j)^{-1}\mathbf{R}_n} \\ &\quad + \left(\frac{1}{n+j[n/M]}\mathbf{R}_n^*(\hat{a}^j I - \mathbf{B}_n^j)^{-1}\mathbf{R}_n\right)_{1\,1}. \end{aligned} \tag{6.4.24}$$

If $\hat{a}^j$ is not an eigenvalue of $\mathbf{B}_n^j$ for all large n, we get from Lemma 6.13

$$\begin{aligned} &\left(\frac{1}{n+j[n/M]}\mathbf{R}_n^*(\hat{a}^j\mathbf{I} - \mathbf{B}_n^j)^{-1}\mathbf{R}_n\right)_{1\,1} \\ &\xrightarrow{\text{a.s.}} 1 + \frac{1}{\hat{a}^j s_{F^{y^j,H}}(\hat{a}^j)} < 1 + \frac{1}{\hat{a} s_a} \quad \text{as } n \to \infty, \end{aligned} \tag{6.4.25}$$

and from Lemma 6.14

$$\limsup_{n\to\infty}\left(\lambda_1^{\frac{1}{n+j[n/M]}\mathbf{R}_n^*(\hat{a}^j\mathbf{I}-\mathbf{B}_n^j)^{-1}\mathbf{R}_n} - \lambda_{[n/M]}^{\frac{1}{n+j[n/M]}\mathbf{R}_n^*(\hat{a}^j\mathbf{I}-\mathbf{B}_n^j)^{-1}\mathbf{R}_n}\right) < \frac{1}{\hat{a}|s_a|} \quad \text{a.s.} \tag{6.4.26}$$

Now, (6.4.24)–(6.4.26) hold for a fixed realization in E_j with respect to the probability measure on $\{\mathbf{Y}_n\}_n$. By Fubini's Theorem and the fact that $\mathrm{P}(E_j) = 1$ (from (1)), we subsequently have (6.4.24)–(6.4.26) holding on the probability space generating $\{\mathbf{B}_n^j\}_n$ and $\{\mathbf{Y}_n\}_n$. Therefore we find

$$\mathrm{P}\left(\lambda_1^{\frac{1}{n+j[n/M]}\mathbf{R}_n^*(\hat{a}^j\mathbf{I}-\mathbf{B}_n^j)^{-1}\mathbf{R}_n} < 1 \quad \text{for all large } n\right) = 1,$$

and since $\mathbf{B}_n^j + \frac{1}{n+j[n/M]}\mathbf{R}_n\mathbf{R}_n^* \sim \frac{n+(j+1)[n/M]}{n+j[n/M]}\mathbf{B}_n^{j+1}$, we get from Lemma 6.18, with probability 1,

$$\lambda_{\ell_n^j+1}^{\mathbf{B}_n^{j+1}} < \hat{a}^j \quad \text{for all large } n.$$

Since $\lambda_{\ell_n^j}^{\mathbf{B}_n^j} \le \lambda_{\ell_n^j}^{\mathbf{B}_n^j + \frac{1}{n+j[n/M]}\mathbf{R}_n\mathbf{R}_n^*}$, we use (6.4.22) to get

$$\mathrm{P}\left(\lambda_{\ell_n^j}^{\mathbf{B}_n^{j+1}} > \hat{b}^j \quad \text{and} \quad \lambda_{\ell_n^j+1}^{\mathbf{B}_n^{j+1}} < \hat{a}^j \quad \text{for all large } n\right) = 1.$$

From (6.4.21) we see that $[\hat{a}^j, \hat{b}^j] \subset [a^{j+1}, b^{j+1}]$. Therefore, combining the event above with E_{j+1}, we conclude that

$$\mathrm{P}\left(\lambda_{\ell_n^j}^{\mathbf{B}_n^{j+1}} > b^{j+1} \quad \text{and} \quad \lambda_{\ell_n^j+1}^{\mathbf{B}_n^{j+1}} < a^{j+1} \quad \text{for all large } n\right) = 1.$$

Therefore, with probability 1, for all large n $[a, b]$ and $[a^K, b^K]$ split the eigenvalues of, respectively, $\mathbf{B}_n$ and $\mathbf{B}_n^K$, having equal amounts on the left-hand sides of the intervals. Finally, from (6.4.23), we get (3).

Chapter 7
Semicircular Law for Hadamard Products

7.1 Sparse Matrix and Hadamard Product

In nuclear physics, since the particles move with very high velocity in a small range, many excited states are seldom observed in very short time instances, and over long time periods there are no excitations. More generally, if a real physical system is not of full connectivity, the random matrix describing the interactions between the particles in the system will have a large proportion of zero elements. In this case, a sparse random matrix provides a more natural and relevant description of the system. Indeed, in neural network theory, the neurons in a person's brain are large in number and are not of full connectivity with each other. Actually, the dendrites connected with one individual neuron are of much smaller number, probably several orders of magnitude, than the total number of neurons. Sparse random matrices are adopted in modeling these partially connected systems in neural network theory.

A sparse or dilute matrix is a random matrix in which some entries will be replaced by 0 if not observed. Sometimes a large portion of entries of the interesting random matrix can be 0's. Due to their special application background, sparse matrices have received special attention in quantum mechanics, atomic physics, neural networks, and many other areas. Some recent works on large sparse matrices and their applications to various areas include, among others, [45, 61, 285] (linear algebra), [48] (neural networks), [62, 89, 143, 197, 218, 292] (algorithms and computing), [207] (financial modeling), [211] (electrical engineering), [216] (biointeractions), and [176, 271] (theoretical physics).

A sparse matrix can be expressed by the Hadamard product (see Section A.3). Let $\mathbf{B}_m = (b_{ij})$ and $\mathbf{D}_m = (d_{ij})$ be two $m \times m$ matrices. Then the Hadamard product $\mathbf{A}_m = (a_{ij})$ with $a_{ij} = b_{ij} d_{ij}$ is denoted by

$$\mathbf{A}_p = \mathbf{B}_m \circ \mathbf{D}_m.$$

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_7,

A matrix $\mathbf{A}_m$ is sparse if the elements d_{ij} of $\mathbf{D}_m$ take values 0 and 1 with $\sum_{i=1}^m P(d_{ij} = 1) = p = o(m)$. The index p usually stands for the level of sparseness; i.e., after performing the Hadamard product, the resulting matrix will have p nonzero elements per row on average.

Several of the papers mentioned above consider a sparse matrix resulting from the removal of entries of a sample covariance matrix. Removing the assumption that the elements of $\mathbf{D}_m$ are Bernoulli trials, this chapter will consider the LSD of general Hadamard products of a normalized sample covariance matrix with a diluted matrix. We shall show that its ESD converges to the semicircular law under certain conditions.

To this end, we make the following assumptions. We remind the reader that the entries of $\mathbf{D}_m$ and $\mathbf{X}_n$ are allowed to depend on n. For brevity, the dependence on n is suppressed.

Assumptions on $\mathbf{D}_m$:

(D1) $\mathbf{D}_m$ is Hermitian.

(D2) $\displaystyle\sum_{i=1}^m p_{ij} = p + o(p)$ uniformly in j, where $p_{ij} = E|d_{ij}^2|$.

(D3) For some $M_2 > M_1 > 0$,

$$\max_j \sum_i \left(\mathrm{E}|d_{ij}|^2 I[|d_{ij}| > M_2] + P(0 < |d_{ij}| < M_1)\right) = o(1) \quad (7.1.1)$$

as $m \to \infty$.

Assumptions on $\mathbf{X}_n$:

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

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

(X2, 1) $\sum_{u=1}^{\infty} \frac{1}{mn}\sum_{ij} \mathrm{E}|x_{ij}^2| I[|x_{ij}| > \eta\sqrt[4]{np}] < \infty$ for any fixed $\eta > 0$,

where u may take $[p]$, m, or n.

(X3) For any $\eta > 0$, $\frac{1}{m}\sum_{i=1}^m P\left[\left|\sum_{k=1}^n (|x_{ik}|^2 - \sigma^2) d_{ii}\right| > \eta\sqrt{np}\right] \to 0. \quad (7.1.2)$

We shall prove the following theorem.

Theorem 7.1. *Assume that conditions (7.1.1) and (7.1.2) hold and that the entries of the matrix* $\mathbf{D}_m$ *are independent of those of the matrix* $\mathbf{X}_n$ $(m \times n)$. *Also, we assume that* $p/n \to 0$ *and* $p \to \infty$.

Then, the ESD $F^{\mathbf{A}_p}$ *tends to the semicircular law as* $[p] \to \infty$*, where* $\mathbf{A}_p = \frac{1}{\sqrt{np}}(\mathbf{X}_n\mathbf{X}_n^* - \sigma^2 n\mathbf{I}_m) \circ \mathbf{D}_m$. *The convergence is in probability if the condition (X2,0) is assumed and the convergence is in the sense of almost* *surely for $[p] \to \infty$ or $m \to \infty$ if condition (X2,1) is assumed for $u = [p]$ or $u = m$, respectively.*

Remark 7.2. 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. If we consider convergence in probability, Theorem 7.1 is true for $p \to \infty$.

Remark 7.3. Conditions (D2) and (D3) imply that $p \le Km$; 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 has to have a lower order than n. This is essential. However, the relation between m and n can be arbitrary.

Remark 7.4. From the proofs given in Sections 7.2 and 7.3, one can see that a.s. convergence is true for $m \to \infty$ in all places except the part involving truncation on the entries of $\mathbf{X}_n$, which was guaranteed by condition (X2,1). Thus, if condition (X2,1) is true for $u = m$, then a.s. convergence is in the sense of $m \to \infty$. Sometimes, it may be of interest to consider a.s. convergence in the sense of $n \to \infty$. Examining the proofs given in Sections 7.2 and 7.3, one finds that to guarantee a.s. convergence for $n \to \infty$, truncation on the entries of $\mathbf{D}_m$ and the removal of diagonal elements require $m/\log n \to \infty$; truncation on the entries of $\mathbf{X}_n$ requires condition (X2,1) be true for $u = n$. As for Theorem 7.1, as remarked in Section 7.3, one may modify the conclusion of (II) to

$$\mathrm{E}|\beta_{nk} - \mathrm{E}\beta_{nk}|^{2\mu} = O(m^{-\mu})$$

for any fixed integer μ, where β_{nk} is defined in Section 7.3. Thus, if $m \ge n^{\delta}$ for some positive constant δ, then a.s. convergence for the ESD after truncation and centralization is true for $n \to \infty$. Therefore, the conclusion of Theorem 7.1 can be strengthened to a.s. convergence as $n \to \infty$ under the additional assumptions that $m \ge n^{\delta}$ and condition (X2,1) is for $u = n$.

Remark 7.5. In Theorem 7.1, if $p = m$ and $d_{ij} \equiv 1$ for all i and j and the entries of $\mathbf{X}_n$ are iid, then the model considered in Theorem 7.1 reduces to that of Bai and Yin [37], where the entries $\mathbf{X}_n$ are assumed to be iid with finite fourth moments. It can be easily verified that the conditions of Theorem 7.1 are satisfied under Bai and Yin's assumption. Thus, Theorem 7.1 contains Bai and Yin's result as a special case.

A slightly different generalization of Bai and Yin's result is the following.

Theorem 7.6. *Suppose that, for each n, the entries of the matrix $\mathbf{X}_n$ are independent complex random variables with a common mean value and variance σ^2. Assume that, for any constant $\delta > 0$,*

$$\frac{1}{p\sqrt{np}} \sum_{jk} \mathrm{E}|x_{jk}^2| I[|x_{jk}| \ge \delta \sqrt[4]{np}] = o(1) \qquad (7.1.3)$$

and

$$\frac{1}{np}\max_{j\le p}\sum_{k=1}^{n}\mathrm{E}|x_{jk}^4|I[|x_{jk}|\le\delta\sqrt[4]{np}]=o(1). \tag{7.1.4}$$

When $p\to\infty$ with $n=n(p)$ and $p/n\to 0$, with probability 1 the ESD of $\mathbf{W}$ *tends to the semicircular law with the scale index σ^2.*

This theorem is not a corollary of Theorem 7.1 because neither of the conditions (X2,0) and (7.1.3) implies the other. But their proofs are very similar and thus we shall omit the details of the proof of Theorem 7.6.

Remark 7.7. If we assume that there is a positive and increasing function $\varphi(x)$ defined on $\mathbb{R}^+$ such that

$$\frac{1}{mn}\sum_{ij}\mathrm{E}|x_{ij}^2|\varphi(|x_{ij}|)I[|x_{ij}|>\eta\sqrt[4]{np}]\to 0 \tag{7.1.5}$$

and

$$\sum_{u=1}^{\infty}1/\varphi(\eta\sqrt[4]{np})<\infty, \tag{7.1.6}$$

then condition (X2,1) holds. If we take $\varphi(x)=x^{4(2\nu-1)}$ for some constant $\frac{1}{2}<\nu<1$, then (7.1.6) is automatically true and (7.1.5) reduces to a condition weaker than the assumption made in Kohrunzhy and Rodgers [177] if we change their notation to $p_{ij}=P(d_{ij}=1)=p/m$ with $p=n^{2\nu-1}$ and $m/n\to c$. Therefore, Theorem 7.1 covers Kohrunzhy and Rodgers [177] as a special case (condition (X3) is automatically true since $P(d_{ii}\neq 0)=0$).

Remark 7.8. The most important contribution of Theorem 7.1 to random matrix theory is to allow the nonhomogeneous and nonzero-one sparseness, and the order of m can be arbitrary between p and n. The conditions on the entries of $\mathbf{X}_n$ are to require some homogeneity on the $\mathbf{X}_n$ matrix. We conjecture that the homogeneity on the $\mathbf{X}_n$ matrix can be relaxed if we require the entries of the $\mathbf{D}_m$ matrix to have certain homogeneity. This problem is under investigation.

7.2 Truncation and Normalization

The strategy of the proof follows along the same lines as in Chapter 2, Section 2.2.

7.2.1 Truncation and Centralization

Truncation on entries of $\mathbf{D}_m$

Define

$$\hat{d}_{ij} = \begin{cases} d_{ij}, & \text{if } M_1 \leq |d_{ij}| \leq M_2, \\ 0, & \text{otherwise,} \end{cases}$$

$\widehat{\mathbf{D}}_m = (\hat{d}_{ij})$, and $\widehat{\mathbf{A}}_p = \frac{1}{\sigma^2\sqrt{np}}(\mathbf{X}_n\mathbf{X}_n^* - \sigma^2 n\mathbf{I}_m) \circ \widehat{\mathbf{D}}_m$.

Lemma 7.9. *Under the assumptions of Theorem 7.1,*

$$\|F^{\widehat{\mathbf{A}}_p} - F^{\mathbf{A}_p}\| \to 0 \text{ a.s. as } \ m \to \infty.$$

Proof. By the rank inequality (Theorem A.43),

$$\begin{aligned} \|F^{\widehat{\mathbf{A}}_p} - F^{\mathbf{A}_p}\| &\leq \frac{1}{m}\text{rank}(\mathbf{S}_m - \sigma^2\mathbf{I}_m) \circ (\mathbf{D}_m - \widehat{\mathbf{D}}_m) \\ &\leq \frac{1}{m}\sum_{ij} I(\{|d_{ij}| > M_2\} \cup \{0 < |d_{ij}| < M_1\}). \end{aligned}$$

By condition (D3) in (7.1.1),

$$\begin{aligned} &\frac{1}{m}\sum_{ij} I(\{|d_{ij}| > M_2\} \cup \{0 < |d_{ij}| < M_1\}) \\ \leq &\frac{1}{m}\sum_{ij} M_2^{-1}\text{E}|d_{ij}|^2 I(|d_{ij}| > M_2) + \text{P}(0 < |d_{ij}| < M_1) = o(1). \end{aligned}$$

Applying Bernstein's inequality, we obtain

$$\begin{aligned} &\text{P}(\|F^{\widehat{\mathbf{A}}_p} - F^{\mathbf{A}_p}\| \geq \varepsilon) \\ &\leq \text{P}\left(\sum_{ij} I[\{|d_{ij}| > M_2\} \cup \{0 < |d_{ij}| < M_1\}] \geq \varepsilon m\right) \\ &\leq 2e^{-bm}, \end{aligned}$$

for any $\varepsilon > 0$, all large n, and some constant $b > 0$. By the Borel-Cantelli lemma, we conclude that

$$\|F^{\widehat{\mathbf{A}}_p} - F^{\mathbf{A}_p}\| \to 0 \text{ a.s. as } m \to \infty.$$

Therefore, in the proof of Theorem 7.1, we may assume that the entries of $\mathbf{D}_m$ are either 0 or bounded by M_1 from 0 and by M_2 from above. We shall still use $\mathbf{D}_m = (d_{ij})$ in the subsequent proofs for brevity.

Removal of the diagonal elements of $\mathbf{A}_p$

For any $\varepsilon > 0$, denote by $\widehat{\mathbf{A}}_p$ the matrix obtained from $\mathbf{A}_p$ by replacing with 0 the diagonal elements whose absolute values are greater than ε and denote by $\widetilde{\mathbf{A}}_p$ the matrix obtained from $\mathbf{A}_p$ by replacing with 0 all diagonal elements.

Lemma 7.10. *Under the assumptions of Theorem 7.1,*

$$\|F^{\widehat{\mathbf{A}}_p} - F^{\mathbf{A}_p}\| \to 0 \text{ a.s. as } m \to \infty,$$

and

$$L(F^{\widehat{\mathbf{A}}_p}, F^{\widetilde{A}_p}) \leq \varepsilon.$$

Proof. The second conclusion of the lemma is a trivial consequence of Theorem A.45. As for the first conclusion, by the rank inequality (Theorem A.43),

$$\|F^{\widehat{\mathbf{A}}_p} - F^{\mathbf{A}_p}\| \leq \frac{1}{m}\sum_{i=1}^{m} I\left[\left|\frac{1}{\sqrt{np}}\sum_{k=1}^{n}(|x_{ik}|^2 - \sigma^2)d_{ii}\right| > \varepsilon\right].$$

By condition (X3) in (7.1.2), we have

$$\sum_{i=1}^{m} P\left[\left|\frac{1}{\sqrt{np}}\sum_{k=1}^{n}(|x_{ik}|^2 - \sigma^2)d_{ii}\right| > \varepsilon\right] = o(m).$$

Here, the reader should note that condition (X3) remains true after the truncation on the d's. By Bernstein's inequality, it follows that, for any constant $\eta > 0$,

$$\begin{aligned} &P(\|F^{\widehat{\mathbf{A}}_p} - F^{\mathbf{A}_p}\| \geq \eta) \\ &\leq P\left(\sum_{i=1}^{m} I\left[\left|\frac{1}{\sqrt{np}}\sum_{k=1}^{n}(|x_{ik}|^2 - \sigma^2)d_{ii}\right| > \varepsilon\right] \geq \eta m\right) \\ &\leq 2e^{-bm} \end{aligned}$$

for some constant $b > 0$. By the Borel-Cantelli lemma, we conclude that

$$\|F^{\widehat{\mathbf{A}}_p} - F^{\mathbf{A}_p}\| \to 0 \text{ a.s. as } m \to \infty.$$

Combining the two conclusions in Lemma 7.10, we have shown that

$$L(F^{\mathbf{A}_p}, F^{\widetilde{\mathbf{A}}_p}) \to 0 \text{ a.s. as } m \to \infty.$$

Hence, in what follows, we can assume that the diagonal elements are 0; i.e., assume $d_{ii} = 0$ for all $i = 1, \cdots, m$.

Truncation and centralization of the entries of $\mathbf{X}_n$

Note that condition (X2,0) in (7.1.2) guarantees the existence of $\eta_n \downarrow 0$ such that

$$\frac{1}{mn\eta_n^2}\sum_{ij}\mathrm{E}|x_{ij}|^2 I(|x_{ij}| \geq \eta_n \sqrt[4]{np}) \to 0.$$

Similarly, if condition (X2,1) holds, there exists $\eta_n \downarrow 0$ such that

$$\sum_u \frac{1}{mn\eta_n^2}\sum_{ij}\mathrm{E}|x_{ij}|^2 I(|x_{ij}| \geq \eta_n \sqrt[4]{np}) < \infty.$$

In the subsequent truncation procedure, we shall not distinguish under which condition the sequence $\{\eta_n\}$ is defined. The reader should remember that, whatever condition is used, the $\{\eta_n\}$ is defined by that condition.

Define $\tilde{x}_{ij} = x_{ij}I(|x_{ij}| \leq \eta_n \sqrt[4]{np}) - \mathrm{E}x_{ij}I(|x_{ij}| \leq \eta_n \sqrt[4]{np})$ and $\hat{x}_{ij} = x_{ij} - \tilde{x}_{ij}$. Also, define $\widetilde{B}_m$ with $\widetilde{B}_{ij} = \frac{1}{\sqrt{np}}\sum_{k=1}^n \tilde{x}_{ik}\bar{\tilde{x}}_{jk}$, and denote its Hadamard product with D_m by $\widetilde{\mathbf{A}}_p$. It is easy to verify that

$$\mathrm{E}|\hat{x}_{ij}|^2 \leq \mathrm{E}|x_{ij}|^2 I(|x_{ij}| \geq \eta_n \sqrt[4]{np}) \tag{7.2.1}$$

and

$$\mathrm{E}|\tilde{x}_{ij}|^2 \leq \sigma^2. \tag{7.2.2}$$

Then, we have the following lemma.

Lemma 7.11. *Under condition (X2,0) in (7.1.2) and other assumptions of Theorem 7.1,*

$$L(F^{\widetilde{\mathbf{A}}_p},\ F^{\mathbf{A}_p}) \to 0 \text{ in probability as } \ m \to \infty.$$

If condition (X2,0) is strengthened to (X2,1), then

$$L(F^{\widetilde{\mathbf{A}}_p},\ F^{\mathbf{A}_p}) \to 0 \text{ a.s. as } \ u \to \infty,$$

where $u = [p],\ m,$ or n in accordance with condition (X2,1).

Proof. By Theorem A.45,

$$\begin{aligned} L^3(F^{\widetilde{\mathbf{A}}_p},\ F^{\mathbf{A}_p}) &\leq \frac{1}{m}\mathrm{tr}[(B_m - \widetilde{B}_m)\circ D_m]^2 \\ &= \frac{1}{mnp}\sum_{i\neq j}\left|\sum_{k=1}^n (x_{ik}\bar{x}_{jk} - \tilde{x}_{ik}\bar{\tilde{x}}_{jk})d_{ij}\right|^2. \end{aligned}$$

By (7.2.1) and (7.2.2), we have

$$\mathrm{E}\left(\frac{1}{mnp}\sum_{i\neq j}\left|\sum_{k=1}^n (x_{ik}\bar{x}_{jk} - \tilde{x}_{ik}\bar{\tilde{x}}_{jk})d_{ij}\right|^2\right)$$

$$\leq \frac{1}{mnp} \sum_{i \neq j} \sum_{k=1}^{n} \mathrm{E}|x_{ik}\bar{x}_{jk} - \tilde{x}_{ik}\bar{\tilde{x}}_{jk}|^2 \mathrm{E}|d_{ij}^2|$$
$$\leq \frac{9\sigma^2}{mnp} \sum_{j=1}^{m} \sum_{k=1}^{n} \mathrm{E}|\hat{x}_{jk}|^2 \sum_{i=1}^{m} p_{ij}$$
$$\leq \frac{20\sigma^2}{mn} \sum_{j=1}^{m} \sum_{k=1}^{n} \mathrm{E}|x_{jk}|^2 I[|x_{jk}| > \eta_n \sqrt[4]{np}].$$

If condition (X2,0) in (7.1.2) holds, then the right-hand side of the inequality above converges to 0 and hence the first conclusion follows.

If condition (X2,1) holds, then the right-hand side of

$$\frac{20}{mn} \sum_{j=1}^{m} \sum_{k=1}^{n} \mathrm{E}|x_{jk}|^2 I[|x_{jk}| > \eta_n \sqrt[4]{np}]$$

is summable. Then, it follows that

$$L^3(F^{\widetilde{\mathbf{A}}_p},\ F^{\mathbf{A}_p}) \to 0 \text{ a.s.}$$

as $u \to \infty$, where u takes $[p]$, m, or n in accordance with the choice of u in (X2,1). The proof of this lemma is complete.

From Lemmas 7.9-7.11, to prove Theorem 7.1 we are allowed to make the following additional assumptions:

$$\begin{aligned} &\text{(i)}\ \ d_{ii} = 0, \quad M_1 I(d_{ij} \neq 0) \leq |d_{ij}| \leq M_2; \\ &\text{(ii)}\ \mathrm{E}x_{ij} = 0, \quad |x_{ij}| \leq \eta_n \sqrt[4]{np}. \end{aligned} \tag{7.2.3}$$

Note that we shall no longer have $\mathrm{E}|x_{ij}|^2 = \sigma^2$ after the truncation and centralization on the X variables. Write $\mathrm{E}|x_{ij}|^2 = \sigma_{ij}^2$. One can easily verify that:

$$\begin{aligned} &\text{(a) For any } i \neq j,\ \mathrm{E}|d_{ij}|^2 \leq p_{ij} \text{ and } \textstyle\sum_\ell \mathrm{E}|d_{i\ell}|^2 = p + o(p). \\ &\text{(b) For any } i \neq j,\ \sigma_{ij}^2 \leq \sigma^2 \text{ and } \tfrac{1}{mn}\textstyle\sum_{ik} \sigma_{ik}^2 \to \sigma^2. \end{aligned} \tag{7.2.4}$$

7.3 Proof of Theorem 7.1 by the Moment Approach

In the last section, we showed that to prove Theorem 7.1 it suffices to prove it under the additional conditions (i) and (ii) in (7.2.3) and (a) and (b) in (7.2.4).

To prove the theorem, we again employ the moment convergence approach. Let β_{nk} and β_k denote the k-th moment of $F^{\hat{\mathbf{A}}_p}$ and the semicircular law $F_{\sigma^2}(x)$ with the scale parameter σ^2.

It was shown in Chapter 2 that

$$\beta_k = \begin{cases} \frac{\sigma^{4s}(2s)!}{s!(s+1)!}, & \text{if } k = 2s, \\ 0, & \text{if } k = 2s+1, \end{cases}$$

and that $\{\beta_k\}$ satisfies the Carleman condition; i.e.,

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

Thus, to complete the proof of the theorem, we need only prove $\beta_{nk} \to \beta_k$ almost surely. By using the Borel-Cantelli lemma, we only need to prove

(I) $\mathrm{E}(\beta_{nk}) = \beta_k + o(1)$,

(II) $\mathrm{E}|\beta_{nk} - \mathrm{E}\beta_{nk}|^4 = O(\frac{1}{m^2})$.

Now, we begin to proceed to the proof of (I) and (II). Write $\mathbf{i} = (i_1, \cdots, i_k)$, $\mathbf{j} = (j_1, \cdots, j_k)$, and

$$\mathcal{I} = \{(\mathbf{i}, \mathbf{j}) : 1 \le i_v \le m,\ 1 \le j_v \le n,\ 1 \le v \le k\}.$$

Then, by definition, we have

$$\beta_{nk} = \frac{1}{mn^{k/2}p^{k/2}} \sum_{(\mathbf{i},\mathbf{j}) \in \mathcal{I}} d_{(\mathbf{i})} X_{(\mathbf{i},\mathbf{j})},$$

where

$$d_{(\mathbf{i})} = d_{i_1 i_2} \cdots d_{i_k i_1},$$

$$X_{(\mathbf{i},\mathbf{j})} = x_{i_1 j_1} \overline{x}_{i_2 j_1} x_{i_2 j_2} \overline{x}_{i_3 j_2} \cdots \overline{x}_{i_k j_{k-1}} x_{i_k j_k} \overline{x}_{i_1 j_k}.$$

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 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 $\overline{x}_{j_v i_{v+1}}$, and a horizontal edge (i_v, i_{v+1}) corresponds to the variable $d_{i_v, i_{v+1}}$. A graph corresponds to the product of the variables corresponding to the edges making up this graph. An example of such graphs is shown in Fig. 7.1. 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

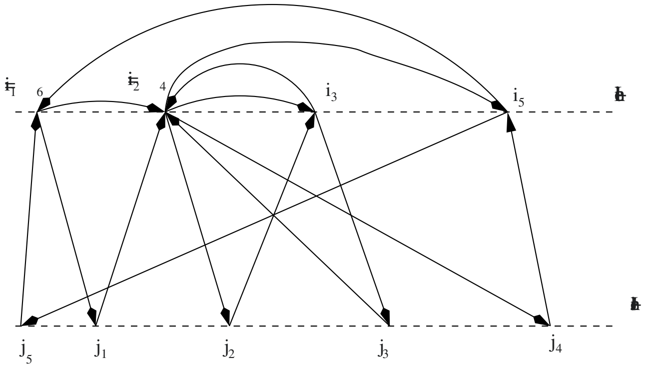

Fig. 7.1 A graph with six I- and six J-vertices

denote it as $\underline{G}(\mathbf{i},\mathbf{j})$. The roof of Fig. 7.1 is shown in Fig. 7.2. By noting that the roof of $G(\mathbf{i},\mathbf{j})$ depends on $\mathbf{i}$ only, we may simplify the notation of roofs as $\overline{G}(\mathbf{i})$.

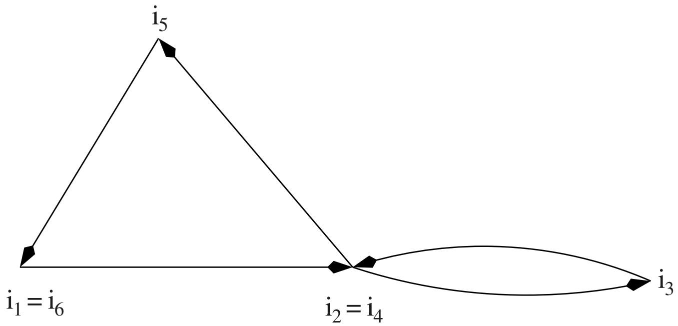

Fig. 7.2 The roof of Fig. 7.1

Two graphs $G(\mathbf{i}_1,\mathbf{j}_1)$ and $G(\mathbf{i}_2,\mathbf{j}_2)$ are said to be isomorphic if one can be converted to the other by a permutation on $(1,\cdots,m)$ and a permutation on $(1,\cdots,n)$. All graphs are classified into isomorphic classes. An isomorphic class is denoted by $\mathcal{G}$. Similarly, two roofs $\overline{G}(\mathbf{i}_1)$ and $\overline{G}(\mathbf{i}_2)$ are said to be isomorphic if one can be converted to the other by a permutation on $(1,\cdots,m)$. An isomorphic roof class is denoted by $\overline{\mathcal{G}}$. For a given $\mathbf{i}$, two graphs $G(\mathbf{i},\mathbf{j}_1)$ and $G(\mathbf{i},\mathbf{j}_2)$ are said to be isomorphic given $\mathbf{i}$ if one can be converted to the other by a permutation on $(1,\cdots,n)$. An isomorphic class given $\mathbf{i}$ is denoted by $\underline{\mathcal{G}}(\mathbf{i})$.

Let r, s, and l denote the number of noncoincident i-vertices, noncoincident j-vertices, and noncoincident vertical edges. Let $\mathcal{G}(r,s,l)$ denote the collection of all isomorphic classes with the numbers r, s, l.

Then, we may rewrite

$$\begin{aligned}\beta_{nk} &= \frac{1}{mn^{k/2}p^{k/2}}\sum_{\mathbf{i},\mathbf{j}} d_{\overline{G}(\mathbf{i})}X_{\underline{G}(\mathbf{i},\mathbf{j})} \\ &= \frac{1}{mn^{k/2}p^{k/2}}\sum_{r,s,l}\sum_{\mathcal{G}\in\mathcal{G}(r,s,l)}\sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} d_{\overline{G}(\mathbf{i})}X_{\underline{G}(\mathbf{i},\mathbf{j})}. \qquad (7.3.1)\end{aligned}$$

Proof of (I). By the notation introduced above,

$$\mathrm{E}(\beta_{nk}) = \frac{1}{mn^{k/2}p^{k/2}}\sum_{r,s,l}\sum_{\mathcal{G}\in\mathcal{G}(r,s,l)}\sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} \mathrm{E}d_{\overline{G}(\mathbf{i})}\mathrm{E}X_{\underline{G}(\mathbf{i},\mathbf{j})}.$$

When $G(\mathbf{i},\mathbf{j})$ contains a single vertical edge, $\mathrm{E}X_{\underline{G}(\mathbf{i},\mathbf{j})} = 0$. When $G(\mathbf{i},\mathbf{j})$ contains a loop (that is, for some $v \le k$, $i_v = i_{v+1}$ (i_{k+1} is understood as i_1)), $d_{\overline{G}(\mathbf{i})} = 0$ since $d_{ii} = 0$ for all $i \le m$.

So, we need only consider the graphs that have no single vertical edges and no loops of horizontal edges. Now, we write

$$\mathrm{E}(\beta_{nk}) = S_1 + S_2 + S_3,$$

where S_1 contains all terms subject to $l < k$ or $r + s \le k$, S_2 contains all terms with $l = k = r + s - 1$ but $s < \frac{1}{2}k$, and S_3 contains all terms with $l = k = r + s - 1$ and $s = \frac{1}{2}k$.

Before evaluating the sums above, we first prove the following lemma.

Lemma 7.12. *For a given r and a given i-index, say i_1, there is a constant K such that, for all $\overline{\mathcal{G}} \in \overline{\mathcal{G}}(r)$,*

$$\left|\sum_{\substack{\overline{G}(\mathbf{i})\in\mathcal{G}\\ \text{fixed } i_1}} \mathrm{E}d_{\overline{G}(\mathbf{i})}\right| \le Kp^{r-1}. \qquad (7.3.2)$$

Consequently, we have

$$\left|\sum_{\overline{G}(\mathbf{i})\in\mathcal{G}} \mathrm{E}d_{\overline{G}(\mathbf{i})}\right| \le Kmp^{r-1}.$$

Proof. If $r = 1$, $\overline{\mathcal{G}}(r) = \emptyset$ since $G(\mathbf{i},\mathbf{j})$ has no loops, and hence (7.3.2) follows trivially. Thus, we only consider the case where $r \ge 2$.

First, let us consider $\mathrm{E}d_{\overline{G}(\mathbf{i})}$. If $\overline{G}(\mathbf{i})$ contains μ_1 horizontal edges with vertices (u,v) and μ_2 horizontal edges with vertices (v,u), then $\mathrm{E}d_{\overline{G}(\mathbf{i})}$ contains a factor $\mathrm{E}d_{u,v}^{\mu_1}\bar{d}_{u,v}^{\mu_2}$ whose absolute value is not larger than $M_2^{\mu_1+\mu_2-2}p_{uv}$ if $\mu_1+\mu_2 \geq 2$ and not larger than $M_1^{-1}p_{uv}$ if $\mu_1+\mu_2 = 1$. Also, we have $|\mathrm{E}d_{u,v}^{\mu_1}\bar{d}_{u,v}^{\mu_2}| \leq M_2^{\mu_1+\mu_2}$ for all cases. That is, each noncoincident horizontal edge of $\overline{G}(\mathbf{i})$ in $\overline{\mathcal{G}}$ corresponds to a factor that is dominated by a constant C and Cp_{uv} for some constant C.

Note that $\overline{G}(\mathbf{i})$ is connected. Thus, for each isomorphic roof class with index r, we may select a tree $T(\mathbf{i})$ from the noncoincident edges of $\overline{G}(\mathbf{i})$ such that any two trees $T(\mathbf{i}_1)$ and $T(\mathbf{i}_2)$ are isomorphic for any two roofs $\overline{G}(\mathbf{i}_1)$ and $\overline{G}(\mathbf{i}_2)$ in the same class. Denote the $r-1$ edges of $T(\mathbf{i})$ by $(u_1,v_1),\cdots,(u_{r-1},v_{r-1})$. Then,

$$|\mathrm{E}d_{\overline{G}(\mathbf{i})}| \leq Cp_{u_1,v_1}\cdots p_{u_{r-1},v_{r-1}}.$$

The inequality above follows by bounding the factors corresponding to edges in the tree by $Cp_{u,v}$ and other factors by C. If $r=2$, then the lemma follows from condition (a) in (7.2.4).

If $r>2$, we use induction. Assume (7.3.2) is true for $r-1$. Since $T(\mathbf{i})$ is a tree, without loss of generality we assume that v_{r-1} is a root other than i_1 of the tree; that is, $v_{r-1} \notin \{u_1,v_1,\cdots,u_{r-2},v_{r-2}\}$, and $i_1, u_{r-1} \in \{u_1,v_1,\cdots,u_{r-2},v_{r-2}\}$. Then, using assumption (D2),

$$\begin{aligned}
&\sum_{u_1,v_1,\cdots,u_{r-1},v_{r-1}} p_{u_1,v_1}\cdots p_{u_{r-1},v_{r-1}}\\
&= \sum_{u_1,v_1,\cdots,u_{r-2},v_{r-2}} p_{u_1,v_1}\cdots p_{u_{r-2},v_{r-2}} \sum_{v_{r-1}} p_{u_{r-1},v_{r-1}}\\
&\leq (p+o(p)) \sum_{u_1,v_1,\cdots,u_{r-2},v_{r-2}} p_{u_1,v_1}\cdots p_{u_{r-2},v_{r-2}}\\
&\leq (p+o(p))^{r-1},
\end{aligned}$$

the last inequality following from the inductive hypothesis. The lemma follows.

Continuing the proof of (I). When $\mathcal{G}$ belongs to $\mathcal{G}(r,s,l)$ with $l<k$ or $r+s \leq k$, for any given $\mathbf{i}$, we have

$$\sum_{G(\mathbf{i},\mathbf{j})\in\underline{\mathcal{G}}(\mathbf{i})} |\mathrm{E}X_{\underline{G}(\mathbf{i},\mathbf{j})}| \leq n^s\sigma^{2l}(\eta_n\sqrt[4]{np})^{2k-2l}. \tag{7.3.3}$$

Let $\overline{\mathcal{G}}(r)$ denote the set of all isomorphic roof classes with r noncoincident i-vertices.

By (7.3.3) and Lemma 7.12, we have

$$|S_1| \leq \frac{1}{m(np)^{\frac{1}{2}k}} \sum_{\substack{r,s,\ l<k\\ \text{or } r+s\leq k}} \sum_{\mathcal{G}\in\mathcal{G}(r,s,l)} O(mp^{r-1})(n^s\sigma^{2l}(\eta_n\sqrt[4]{np})^{2k-2l})$$

$$= \begin{cases} o(n^{-\frac{1}{2}l+s}p^{-\frac{1}{2}l+r-1}), & \text{if } l < k, \\ o(n^{-\frac{1}{2}k+s}p^{-\frac{1}{2}k+r}), & \text{if } r+s \le l = k. \end{cases} \tag{7.3.4}$$

Noting that each j_v vertex connects with at least two noncoincident vertical edges since $G(\mathbf{i},\mathbf{j})$ does not have loops, we conclude that $l \ge 2s$. Since the vertical edges form a connected graph, we conclude that $r+s \le l+1$ for graphs in S_1. Therefore, we have

$$|S_1| \le \left\{ \begin{array}{ll} o((p/n)^{\frac{1}{2}l-s}), & \text{if } l < k \\ o((p/n)^{\frac{1}{2}k-s}), & \text{if } r+s \le l = k \end{array} \right\} = o(1).$$

By the same argument as for estimating S_1, we also have

$$|S_2| \le O((p/n)^{\frac{1}{2}k-s}) = o(1),$$

where the second inequality follows from the fact that $s < \frac{1}{2}k$.

Note that if k is odd and $G(\mathbf{i},\mathbf{j})$ does not have single vertical edges, it is impossible to have $l = k$ and $s = \frac{1}{2}k$ since $2s \le l \le k$. That is, all terms of $\mathrm{E}\beta_{nk}$ are either in S_1 or S_2. We have now proved that

$$\mathrm{E}\beta_{nk} \to 0.$$

Therefore, in what follows, we only need to consider S_3 in the case where k is even.

Now let us evaluate S_3 for the case $l = k = 2s = 2r-2$. In this case, each noncoincident j-vertex connects exactly with two noncoincident edges and the noncoincident vertical edges from a tree. As discussed in Chapter 3, each down edge must be coincident with an up edge.

Denote the noncoincident vertical edges by $\{(u_1,v_1),\cdots,(u_k,v_k)\}$. Then

$$\mathrm{E}X_{\underline{G}(\mathbf{i},\mathbf{j})} = \prod_{j=1}^{k} \sigma_{u_j,v_j}^2,$$

and hence, for each isomorphic class $\mathcal{G}$, we have

$$\sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} \mathrm{E}d_{\overline{G}(\mathbf{i})} \prod_{j=1}^{k} \sigma_{u_j,v_j}^2.$$

We first show that

$$\frac{1}{m(np)^s} \sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} \mathrm{E}d_{\overline{G}(\mathbf{i})} \prod_{j=1}^{k} \sigma_{u_j,v_j}^2 = \frac{1}{m(np)^s} \sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} \mathrm{E}d_{\overline{G}(\mathbf{i})}\sigma^{2k} + o(1). \tag{7.3.5}$$

By condition (b) in (7.2.4) and Lemma 7.12, we have

$$
\begin{aligned}
0 &\le \frac{1}{m(np)^s} \sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} |\mathrm{E}d_{\overline{G}(\mathbf{i})}| \left[\sigma^{2k} - \prod_{j=1}^{k} \sigma^2_{u_j,v_j}\right] \\
&\le \frac{1}{m(np)^s} \sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} |\mathrm{E}d_{\overline{G}(\mathbf{i})}| \sum_{\ell=1}^{k} \left[\sigma^{2(s-\ell)}(\sigma^2 - \sigma^2_{u_\ell,v_\ell}) \prod_{j=1}^{\ell-1} \sigma^2_{u_j,v_j}\right] \\
&\le \frac{1}{m(np)^s} \sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} |\mathrm{E}d_{\overline{G}(\mathbf{i})}| \sum_{\ell=1}^{k} \left[\sigma^{2(s-1)}(\sigma^2 - \sigma^2_{u_\ell,v_\ell})\right] \\
&\le \frac{\sigma^{2(k-1)}}{mnp^s} \sum_{\overline{G}(\mathbf{i})\in\overline{\mathcal{G}}} |\mathrm{E}d_{\overline{G}(\mathbf{i})}| \sum_{\ell=1}^{k} \sum_{u_\ell} (\sigma^2 - \sigma^2_{u_\ell,v_\ell}) \\
&\le \sum_{\ell=1}^{k} \frac{K\sigma^{2(k-1)}}{mn} \sum_{v_\ell} \sum_{u_\ell} (\sigma^2 - \sigma^2_{u_\ell,v_\ell}) \to 0,
\end{aligned}
$$

from which (7.3.5) 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, if $(i_\ell, i_{\ell+1})$ 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 $\{i_1, \cdots, i_\ell\}$), there is a vertex j_ℓ such that (i_ℓ, j_ℓ) 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 a down edge (i_ν, j_ν) coincident with the up edge $(j_\ell, i_{\ell+1})$; that is, $i_\nu = i_{\ell+1} = v_2$. 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 connect to at least three noncoincident vertical edges, which violates the assumption that $s = k/2$. That is, we have proved that $i_{\nu+1} = i_\ell = v_1$ and hence 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 Fig. 7.3). Since the number of noncoincident i vertices is $r = s + 1 = \frac{1}{2}k + 1$, the noncoincident edges of $\overline{G}(\mathbf{i})$ form a tree of $s = \frac{1}{2}k$ edges.

Therefore,

$$
\mathrm{E}d_{\overline{G}(\mathbf{i})} = \prod_{\ell=1}^{s} p_{e_\ell},
$$

where $e_\ell, \ell = 1, \cdots, s$, are the edges of the tree of noncoincident horizontal edges.

By (7.12) and (7.3.5), we have

$$
\begin{aligned}
&\frac{1}{m(np)^s} \sum_{G(\mathbf{i},\mathbf{j})\in\mathcal{G}} \mathrm{E}d_{\overline{G}(\mathbf{i})} \mathrm{E}X_{G(\mathbf{i},\mathbf{j})} \\
&= \frac{\sigma^{2k}}{mp^s} \sum_{\mathcal{G}(\mathbf{i})\in\overline{\mathcal{G}}} \mathrm{E}d_{\overline{G}(\mathbf{i})} + o(1) \\
&= \sigma^{2k} + o(1).
\end{aligned}
$$

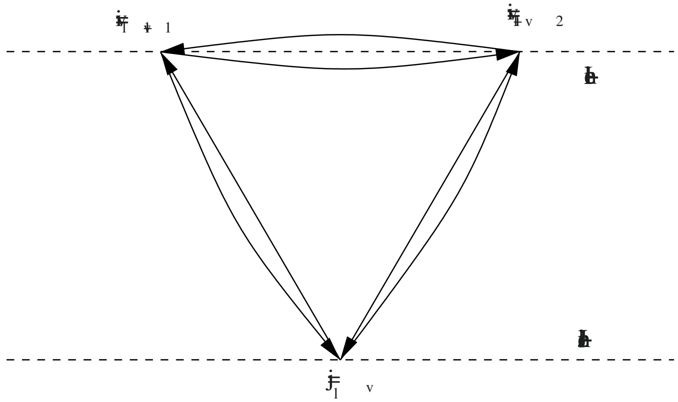

Fig. 7.3 (v_1, v_2) coincides with (v_2, v_1).

Recall that the number of isomorphic classes with indices $k = 2s$ has been computed in Chapter 2; that is,

$$\frac{(2s)!}{s!(s+1)!}.$$

Conclusion (I) follows.

Proof of (II). Write

$$\begin{aligned} & \mathrm{E}|\beta_{nk} - \mathrm{E}\beta_{nk}|^4 \qquad (7.3.6) \\ &= \frac{1}{m^4 n^{2k} p^{2k}} \sum_{\substack{\mathbf{i}_\ell, \mathbf{j}_\ell \\ \ell=1,2,3,4}} \mathrm{E} \prod_{\ell=1}^{4} [d_{\overline{G}(\mathbf{i}_\ell)} X_{G(\mathbf{i}_\ell, \mathbf{j}_\ell)} - \mathrm{E} d_{\overline{G}(\mathbf{i}_\ell)} \mathrm{E} X_{G(\mathbf{i}_\ell, \mathbf{j}_\ell)}], \end{aligned}$$

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 (I).

If $G(\mathbf{i}_\ell, \mathbf{j}_\ell)$ has no edges coincident with edges of the other three, then the corresponding term is 0 by independence. Furthermore, the term is also 0 if $\bigcup_{\ell=1}^{4} G(\mathbf{i}_\ell, \mathbf{j}_\ell)$ contains a single vertical edge or a loop. We need to consider the following two cases:

(1) The four graphs are connected together through a coincident edge.

(2) $\bigcup_{\ell=1}^{4} G(\mathbf{i}_\ell, \mathbf{j}_\ell)$ has two separated pieces; that is, two graphs are connected and the other two are connected.

Split $\mathrm{E}|\beta_{nk} - \mathrm{E}\beta_{nk}|^4 = S_I + S_{II}$ according to the two cases.

For these two cases, suppose $\bigcup_{\ell=1}^{4} G(\mathbf{i}_\ell, \mathbf{j}_\ell)$ contains r noncoincident i-vertices, s noncoincident j-vertices, and l noncoincident vertical edges. Similar to the proof of (I), we have

$$S_I \le \sum_{r,s,l} \frac{1}{m^4 n^{2k} p^{2k}} Kmp^{r-1}(\eta_n \sqrt[4]{np})^{8k-2l} n^s$$
$$= K \sum_{r,s,l} m^{-3} p^{-1} n^{-\frac{1}{2}l+s} p^{-\frac{1}{2}l+r}.$$

For graphs corresponding to terms in S_I, we have $2s \le l$ and $r + s \le l + 1$. We obtain

$$|S_I| \le Km^{-3}(p/n)^{\frac{1}{2}l-s} = O(m^{-3}).$$

Similarly, we have

$$S_{II} \le \sum_{r,s,l} \frac{1}{m^4 n^{2k} p^{2k}} Km^2 p^{r-2}(\eta_n \sqrt[4]{np})^{8k-2l} n^s$$
$$= K \sum_{r,s,l} m^{-2} p^{-1} n^{-\frac{1}{2}l+s} p^{-\frac{1}{2}l+r}.$$

For graphs corresponding to terms in S_{II}, we have $2s \le l$ and $r + s \le l + 2$. We obtain

$$|S_{II}| \le Km^{-2}(p/n)^{\frac{1}{2}-s} = O(m^{-2}).$$

Combining the above, (II) is proved.

Remark 7.13. Using the same approach in proving (II), one can easily show that

$$\mathrm{E}|\beta_{nk} - \mathrm{E}\beta_{nk}|^{2\mu} = O(m^{-\mu})$$

for any fixed integer μ. This is useful when considering the a.s. convergence when $n \to \infty$.

These estimates may be used to strengthen the almost sure convergence for $m \to \infty$ to $n \to \infty$ when m is at least as large as a power of n.

Chapter 8
Convergence Rates of ESD

In applications of asymptotic theorems of spectral analysis of large dimensional random matrices, one of the important problems is the convergence rate of the ESD. It had been puzzling probabilists for a long time until the papers of Bai [16, 17] were published. The moment approach to establishing limiting theorems for spectral analysis of large dimensional random matrices is to show that each moment of the ESD tends to a nonrandom limit. This proves the existence of the LSD by applying the Carleman criterion. This method successfully established the existence of the LSD of large dimensional Wigner matrices, sample covariance matrices, and multivariate F-matrices. However, this method cannot give any convergence rate. In Bai [16], three inequalities were established in terms of the difference of Stieltjes transforms (see Chapter B). In this chapter, we shall apply these inequalities to establish the convergence rates for the ESD of large Wigner and sample covariance matrices.

8.1 Convergence Rates of the Expected ESD of Wigner Matrices

In this section, we consider the convergence rate of ESD of the complex Wigner matrix $\mathbf{W}_n = (x_{ij},\ i, j = 1, \cdots, n)$, whose entries may depend on n, but the index n is suppressed for brevity. Also, we assume the following conditions hold:

$$\begin{aligned} &\text{(i) } \mathrm{E}x_{ij} = 0, \quad x_{ji} = \bar{x}_{ij}, \quad \text{for all } 1 \le i \le j \le n. \\ &\text{(ii) } \mathrm{E}|x_{ij}^2| = 1, \quad \text{for all } 1 \le i < j \le n; \\ &\qquad \mathrm{E}|x_{ii}^2| = \sigma^2, \quad \text{for all } 1 \le i \le n. \\ &\text{(iii) } \sup_n \sup_{1 \le i < j \le n} \mathrm{E}|x_{ij}^6|, \mathrm{E}|x_{ii}^3| \le M < \infty. \end{aligned} \tag{8.1.1}$$

For convenience, we assume that $\sigma^2 \le M$ in what follows.

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_8,

Remark 8.1. The moment requirements (iii) are needed for maintaining the convergence rate after the truncation step.

Here and in what follows, F_n denotes the ESD of $\frac{1}{\sqrt{n}}\mathbf{W}_n$. Under the conditions in (8.1.1), it follows from Theorem 2.9 that $F_n \xrightarrow{w} F$ almost surely, where F is the well-known semicircular law,

$$F(x) = \frac{1}{2\pi}\int_{-\infty}^{x} \sqrt{4-y^2}\, I_{[-2,2]}(y)\mathrm{d}y. \tag{8.1.2}$$

In this section, we shall establish the following theorem.

Theorem 8.2. *Under assumptions (8.1.1), we have*

$$\|\mathrm{E}F_n - F\| = O(n^{-1/2}). \tag{8.1.3}$$

The proof of the theorem, which relies strongly on those inequalities on Stieltjes transforms found in Appendix B, is similar to what was done in Chapter 2 but requires more accurate estimates on the remainder term of the Stieltjes transform of the ESD. To this end, the error quantity δ (see (8.1.20)) should be expanded into additional terms. Also, the truncation position on the x-variables needs to be lowered to $n^{1/4}$ while allowing the renormalization to maintain the same convergence rate.

8.1.1 Lemmas on Truncation, Centralization, and Rescaling

Lemma 8.3. *Let $\widetilde{\mathbf{W}}_n$ denote the matrix whose entries are $\frac{1}{\sqrt{n}}x_{ij}I(|x_{ij}| \le n^{1/4})$ for all i, j. Then, we have*

$$\sqrt{n}\mathrm{E}\|F_n - \widetilde{F}_n\| \le M, \tag{8.1.4}$$

where F_n and $\widetilde{F}_n$ are the ESDs of $\mathbf{W}_n$ and $\widetilde{\mathbf{W}}_n$, respectively.

Also, for any $\eta > 0$,

$$\limsup_{n\to\infty} n^{\frac{1}{2}-\eta}\|F_n - \widetilde{F}_n\| = 0, \quad a.s. \tag{8.1.5}$$

If $\{|x_{ii}|^2\}$ and $\{|x_{ij}|^6,\ i < j \le n\}$ are uniformly integrable, then the right-hand side of (8.1.4) can be improved to zero.

Proof. By Theorem A.43, we have

$$\|F_n - \widetilde{F}_n\| \le \frac{1}{n}\mathrm{rank}(\mathbf{W}_n - \widetilde{\mathbf{W}}_n)$$

$$\leq \frac{1}{n}\sum_{i\neq j} I(|x_{ij}| > n^{1/4}) + \frac{1}{n}\sum_{i=1}^{n} I(|x_{ii}| > n^{1/4}).$$

Therefore,

$$\begin{aligned}
\mathrm{E}\|F_n - \widetilde{F}_n\| &\leq \frac{1}{n}\sum_{i\neq j} \mathrm{P}(|x_{ij}| > n^{1/4}) + \frac{1}{n}\sum_{i=1}^{n} \mathrm{P}(|x_{ii}| > n^{1/4}) \\
&\leq n^{-5/2}\sum_{i\neq j} \mathrm{E}|x_{ij}|^6 + n^{-3/2}\sum_{i=1}^{n} \mathrm{E}|x_{ii}|^2 \\
&\leq Mn^{-1/2}.
\end{aligned}$$

From the estimate above and Bernstein's inequality, the second conclusion follows. The proof of Lemma 8.3 is complete.

Lemma 8.4. *Let* $\widehat{\mathbf{W}}_n$ *denote the matrix whose entries are* $\frac{1}{\sqrt{n}}[x_{ij}I(|x_{ij}| \leq n^{1/4}) - \mathrm{E}x_{ij}I(|x_{ij}| \leq n^{1/4})]$ *for all* i, j. *Then, we have*

$$L(\widehat{F}_n, \widetilde{F}_n) \leq M^{2/3}n^{-1/2}, \tag{8.1.6}$$

where $L(\cdot,\cdot)$ *denotes the Levy distance between distribution functions and* $\widehat{F}_n$ *is the ESD of* $\widehat{\mathbf{W}}_n$.

Proof. By Corollary A.41, we have

$$\begin{aligned}
&L^3(\widehat{F}_n, \widetilde{F}_n) \leq \frac{1}{n}\mathrm{tr}(\widehat{\mathbf{W}}_n - \widetilde{\mathbf{W}}_n)^2 \\
&= \frac{1}{n^2}\sum_{i\neq j} |\mathrm{E}x_{ij}I(|x_{ij}| \leq n^{1/4})|^2 + \frac{1}{n^2}\sum_{i=1}^{n} |\mathrm{E}x_{ii}I(|x_{ii}| \leq n^{1/4})|^2 \\
&\leq \frac{1}{n^{7/2}}\sum_{i\neq j} \mathrm{E}^2|x_{ij}|^6 + \frac{1}{n^{5/2}}\sum_{i=1}^{n} \mathrm{E}^2|x_{ii}|^2 \\
&\leq M^2n^{-3/2}.
\end{aligned}$$

The proof is done.

Lemma 8.5. *Let* $\widetilde{\widetilde{\mathbf{W}}}$ *denote the matrix whose entries are* $\frac{1}{\sqrt{n}}\sigma_{ij}^{-1}[x_{ij}I(|x_{ij}| \leq n^{1/4}) - \mathrm{E}x_{ij}I(|x_{ij}| \leq n^{1/4})]$ *for* $i \neq j$ *and* $\frac{1}{\sqrt{n}}\sigma\sigma_{ii}^{-1}[x_{ii}I(|x_{ii} \leq n^{1/4}) - \mathrm{E}x_{ii}I(|x_{ii} \leq n^{1/4})]$, *where* σ_{ij}^2 *are the variances of the truncated variables. Then, we have*

$$\mathrm{E}L(\widehat{F}_n, \widetilde{\widetilde{F}}_n) \leq 2^{1/3}M^{2/3}n^{-1/2}, \tag{8.1.7}$$

$$\limsup_{n\to\infty} \sqrt{n}L(\widehat{F}_n, \widetilde{\widetilde{F}}_n) \leq 2^{1/3}M^{2/3} \quad \text{a.s.}, \tag{8.1.8}$$

where $\widetilde{\widehat{F}}_n$ *is the ESD of* $\widetilde{\widehat{\mathbf{W}}}_n$.

Proof. By Corollary A.41, we have

$$L^3(\widehat{F}_n, \widetilde{\widehat{F}}_n) \le \frac{1}{n}\mathrm{tr}(\widehat{\mathbf{W}}_n - \widetilde{\widehat{\mathbf{W}}}_n)^2$$
$$= \frac{1}{n^2}\sum_{i\neq j}(1-\sigma_{ij}^{-1})^2|x_{ij}I(|x_{ij}|\le n^{1/4}) - \mathrm{E}x_{ij}I(|x_{ij}|\le n^{1/4})|^2$$
$$+\frac{1}{n^2}\sum_{i=1}^{n}(1-\sigma\sigma_{ii}^{-1})^2|x_{ii}I(|x_{ii}|\le n^{1/4}) - \mathrm{E}x_{ii}I(|x_{ii}|\le n^{1/4})|^2.$$

Thus,

$$\begin{aligned}\mathrm{E}L^3(\widehat{F}_n, \widetilde{\widehat{F}}_n) &\le \frac{1}{n^2}\sum_{i\neq j}(1-\sigma_{ij})^2 + \frac{1}{n^2}\sum_{i=1}^{n}(\sigma-\sigma_{ii})^2\\ &\le \frac{1}{n^2}\sum_{i\neq j}(1-\sigma_{ij}^2)^2 + \frac{1}{n^2}\sum_{i=1}^{n}(\sigma^2-\sigma_{ii}^2)^2\\ &\le 2M^2n^{-3/2},\end{aligned}$$

where we have used the fact that, for all large n, $\sigma_{ij}(1+\sigma_{ij}) \ge 1$ and hence

$$\begin{aligned}(1-\sigma_{ij}^2)^2 &\le (\mathrm{E}|x_{ij}^2|I(|x_{ij}|>n^{1/4}) + \mathrm{E}^2|x_{ij}|I(|x_{ij}|>n^{1/4}))^2\\ &\le 2M^2n^{-2}\end{aligned}$$

and

$$\begin{aligned}(\sigma^2-\sigma_{ii}^2)^2 &\le (\mathrm{E}|x_{ii}^2|I(|x_{ii}|>n^{1/4}) + \mathrm{E}^2|x_{ii}|I(|x_{ii}|>n^{1/4}))^2\\ &\le 2M^2n^{-1}.\end{aligned}$$

The proof of (8.1.7) is done. Conclusion (8.1.8) follows from the fact that

$$\mathrm{Var}\Bigg(\frac{1}{\sqrt{n}}\sum_{i\neq j}(1-\sigma_{ij}^{-1})^2|x_{ij}I(|x_{ij}|\le n^{1/4}) - \mathrm{E}x_{ij}I(|x_{ij}|\le n^{1/4})|^2$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\sigma\sigma_{ij}^{-1})^2|x_{ii}I(|x_{ii}|\le n^{1/4}) - \mathrm{E}x_{ii}I(|x_{ii}|\le n^{1/4})|^2\Bigg)$$
$$\le \frac{16M^2}{n}\left(\sum_{i\neq j}(1-\sigma_{ij}^{-1})^4 + \sum_{i=1}^{n}(1-\sigma\sigma_{ii}^{-1})^4\right)$$
$$\le 64M^4n^{-2}.$$

8.1.2 Proof of Theorem 8.2

By Lemma B.18 with $D = 1/\pi$ and $\alpha = 1$, we know that $L(F_n, F)$ and $\|F_n - F\|$ have the same order if F is the distribution function of the semicircular law. Now, applying Lemmas 8.3, 8.4, and 8.5, to prove Theorem 8.2 for the general case, it suffices to prove it for the truncated, centralized, and rescaled version. Therefore, we shall assume that the entries of the Wigner matrix are truncated at the positions given in Lemma 8.3 and then centralized and rescaled.

Define

$$\Delta = \|\mathrm{E}F_n - F\|, \tag{8.1.9}$$

where F_n is the ESD of $\frac{1}{\sqrt{n}}\mathbf{W}_n$ and F is the distribution function of the semicircular law.

Recall that we found in Chapter 2 the Stieltjes transform of the semicircular law, which is given by

$$s(z) = -\frac{1}{2}(z - \sqrt{z^2 - 4}). \tag{8.1.10}$$

Here, the reader is reminded that the square root of a complex number is defined to be the one with a positive imaginary part.

Then, it is easy to verify that $s(z)(-\frac{1}{2}(z + \sqrt{z^2 - 4})) = 1$ and $|s(z)| < |(-\frac{1}{2}(z+\sqrt{z^2 - 4}))|$ since both the real and imaginary parts of z and $\sqrt{z^2 - 4}$ have the same signs. Hence, for any $z \in \mathbb{C}^+$,

$$|s(z)| < 1. \tag{8.1.11}$$

Now, we begin to prove the theorem by using the inequality of Theorem B.14.

Let u and $v > 0$ be real numbers and let $z = u + iv$. Set

$$s_n(z) = \int_{-\infty}^{\infty} \frac{1}{x - z} F_n(x) = \frac{1}{n}\mathrm{tr}(\mathbf{W}_n - z\mathbf{I}_n)^{-1}. \tag{8.1.12}$$

By (8.1.12) and the inverse matrix formula (see (A.1.8)),

$$\begin{aligned}
\mathrm{E}s_n(z) &= \frac{1}{n}\sum_{k=1}^{n} \mathrm{E}\frac{1}{\frac{1}{\sqrt{n}}x_{kk} - z - \frac{1}{n}\alpha_k'(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\alpha_k} \\
&= \frac{1}{n}\sum_{k=1}^{n} \mathrm{E}\frac{1}{\varepsilon_k - z - \mathrm{E}s_n(z)} \\
&= -\frac{1}{z + \mathrm{E}s_n(z)} + \delta,
\end{aligned} \tag{8.1.13}$$

where $\alpha_k' = (x_{1k}, \cdots, x_{k-1,k}, x_{k+1,k}, \cdots, x_{nk})$, $\mathbf{W}_n(k)$ is the matrix obtained from $\mathbf{W}_n$ by deleting the k-th row and k-th column,

$$\varepsilon_k = \frac{1}{\sqrt{n}} x_{kk} - \frac{1}{n}\alpha_k'(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\alpha_k + \mathrm{E}s_n(z), \tag{8.1.14}$$

and

$$\delta = \delta_n = -\frac{1}{n}\sum_{k=1}^{n} \mathrm{E}\frac{\varepsilon_k}{(z + \mathrm{E}s_n(z))(z + \mathrm{E}s_n(z) - \varepsilon_k)}. \tag{8.1.15}$$

Solving the quadratic equation (8.1.13), we obtain

$$s_{(1)}(z), s_{(2)}(z) = -\frac{1}{2}(z - \delta \pm \sqrt{(z+\delta)^2 - 4}). \tag{8.1.16}$$

As analyzed in Chapter 2, we should have

$$\mathrm{E}s_n(z) = s_{(2)}(z) = -\frac{1}{2}(z - \delta - \sqrt{(z+\delta)^2 - 4}). \tag{8.1.17}$$

If $\Im(z + \delta) > 0$, we can also write

$$\mathrm{E}s_n(z) = \delta + s(z + \delta). \tag{8.1.18}$$

We shall show that (8.1.18) is true for all $z \in \mathbb{C}^+$. We shall prove this by showing that $\mathbb{D} = \mathbb{C}^+$, where

$$\mathbb{D} = \{z \in \mathbb{C}^+, \Im(z + \delta(z)) > 0\}.$$

At first, we see that $\delta \to 0$ as $\Im z \to \infty$. That is, $z \in \mathbb{D}$ if $\Im(z)$ is large.

If $\mathbb{D} \neq \mathbb{C}^+$, then there is a point in $\mathbb{C}^+\mathbb{D}^c$, say z_1. Let $z_0 \in \mathbb{D}$. Let z_2 be a point in the intersection of $\partial\mathbb{D}$ and the segment connecting z_0 and z_1. By the continuity of $\delta(z)$ in z, we have $\Im(z_2 + \delta(z_2)) = 0$. By (8.1.13), we obtain

$$z_2 + \mathrm{E}s_n(z_2) + \frac{1}{z_2 + \mathrm{E}s_n(z_2)} = z_2 + \delta(z_2),$$

in which the right-hand side is a real number. We conclude that

$$|z_2 + \mathrm{E}s_n(z_2)| = 1.$$

Since $z_2 \in \partial\mathbb{D}$, there are $z_m \in \mathbb{D}$ such that $z_m \to z_2$. Then, by (8.1.17), we have

$$\begin{aligned}
&z_2 + \mathrm{E}s_n(z_2) \\
&= \lim_m (z_m + \mathrm{E}s_n(z_m)) = \lim_m (z_m + \delta(z_m) + s(z_m + \delta(z_m)) \\
&= -\lim_m s_{(1)}(z_m + \delta(z_m)) = -s_{(1)}(z_2 + \delta(z_2)).
\end{aligned}$$

The two identities above imply that $|s_{(1)}(z_2+\delta(z_2))| = 1$, which implies that $z_2 + \delta(z_2) = \pm 2$ and that $s_{(1)}(z_2 + \delta(z_2)) = \pm 1$. Again, using the identity above, $z_2 + \mathrm{E}s_n(z_2) = \pm 1$, a real number, which violates the assumption that $z_2 \in \mathbb{C}^+$ since $\Im(z_2 + \mathrm{E}s_n(z_2) > 0$.

We shall proceed with our proofs using the following steps. Prove that $|\delta|$ is "small" both in its absolute value and in the integral of its absolute value with respect to u. Then, find a bound of $s_n(z) - s(z)$ in terms of δ. First, let us begin to estimate $|\delta|$.

For brevity, define

$$\begin{aligned} b_n &= b_n(z) =: (z + \mathrm{E}s_n(z))^{-1} = -\mathrm{E}s_n(z) + \delta, \\ \beta_k &= \beta_k(z) =: (z + \mathrm{E}s_n(z) - \varepsilon_k)^{-1}. \end{aligned}$$

By (8.1.18), we have $b_n(z) = -s(z + \delta)$, and hence, by (8.1.11),

$$|b_n(z)| < 1 \text{ for all } z \in \mathbb{C}^+. \tag{8.1.19}$$

By (8.1.15), we have

$$\begin{aligned} |\delta| &= \left| \frac{1}{n} \sum_{k=1}^{n} \mathrm{E}(\beta_k \varepsilon_k) \right| \\ &= \left| \frac{1}{n} \sum_{k=1}^{n} (b_n^2 \mathrm{E}\varepsilon_k + b_n^3 \mathrm{E}\varepsilon_k^2 + b_n^4 \mathrm{E}\varepsilon_k^3 + b_n^4 \mathrm{E}\varepsilon_k^4 \beta_k) \right| \\ &\le \frac{1}{n} \sum_{k=1}^{n} |\mathrm{E}\varepsilon_k| + \mathrm{E}|\varepsilon_k^2| + \mathrm{E}|\varepsilon_k^3| + v^{-1}\mathrm{E}|\varepsilon_k^4| \\ &= : J_1 + J_2 + J_3 + J_4. \end{aligned} \tag{8.1.20}$$

By Lemma 8.6,

$$J_1 \le \frac{1}{nv}. \tag{8.1.21}$$

Applying Lemmas 8.6 and 8.7 to be given in Subsection 8.1.3, we obtain, for all large n and some constant C,

$$|J_2| \le \frac{1}{n} \sum_{k=1}^{n} \mathrm{E}|\varepsilon_k|^2 \le \frac{C(v + \Delta)}{nv^2}. \tag{8.1.22}$$

Similar to (8.1.27), we have, for some constant C,

$$\begin{aligned} &\mathrm{E}|\varepsilon_k^4| = C[\mathrm{E}|\varepsilon_k - \mathrm{E}\varepsilon_k|^4 + |\mathrm{E}^4(\varepsilon_k)|] \\ &= \frac{C}{n^2}\mathrm{E}|x_{kk}|^4 + \frac{C}{n^4}\mathrm{E}\Big|\alpha_k'(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\alpha_k \\ &\quad -\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\Big|^4 + \frac{C}{n^4}\mathrm{E}\Big|\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1} \end{aligned}$$

$$-\mathrm{E}\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\Big|^4 + C|\mathrm{E}^4(\varepsilon_k)|. \tag{8.1.23}$$

We shall use the following estimates:

(i) $n^{-2}\mathrm{E}|x_{kk}|^4 \le n^{-3/2}\sigma^2$ (by truncation).

(ii)
$$\begin{aligned}
&n^{-4}\mathrm{E}\Big|\alpha_k'(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\alpha_k - \mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\Big|^4\\
&\le Cn^{-4}\Big[\mathrm{E}\nu_8\mathrm{tr}\big((\mathbf{W}_n(k) - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1}\big)^{-2}\\
&\quad +\big(\nu_4\mathrm{tr}((\mathbf{W}_n(k) - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})^{-1}\big)^2\Big] \text{ (Lemma B.26)}\\
&\le Cn^{-4}\mathrm{E}\Big(n^{1/2}v^{-3}\Im\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\\
&\quad +v^{-2}\mathrm{tr}\Big(\Im\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\Big)^2\Big) \qquad \text{(by } \nu_8 \le n^{1/2}\text{)}\\
&\le Cn^{-4}\mathrm{E}\left(n^{1/2}v^{-3}\left[n\Im s_n(z) + v^{-1}\right] + v^{-2}n^2(\Im s_n(z))^2 + v^{-4}\right)\\
&\le Cn^{-4}\Big(n^{1/2}v^{-3}[n(|\mathrm{E}s_n(z) - s(z)| + |s(z)|) + v^{-1}]\\
&\quad +v^{-2}n^2\big(\mathrm{E}|s_n(z) - \mathrm{E}s_n(z)|^2 + |\mathrm{E}s_n(z) - s(z)|^2 + |s(z)|^2\big)\\
&\quad +v^{-4}\Big)\\
&\le Cn^{-1}v^{-2}\left(n^{-1/2}(v+\varDelta) + (v+\varDelta)^2\right) \quad \text{(by Lemma 8.7).}
\end{aligned}$$

(iii)
$$\begin{aligned}
&n^{-4}\mathrm{E}\Big|\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1} - \mathrm{E}\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\Big|^4\\
&\le Cn^{-4}\left[n^4\mathrm{E}|s_n(z) - \mathrm{E}s_n(z)|^4 + v^{-4}\right]\\
&\le Cn^{-1}v^{-2}[n^{-1/2}(v+\varDelta) + (v+\varDelta)^2 + n^{-2}] \text{ (Lemma 8.7).}
\end{aligned}$$

Substituting these into (8.1.23), we obtain

$$\mathrm{E}|\varepsilon_k^4| \le C(n^{-3/2} + n^{-1}v^{-2}(v+\varDelta)^2),$$

which implies that

$$|J_4| \le Cv^{-1}(n^{-3/2} + n^{-1}v^{-2}(v+\varDelta)^2). \tag{8.1.24}$$

By the elementary inequality $|a|^3 \le \frac{1}{2}(|a|^2 + |a|^4)$, we notice that

$$\mathrm{E}|\varepsilon_k|^3 \le \frac{1}{2}(\mathrm{E}|\varepsilon_k|^2 + \mathrm{E}|\varepsilon_k|^4),$$

$$\begin{aligned}|J_3| &\le \frac{1}{n}\sum_{k=1}^{n}(\mathrm{E}|\varepsilon_k|^2+\mathrm{E}|\varepsilon_k|^4)\\ &\le C(v+\varDelta)n^{-1}v^{-2}.\end{aligned}$$

Therefore, we obtain

$$|\delta| \le C_0\left(\frac{1}{nv}+\frac{v+\varDelta}{nv^2}+\frac{n^{-1/2}(v+\varDelta)+(v+\varDelta)^2}{nv^3}\right). \tag{8.1.25}$$

By Lemma 8.8, if $|\delta| < v$, then $\varDelta < C_1 v$. Choose $M > C_0(2+C_1)^2$ and consider the set

$$\mathcal{E}_v = \left\{v \in \left[\sqrt{M/n}, \frac{1}{3}\right],\ |\delta| < v\right\}.$$

First, choose $v_0 = (9C_0/n)^{1/3}$ (which is less than $\frac{1}{3}$ for all large n). Since $\varDelta \le 1$, we have

$$|\delta| \le C_0\left(\frac{1}{nv^3}+\frac{2}{nv^3}+\frac{6}{nv^3}\right) < v_0.$$

Thus, $v_0 \in \mathcal{E}_v$. Now, let $v_1 = \inf \mathcal{E}_v$. We show that $v_1 = \sqrt{M/n}$.

If that is not the case, assume that $nv_1^2 > M + \omega_0$. Choose $\omega_1 = \min\{\omega_0/(4\sqrt{n}v_1), \omega_0/(24C_0C_1)\}$ and define $v_2 = v_1 - \omega_1/\sqrt{n}$. Then, by Lemma 8.8,

$$\varDelta < C_1\Big(v_2 + \frac{2\omega_1}{\sqrt{n}}\Big).$$

Consequently, letting $z = u + iv_2$, we then have

$$\begin{aligned}|\delta| &\le C_0\bigg(\frac{1}{nv_2}+\frac{2(v_2+C_1(v_2+2\omega_1/\sqrt{n}))}{nv_2^2}\\ &\qquad +\frac{(v_2+C_1(v_2+2\omega_1/\sqrt{n}))^2}{nv_2^3}\bigg)\\ &\le C_0\left(\frac{(2+C_1)^2+12C_1\omega_1}{nv_2}\right)\\ &\le \frac{M+\omega_0/2}{nv_2} \le \frac{nv_1^2-\omega_0/2}{nv_2^2}v_2 < v_2.\end{aligned}$$

This shows that $v_2 \in \mathcal{E}_v$, which contradicts the definition of v_1.

Finally, applying Lemma 8.8, Theorem 8.2 is proved.

8.1.3 Some Lemmas on Preliminary Calculation

Lemma 8.6. *Under the conditions of Theorem 8.2, we have*

(i) $|\mathrm{E}\varepsilon_k| \le 1/nv$,

(ii) $\mathrm{E}|\varepsilon_k|^2 \le Cn^{-1}\mathrm{E}|s_n(z) - \mathrm{E}s_n(z)|^2 + C(v+\Delta)n^{-1}v^{-2}$.

Proof. Recalling the definition of ε_k in (8.1.14) and applying (A.1.12), we obtain

$$\begin{aligned}|\mathrm{E}\varepsilon_k| &= \frac{1}{n}\left|\mathrm{E}\left[\mathrm{tr}(\mathbf{W}_n - z\mathbf{I}_n)^{-1} - \mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\right]\right| \\ &\le \frac{1}{nv}. \end{aligned} \tag{8.1.26}$$

Next, we estimate $\mathrm{E}|\varepsilon_k^2|$. Recalling definition (8.1.14), we have

$$\begin{aligned} &\mathrm{E}|\varepsilon_k^2| = \mathrm{E}|\varepsilon_k - \mathrm{E}\varepsilon_k|^2 + |\mathrm{E}^2(\varepsilon_k)| \\ &= \frac{1}{n}\mathrm{E}|x_{kk}|^2 + \frac{1}{n^2}\mathrm{E}\Big|\alpha_k'(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\alpha_k \\ &\quad -\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\Big|^2 + \frac{1}{n^2}\mathrm{E}\Big|\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1} \\ &\quad -\mathrm{E}\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\Big|^2 + |\mathrm{E}^2(\varepsilon_k)|. \end{aligned} \tag{8.1.27}$$

Then, by Lemma B.26, we have

$$\begin{aligned} &\frac{1}{n^2}\mathrm{E}\left|\alpha_k'(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\alpha_k - \mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\right|^2 \\ &\le \frac{C}{n^2}\mathrm{E}\mathrm{tr}\left((\mathbf{W}_n(k) - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1}\right)^{-1} \\ &= \frac{C}{n^2 v}\Im\left(\mathrm{E}\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\right) \\ &\le \frac{C}{n^2 v}\left|\mathrm{E}\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1} - \mathrm{tr}(\mathbf{W}_n - z\mathbf{I}_n)^{-1}\right| \\ &\quad + \frac{C}{nv}\Big[|\mathrm{E}s_n(z) - s(z)| + |s(z)|\Big] \\ &\le \frac{C(v+\Delta)}{nv^2}. \end{aligned} \tag{8.1.28}$$

Here the estimate of the first term follows from (A.1.12) and that of the second term follows from Lemma B.22.

Again, by (A.1.12), we have

$$\begin{aligned} &\frac{1}{n^2}\mathrm{E}\Big|\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1} - \mathrm{E}\mathrm{tr}(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\Big|^2 \\ &\le \frac{2}{n^2}\mathrm{E}\Big|\mathrm{tr}(\mathbf{W}_n - z\mathbf{I}_n)^{-1} - \mathrm{E}\mathrm{tr}(\mathbf{W}_n - z\mathbf{I}_n)^{-1}\Big|^2 + \frac{8}{n^2v^2} \\ &= \frac{2}{n}|s_n(z) - \mathrm{E}s_n(z)|^2 + \frac{8}{n^2v^2}. \end{aligned} \tag{8.1.29}$$

Finally, by (8.1.26)–(8.1.29), the second conclusion of the lemma follows.

Lemma 8.7. *Assume that $v > n^{-1/2}$. Under the conditions of Theorem 8.2, for all $\ell \geq 1$,*

$$\mathrm{E}|s_n(z) - \mathrm{E}s_n(z)|^{2\ell} \leq Cn^{-2\ell}v^{-4\ell}(\Delta + v)^{\ell}.$$

Proof. Let

$$\begin{aligned}\gamma_k &= \mathrm{E}_{k-1}\mathrm{tr}(\mathbf{W}_n - z\mathbf{I}_n)^{-1} - \mathrm{E}_k\mathrm{tr}(\mathbf{W}_n - z\mathbf{I}_n)^{-1} \\ &= \mathrm{E}_{k-1}\sigma_k - \mathrm{E}_k\sigma_k,\end{aligned}$$

where E_k denotes the conditional expectation given $\{x_{ij}, k+1 \leq i < j \leq n\}$ and

$$\begin{aligned}\sigma_k &= \left[\mathrm{tr}(\mathbf{W}_n - z\mathbf{I}_n)^{-1} - (\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\right] \\ &= \beta_k\left(1 + \frac{1}{n}\boldsymbol{\alpha}_k^*(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k\right). \qquad (8.1.30)\end{aligned}$$

Note that $\{\gamma_k\}$ forms a martingale difference sequence and

$$s_n(z) - \mathrm{E}s_n(z) = \frac{1}{n}\sum_{k=1}^{n}\gamma_k.$$

Since $2\ell > 1$, by Lemma 2.13, we have

$$\begin{aligned}&\mathrm{E}|s_n(z) - \mathrm{E}s_n(z)|^{2\ell} \\ &\leq \frac{C}{n^{2\ell}}\left[\mathrm{E}\left(\sum_{k=1}^{n}\mathrm{E}_k|\gamma_k|^2\right)^{\ell} + \sum_{k=1}^{n}\mathrm{E}|\gamma_k|^{2\ell}\right]. \qquad (8.1.31)\end{aligned}$$

By (A.1.12), we have

$$|\sigma_k| \leq v^{-1}. \qquad (8.1.32)$$

Write

$$\begin{aligned}\mathbf{A}_k &= (\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-2}, \\ \tilde{\varepsilon}_k &= n^{-1/2}x_{kk} - n^{-1}\alpha_k^*(\mathbf{W}_n(k) - z\mathbf{I}_{n-1})^{-1}\alpha_k + s_n(z), \\ \tilde{b}_n &= \frac{1}{z + s_n(z)}.\end{aligned}$$

Recall that $\beta_k = -\tilde{b}_n - \tilde{b}_n\beta_k\tilde{\varepsilon}_k$. Similar to (8.1.19), one may prove that

$$|\tilde{b}_n| < 1.$$

Substituting these into (8.1.30) and noting that

$$\begin{aligned}&\mathrm{E}_{k-1}\left[1+\frac{1}{n}\boldsymbol{\alpha}_k^*\mathbf{A}_k\boldsymbol{\alpha}_k\right]-\mathrm{E}_k\left[1+\frac{1}{n}\boldsymbol{\alpha}_k^*\mathbf{A}_k\boldsymbol{\alpha}_k\right]\\&=\frac{1}{n}\mathrm{E}_{k-1}[\boldsymbol{\alpha}_k^*\mathbf{A}_k\boldsymbol{\alpha}_k-\mathrm{tr}(\mathbf{A}_k)],\end{aligned}$$

we may rewrite

$$\gamma_k=-\frac{1}{n}\mathrm{E}_{k-1}\tilde{b}_n(\boldsymbol{\alpha}_k^*\mathbf{A}_k\boldsymbol{\alpha}_k-\mathrm{tr}(\mathbf{A}_k))+[\mathrm{E}_{k-1}\tilde{b}_n(\sigma_k\tilde{\varepsilon}_k)-\mathrm{E}_k\tilde{b}_n(\sigma_k\tilde{\varepsilon}_k)]. \quad (8.1.33)$$

Employing Lemma B.26, we have

$$\begin{aligned}\mathrm{E}_k|\gamma_k|^2&\le\frac{2}{n^2}\mathrm{E}_k|\tilde{b}_n(\boldsymbol{\alpha}_k^*\mathbf{A}_k\boldsymbol{\alpha}_k-\mathrm{tr}(\mathbf{A}_k))|^2+\frac{2}{v^2}\mathrm{E}_k|\tilde{b}_n\tilde{\varepsilon}_k|^2\\&\le\frac{2}{n^2}\mathrm{E}_k|\boldsymbol{\alpha}_k^*\mathbf{A}_k\boldsymbol{\alpha}_k-\mathrm{tr}(\mathbf{A}_k)|^2+\frac{2}{v^2}\mathrm{E}_k|\tilde{\varepsilon}_k|^2\\&\le\frac{C}{n^2}\mathrm{E}_k(\mathrm{tr}(\mathbf{A}_k\mathbf{A}_k^*))+v^{-2}\Bigg[n^{-1}\mathrm{E}|x_{kk}|^2\\&\quad+\mathrm{E}_k\bigg|\boldsymbol{\alpha}_k^*(\mathbf{W}_n(k)-z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k\\&\quad-\frac{1}{n}\mathrm{tr}(\mathbf{W}_n(k)-z\mathbf{I}_{n-1})^{-1}\bigg|^2\\&\quad+\mathrm{E}_k\left|\frac{1}{n}\mathrm{tr}(\mathbf{W}_n-z\mathbf{I}_{n-1})^{-1}-s_n(z)\right|^2\Bigg]\\&\le\frac{C}{nv^3}\left[\mathrm{E}_k(\Im s_n(z))+v+n^{-1}v^{-1}\right]. \qquad (8.1.34)\end{aligned}$$

Thus, we have, by noting $v^2>n^{-1}$,

$$\begin{aligned}&\text{The first term on the right-hand side of (8.1.31)}\\&\le\frac{C}{n^{2\ell}v^{4\ell}}\left[v^\ell\mathrm{E}(\Im s_n(z))^\ell+v^\ell\right]. \qquad (8.1.35)\end{aligned}$$

Furthermore, by Lemma B.26 and the fact that $\nu_{4\ell}\le Cn^{\ell-1}$,

$$\begin{aligned}&\mathrm{E}\left|\frac{1}{n}[(\boldsymbol{\alpha}_k^*\mathbf{A}_k\boldsymbol{\alpha}_k-\mathrm{tr}(\mathbf{A}_k)]\right|^{2\ell}\\&\le C_\ell n^{-2\ell}\left[\nu_{4\ell}\mathrm{E}\mathrm{tr}[(\mathbf{A}_k\mathbf{A}_k^*)^\ell]+\nu_4^\ell\mathrm{E}[\mathrm{tr}(\mathbf{A}_k\mathbf{A}_k^*)]^\ell\right]\\&\le C_\ell\left[v^{-4\ell+1}n^{-\ell}\mathrm{E}(\Im s_n(z))+n^{-\ell}v^{-3\ell}\mathrm{E}(\Im s_n(z))^\ell\right]. \qquad (8.1.36)\end{aligned}$$

Similarly, by noting $\mathrm{E}|x_{kk}|^{2\ell}\le\sigma^2n^{(\ell-1)/2}$, we have

$$\begin{aligned}&\mathrm{E}|\tilde{\varepsilon}_k|^{2\ell}\le C\Bigg[\mathrm{E}|n^{-1/2}x_{kk}|^{2\ell}\\&+\mathrm{E}\left|n^{-1}\left[\boldsymbol{\alpha}_k^*(\mathbf{W}_n(k)-z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k-\mathrm{tr}(\mathbf{W}_n(k)-z\mathbf{I}_{n-1})^{-1}\right]\right|^{2\ell}\end{aligned}$$

$$+n^{-2\ell}\mathrm{E}\Big|\mathrm{tr}\big(\mathbf{W}_n(k)-z\mathbf{I}_{n-1}\big)^{-1}-\mathrm{tr}(\mathbf{W}_n-z\mathbf{I}_n)^{-1}\Big|^{2\ell}\Big]$$
$$\leq C\Big[n^{-(\ell+1)/2}+n^{-\ell}v^{-2\ell+1}\mathrm{E}(\Im s_n(z))$$
$$+n^{-\ell}v^{-\ell}\mathrm{E}[\Im(s_n(z))]^{\ell}+n^{-2\ell}v^{-2\ell}\Big]. \tag{8.1.37}$$

Thus, by the two estimates above and recalling (8.1.33), we have

$$\text{The second term on the right-hand side of (8.1.31)}$$
$$\leq \frac{C}{n^{2\ell}v^{4\ell}}\big[n^{-\ell+1}v\mathrm{E}(\Im s_n(z))+v^{\ell}n^{-\ell+1}\mathrm{E}(\Im s_n(z))^{\ell}$$
$$+v^{2\ell}n^{-(\ell-1)/2}\big]. \tag{8.1.38}$$

Substituting (8.1.35) and (8.1.38) into (8.1.31), we obtain

$$\mathrm{E}|s_n(z)-\mathrm{E}s_n(z)|^{2\ell}$$
$$\leq \frac{C}{n^{2\ell}v^{4\ell}}\left[n^{-\ell+1}v\mathrm{E}(\Im s_n(z))+v^{\ell}\mathrm{E}(\Im s_n(z))^{\ell}+v^{\ell}\right]. \tag{8.1.39}$$

First, we note that

$$0<\mathrm{E}\Im s_n(z)\leq|\mathrm{E}s_n(z)-s(z)|+|s(z)|\leq\Delta/v+1.$$

The lemma then follows if $\ell=1$.

To treat the term $\mathrm{E}(\Im s_n(z))^{\ell}$ when $\ell>1$, we need to employ induction.

Now, we extend the conclusion to the case where $\frac{1}{2}<\ell<1$. Applying Lemma 2.12, we obtain

$$\begin{aligned}\mathrm{E}|s_n(z)-\mathrm{E}s_n(z)|^{2\ell} &\leq Cn^{-2\ell}\mathrm{E}\left(\sum_{k=1}^{n}|\gamma_k|^2\right)^{\ell}\\ &\leq Cn^{-2\ell}\left(\sum_{k=1}^{n}\mathrm{E}|\gamma_k|^2\right)^{\ell}\\ &\leq Cn^{-2\ell}\left(\sum_{k=1}^{n}n^{-1}v^{-3}\mathrm{E}\Im s_n(z)+v\right)^{\ell}\\ &\leq Cn^{-2\ell}(v^{-4\ell}(v+\Delta)+n^{\ell}v^{\ell})<C.\end{aligned}$$

This shows that when $1<\ell<2$,

$$\begin{aligned}\mathrm{E}(\Im s_n(z))^{\ell} &\leq 2\mathrm{E}|s_n(z)-\mathrm{E}s_n(z)|^{\ell}+2(\mathrm{E}\Im s_n(z))^{\ell}\\ &\leq C(1+(1+\Delta/v)^{\ell}).\end{aligned}$$

This, together with (8.1.39), implies that the lemma holds for $1\leq\ell<2$.

Then, the lemma follows by induction and (8.1.39). The proof of the lemma is complete.

Lemma 8.8. *If* $|\delta| < v < \frac{1}{3}$ *for all* $z = u + iv$*, then* $\Delta \le Cv$*.*

Proof. From (8.1.10) and (8.1.17), we have

$$\mathrm{E}s_n(z) - s(z) = \frac{1}{2}\delta\left[1 + \frac{2z+\delta}{\sqrt{z^2-4} + \sqrt{(z+\delta)^2-4}}\right].$$

By the convention for the square root of complex numbers, we know that the real parts of $\sqrt{z^2-4}$ and $\sqrt{(z+\delta)^2-4}$ are $\mathrm{sgn}(uv) = \mathrm{sgn}(u)$ and $\mathrm{sgn}(u + \Re\delta)(v + \Im\delta) = \mathrm{sgn}(u + \Re\delta)$. Therefore, if $|u| > v > |\delta|$, then both the real and imaginary parts of $\sqrt{z^2-4}$ and $\sqrt{(z+\delta)^2-4}$ have the same signs. Therefore, for $v < |u| < 16$, we have

$$|\mathrm{E}s_n(z) - s(z)| \le \frac{1}{2}|\delta|\left[1 + \frac{32}{\sqrt{|u^2-4|}}\right].$$

When $|u| < v < \frac{1}{3}$, then

$$|\mathrm{E}s_n(z) - s(z)| \le \frac{1}{2}|\delta|\left[1 + \frac{32}{\Im(\sqrt{z^2-4})}\right] \le \frac{1}{2}|\delta|\left[1 + \frac{32}{2-2v}\right].$$

This simply implies that

$$\int_{16}^{-16} |\mathrm{E}s_n(z) - s(z)|du \le \eta v,$$

where $\eta = \frac{1}{2}\int_{-16}^{16}\left[1 + \frac{32}{\sqrt{|u^2-4|}}\right]du + 1$. Then, the lemma follows from Corollary B.15.

8.2 Further Extensions

In this section, we shall consider the convergence rate in probability and almost surely.

Theorem 8.9. *Under conditions (i)–(iii) in (8.1.1), we have*

$$\|F^{(1/\sqrt{n}\mathbf{W})} - F\| = O_p(n^{-2/5}), \tag{8.2.1}$$

$$\|F^{(1/\sqrt{n}\mathbf{W})} - F\| = O_{\mathrm{a.s.}}(n^{-2/5+\eta}), \quad \textit{for any } \eta > 0. \tag{8.2.2}$$

In the proof of Theorem 8.2, we have already proved that we may assume that the elements of the matrix $\mathbf{W}$ are truncated at $n^{1/4}$ and then recentralized and rescaled. Then, by Corollary A.41, to prove (8.2.1), we only need to show that, for $z = u + iv$, $v = n^{-2/5}$,

$$\mathrm{E}\int_{-16}^{16} |s_n(z) - s(z)|du = O(v).$$

In the proof of Theorem 8.2, we have proved that

$$\int_{-16}^{16} |\mathrm{E}s_n(z) - s(z)|du = O(v).$$

Thus, to prove Theorem 8.9, one only needs to show that

$$\int_{-16}^{16} |s_n(z) - \mathrm{E}s_n(z)|du = O(v).$$

This follows from Lemma 8.7 and the following argument: for $v = n^{-2/5}$,

$$\int_{-16}^{16} |s_n(z) - \mathrm{E}s_n(z)|^2 du \le Cn^{-2}v^{-3} = O(v^2).$$

Conclusion (8.2.2) follows from the argument that, for $v = n^{-2/5+\eta}$,

$$\int_{-16}^{16} |s_n(z) - \mathrm{E}s_n(z)|^{2\ell} du \le Cn^{-2\ell}v^{-3\ell} = O(v^{2\ell}n^{-5\ell\eta}).$$

Here, we choose ℓ such that $5\ell\eta > 1$. Thus, Theorem 8.9 is proved.

8.3 Convergence Rates of the Expected ESD of Sample Covariance Matrices

In this section, we shall establish some convergence rates for the expected ESD of the sample covariance matrices.

8.3.1 Assumptions and Results

Let $\mathbf{S}_n = n^{-1}\mathbf{X}_n\mathbf{X}_n^* : \ p \times p$, where $\mathbf{X}_n = (x_{ij}(n), i = 1, \cdots, p, \ j = 1, \cdots, n)$. Assume that the following conditions hold:

$$\begin{aligned} &\text{(i) For each } n,\ x_{ij}(n) \text{ are independent.} \\ &\text{(ii) } \mathrm{E}x_{ij}(n) = 0 \ \textit{and}\ \mathrm{E}|x_{ij}^2(n)| = 1 \ \text{ for all } \ i,\ j. \\ &\text{(iii) } \sup_n \sup_{i,j} \mathrm{E}|x_{ij}^6(n)| < \infty. \end{aligned} \tag{8.3.1}$$

Throughout this section, for brevity, we shall drop the index n from the entries of $\mathbf{X}_n$ and $\mathbf{S}_n$.

Denote by F_p the ESD of the matrix $\mathbf{S}_n$. Under the conditions in (8.3.1), it is well known (see Theorem 3.10) that $F_p \xrightarrow{w} F_y$ a.s., where $y = \lim_{n\to\infty}(p/n) \in (0,\infty)$ and F_y is the limiting spectral distribution of F_p, known as the Marčenko-Pastur [201] distribution, which has a mass of $1 - y^{-1}$ at the origin when $y > 1$ and has density

$$F_y'(x) = \frac{1}{2xy\pi}\sqrt{4y - (x - y - 1)^2}I_{[a,b]}(x), \tag{8.3.2}$$

with $a = a(y) = (1 - \sqrt{y})^2$ and $b = b(y) = (1 + \sqrt{y})^2$. In this section, we shall establish the following theorem.

Theorem 8.10. *Under the assumptions in (8.3.1), we have*

$$\|\mathrm{E}F_p - F_{y_p}\| = \begin{cases} O\left(n^{-1/2}a^{-1}\right), & \text{if } a > n^{-1/3}, \\ O(n^{-1/6}), & \text{otherwise,} \end{cases} \tag{8.3.3}$$

where $y_p = p/n \le 1$.

Remark 8.11. Because the convergence rate of $|y_p - y|$ may be arbitrarily slow, it is impossible to establish any rate for the convergence of $\|\mathrm{E}F_p - F_y\|$ if we know nothing about the convergence rate of $|y_p - y|$. Conversely, if we know the convergence rate of $|y_p - y|$, then from (8.3.3) we can easily derive a convergence rate for $\|\mathrm{E}F_p - F_y\|$. This is the reason why F_{y_p}, instead of the limit distribution F_y, is used in Theorem 8.10.

Remark 8.12. If $y_p > 1$, consider the sample covariance matrix $\frac{1}{p}\mathbf{X}^*\mathbf{X}$ whose ESD is denoted by $G_n(x)$. Noting that the matrices $\mathbf{X}\mathbf{X}^*$ and $\mathbf{X}^*\mathbf{X}$ have the same set of nonzero eigenvalues, we have the relation

$$F_p(x) = y_p^{-1}G_n(y_p^{-1}x) + (1 - y_p^{-1})\delta(x).$$

Therefore, we have

$$\|F_p - F_{y_p}\| = y_p^{-1}\|G_n - F_{1/y_p}\|.$$

Therefore, the convergence rate for the case $y_p > 1$ can be derived from Theorem 8.10 with $y_p < 1$. This is the reason we only consider the case where $y_p \le 1$ in Theorem 8.10.

To better understand the notation of the convergence rates, let us see the following special cases.

Corollary 8.13. *Assume the conditions of Theorem 8.10 hold.*

If $y_p = (1 - \delta)^2$ *for some constant* $\delta > 0$, *then* $\|\mathrm{E}F_p - F_{y_p}\| = O(n^{-1/2})$.

If $y_p \ge (1 - n^{-1/6})^2$, *then* $\|\mathrm{E}F_p - F_{y_p}\| = O(n^{-1/6})$.

If $y_p = (1 - n^{-\eta})^2$ *for some* $0 < \eta < \frac{1}{6}$, *then* $\|\mathrm{E}F_p - F_{y_p}\| = O(n^{-(1-4\eta)/2})$.

For brevity of notation, we shall drop the index p from y_p in the remainder of this section.

The steps in the proof follow along the same lines as in Theorem 8.2.

8.3.2 Truncation and Centralization

We first truncate the variables x_{ij} at $n^{1/4}$ and then centralize and rescale the variables. Let $\widetilde{\mathbf{X}}_n$ and $\mathbf{D}_n$ denote the $p \times n$ matrix of the truncated variables $\tilde{x}_{ij} = x_{ij} I(|x_{ij}| < n^{1/4})$ and that of the rescalers, i.e., $\mathbf{D}_n = \left(\sigma_{jk}^{-1}\right)_{p \times n}$ with $\sigma_{jk}^2 = \mathrm{var}(\tilde{x}_{ij})$. Write $\widehat{\mathbf{X}}_n = \widetilde{\mathbf{X}}_n - \mathrm{E}\widetilde{\mathbf{X}}_n$ and $\mathbf{Y}_n = \widehat{\mathbf{X}}_n \circ \mathbf{D}_n$, where $\circ$ denotes the Hadamard product of matrices. Further, denote by $F_p^{(t)}$, $F_p^{(tc)}$, and $F_p^{(s)}$ the ESDs of the sample covariance matrices $\frac{1}{n}\widetilde{\mathbf{X}}_n\widetilde{\mathbf{X}}_n^*$, $\frac{1}{n}\widehat{\mathbf{X}}_n\widehat{\mathbf{X}}_n^*$, and $\frac{1}{n}\mathbf{Y}_n\mathbf{Y}_n^*$, respectively. Then, by the rank inequality (see Theorem A.44) and the norm inequality (see Theorem A.47), we have

$$\|F_p - F_p^{(t)}\| \le \frac{1}{p}\sum_{jk} I_{\{|x_{jk}| \ge n^{1/4}\}},$$

$$L(F_p^{(t)}, F_p^{(tc)}) \le 2\left\|\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n\right\| \left\|\frac{1}{\sqrt{n}}\mathrm{E}(\widetilde{\mathbf{X}}_n)\right\| + \left\|\frac{1}{\sqrt{n}}\mathrm{E}(\widetilde{\mathbf{X}}_n)\right\|^2,$$

and

$$\begin{aligned} &L(F_p^{(tc)}, F_p^{(s)}) \\ &\le 2\left\|\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n\right\|^2 \left[\left\|\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n \circ (\mathbf{D}_n - \mathbf{J})\right\| + \left\|\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n \circ (\mathbf{D}_n - \mathbf{J})\right\|^2\right], \end{aligned}$$

where $\mathbf{J}$ is the $p \times n$ matrix of all entries 1.

Similar to the proof of Lemma 8.3, under condition (8.3.1), applying Bernstein's inequality, one can show that, for any $\eta > 0$,

$$\begin{aligned} \mathrm{E}\|F_p - F_p^{(t)}\| &\le p^{-1}\sum_{ij} P(|x_{ij}| > n^{1/4}) = O(n^{-1/2}), \\ \|F_p - F_p^{(t)}\| &\le \frac{1}{p}\sum_{jk} I_{\{|x_{jk}| \ge n^{1/4}\}} = o_{\text{a.s.}}(n^{-1/2+\eta}), \ \text{a.s.} \end{aligned}$$

By Theorem 5.9,

$$\limsup \left\|\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n\right\| \le (1 + \sqrt{y}), \ \text{a.s.},$$

and by elementary calculus, one gets

$$\left\| \frac{1}{\sqrt{n}} \mathrm{E}(\widetilde{\mathbf{X}}_n) \right\| \leq \sqrt{n} \max_{jk} \mathrm{E}|x_{jk}| I_{\{|x_{jk}| \geq n^{1/4}\}} = O(n^{-3/4})$$

and, by the fact that $\max |1 - \sigma_{jk}| = O(n^{-1})$,

$$\begin{aligned}
&\left\| \frac{1}{\sqrt{n}} \widehat{\mathbf{X}}_n \circ (\mathbf{D}_n - \mathbf{J}) \right\|^2 \\
&\leq \frac{1}{n} \sum_{jk} |\widehat{x}_{jk}|^2 (\sigma_{jk}^{-1} - 1)^2 \\
&\leq \frac{1}{n} \sum_{jk} |\widehat{x}_{jk}|^2 \max_{jk} |\sigma_{jk}^{-1} - 1|^2 \\
&= O_{\text{a.s.}}(n^{-1}).
\end{aligned}$$

These show that

$$\begin{aligned}
L(F_p^{(t)}, F_p^{(tc)}) &= O_{\text{a.s.}}(n^{-3/4}), \\
L(F_p^{(tc)}, F_p^{(s)}) &= O_{\text{a.s.}}(n^{-1/2}).
\end{aligned}$$

Applying Lemmas B.19 and 8.14, given in Section 8.4, we have

$$\|F_p - F_y\| \leq C \max\left\{ \|F^{(s)} - F_y\|, \frac{1}{\sqrt{na} + \sqrt[4]{n}} \right\}.$$

Thus, to prove Theorem 8.10 and Corollary 8.13, one can assume that the entries of $\mathbf{X}_n$ are truncated at $n^{-1/4}$, recentralized, and then rescaled.

8.3.3 Proof of Theorem 8.10

In Chapter 3, we derived that the Stieltjes transform of the M-P law with index y is given by

$$s_y(z) = -\frac{y + z - 1 - \sqrt{(1 + y - z)^2 - 4y}}{2yz}. \tag{8.3.4}$$

Because M-P distributions are weakly continuous in y, letting $y \uparrow 1$, we obtain

$$s_1(z) = -\frac{z - \sqrt{z^2 - 4z}}{2z}. \tag{8.3.5}$$

We point out that (8.3.4) is still true when $y > 1$, which can be easily derived through the dual case where $y < 1$.

Set

$$s_p(z) = \frac{1}{p} \mathrm{tr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1}, \tag{8.3.6}$$

where $z = u + iv$ with $v > 1/\sqrt{n}$. Similar to the proof of (3.3.2) in Chapter 3, one may show that

$$\begin{aligned}
\mathrm{E}s_p(z) &= \frac{1}{p}\sum_{k=1}^{p} \mathrm{E}\frac{1}{s_{kk} - z - n^{-2}\boldsymbol{\alpha}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k} \\
&= \frac{1}{p}\sum_{k=1}^{p} \mathrm{E}\frac{1}{\varepsilon_k + 1 - y - z - yz\mathrm{E}s_p(z)} \\
&= -\frac{1}{z + y - 1 + yz\mathrm{E}s_p(z)} + \delta, \qquad (8.3.7)
\end{aligned}$$

where

$$\begin{aligned}
s_{kk} &= \frac{1}{n}\sum_{j=1}^{n} |x_{kj}^2|, \\
\mathbf{S}_{nk} &= \frac{1}{n}\mathbf{X}_{nk}\mathbf{X}_{nk}^*, \\
\boldsymbol{\alpha}_k &= \mathbf{X}_{nk}\bar{\mathbf{x}}_k, \\
\varepsilon_k &= (s_{kk} - 1) + y + yz\mathrm{E}s_p(z) - \frac{1}{n^2}\boldsymbol{\alpha}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}\boldsymbol{\alpha}_k, \\
\delta = \delta_p &= -\frac{1}{p}\sum_{k=1}^{p} b_n \mathrm{E}\beta_k\varepsilon_k, \\
b_n = b_n(z) &= \frac{1}{z + y - 1 + yz\mathrm{E}s_p(z)}, \\
\beta_k = \beta_k(z) &= \frac{1}{z + y - 1 + yz\mathrm{E}s_p(z) - \varepsilon_k}, \qquad (8.3.8)
\end{aligned}$$

and $\mathbf{X}_{nk}$ is the $(p-1)\times n$ matrix obtained from $\mathbf{X}_n$ with its k-th row removed and $\mathbf{x}_k'$ is the k-th row of $\mathbf{X}_n$.

In Chapter 3, it was proved that one of the roots of equation (8.3.7) is

$$\mathrm{E}s_p(z) = -\frac{1}{2yz}\left(z + y - 1 - yz\delta - \sqrt{(z + y - 1 + yz\delta)^2 - 4yz}\right). \qquad (8.3.9)$$

By Lemma 8.16, we need only estimate $s_p(z) - s_y(z)$ for $z = u + iv, v > 0, |u| < A$, where A is a constant chosen according to (B.2.10).

As done in the proof of Theorem 8.2, we mainly concentrate on finding a bound for $|\delta|$ and postpone the technical proofs to the next subsection.

We now proceed to estimate $|\delta_n|$. First, we note that

$$|\beta_k| = \frac{1}{|z + y - 1 + yz\mathrm{E}s_p(z) - \varepsilon_k|} \leq v^{-1} \qquad (8.3.10)$$

by the fact that

$$\begin{aligned}&\Im(z+y-1+yzs_p(z)-\varepsilon_k)\\&=\Im\left(s_{kk}-z-\frac{1}{n^2}\boldsymbol{\alpha}_k^*(\mathbf{S}_{nk}-z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k\right)<-v.\end{aligned}$$

Then, by (8.3.8) and (8.3.10), we have

$$|\delta|\le\frac{1}{p}\sum_{k=1}^{p}(|b_n^2||\mathrm{E}\varepsilon_k|+|b_n^3|\mathrm{E}|\varepsilon_k|^2+|b_n^4|\mathrm{E}|\varepsilon_k|^3+|b_n|^4v^{-1}\mathrm{E}|\varepsilon_k^4|).\qquad(8.3.11)$$

In (3.3.15), it was proved that

$$|\mathrm{E}\varepsilon_k|\le\frac{C}{nv},\qquad(8.3.12)$$

where the constant C may take the value $yA+1$.

Next, the estimation of $\mathrm{E}|\varepsilon_k|^2$ needs to be more precise than in Chapter 3. By writing $\varDelta=\|\mathrm{E}F_p-F_y\|$, we have

$$\mathrm{E}|\varepsilon_k^2|\le\frac{C}{n}+R_1+R_2+|\mathrm{E}(\varepsilon_k)|^2,\qquad(8.3.13)$$

where the first term is a bound for $\mathrm{E}|s_{kk}-1|^2$,

$$\begin{aligned}R_1&=n^{-4}\mathrm{E}\left|\boldsymbol{\alpha}_k^*(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}\boldsymbol{\alpha}_k\right.\\&\qquad\left.-\mathrm{E}[(\boldsymbol{\alpha}_k'(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}\boldsymbol{\alpha}_k|\mathbf{X}_{nk})\right|^2\\&=n^{-4}\mathrm{E}\left|\mathbf{x}_k'\mathbf{X}_{nk}^*(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}\mathbf{X}_{nk}\bar{\mathbf{x}}_k\right.\\&\qquad\left.-\mathrm{tr}(\mathbf{X}_{nk}^*(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}\mathbf{X}_{nk})\right|^2\\&\le Cn^{-4}\mathrm{tr}(\mathbf{X}_{nk}^*(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}\mathbf{X}_{nk}\mathbf{X}_{nk}^*(\mathbf{S}_{nk}-\bar{z}\mathbf{I}_{p-1})^{-1}\mathbf{X}_{nk})\\&\qquad\text{(by Lemma B.26)}\\&=C\left(\frac{1}{n}+\frac{|u|^2}{n^2}\mathrm{Etr}((\mathbf{S}_{nk}-u\mathbf{I}_{p-1})^2+v^2\mathbf{I}_{p-1})^{-1}\right)\\&=C\left(\frac{1}{n}+\frac{|u|^2}{n^2v}\mathrm{E}\Im\mathrm{tr}(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}\right)\\&\le C\left(\frac{1}{n}+\frac{|u|^2}{n^2v}\mathrm{E}\Im\mathrm{tr}(\mathbf{S}_{n}-z\mathbf{I}_{p})^{-1}+\frac{2}{n^2v^2}\right)\\&\le C\left(\frac{1}{n}+\frac{|u|^2}{nv}[|\mathrm{E}s_p(z)-s_y(z)|+|s_y(z)|]\right)\\&\le C\left(\frac{1}{n}+\frac{|u|^2}{nv^2}(\varDelta+v/\sqrt{y}v_y)\right)\quad\text{(by Lemma B.22)},\end{aligned}\qquad(8.3.14)$$

where $v_y = \sqrt{a} + \sqrt{v} = 1 - \sqrt{y} + \sqrt{v}$, and

$$\begin{aligned}
R_2 &= \frac{|z|^2}{n^2} \mathrm{E}\left|\mathrm{tr}(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1} - \mathrm{Etr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1}\right|^2 \\
&\leq \frac{2|z|^2}{n^2}\left[\mathrm{E}\left|\mathrm{tr}(\mathbf{S}_n - z\mathbf{I}_{p-1})^{-1} - \mathrm{Etr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1}\right|^2 + \frac{1}{v^2}\right] \\
&= 2|z|^2\left[\mathrm{E}\left|s_p(z) - \mathrm{E}s_p(z)\right|^2 + \frac{1}{n^2v^2}\right] \\
&\leq C|z|^2\left[n^{-2}v^{-4}(\Delta + v/v_y) + \frac{1}{n^2v^2}\right] \quad \text{(by Lemma 8.20)} \\
&\leq \frac{C|z|^2}{n^2v^4}(\Delta + v/v_y). \qquad (8.3.15)
\end{aligned}$$

Substituting (8.3.12) and the estimates for R_1 and R_2 into (8.3.13), we obtain

$$\mathrm{E}|\varepsilon_k|^2 \leq C\left(\frac{1}{n} + \frac{|z|^2}{nv^2}(\Delta + v/v_y)\right).$$

We now estimate $\mathrm{E}|\varepsilon_k|^4$. At first, by noting $\nu_8 \leq Cn^{1/2}$, we have

$$\mathrm{E}|s_{kk} - 1|^4 \leq Cn^{-4}[\nu_8 n + \nu_4^2 n^2] \leq C/n^2.$$

Employing Lemma B.26, we obtain

$$\begin{aligned}
&n^{-8}\mathrm{E}\left|\boldsymbol{\alpha}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}\boldsymbol{\alpha}_k - \mathrm{tr}(\mathbf{X}_{nk}^*(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}\mathbf{X}_{nk})\right|^4 \\
&\leq Cn^{-4}\mathrm{E}\left[\nu_8\mathrm{tr}((\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-2}\mathbf{S}_{nk}^4(\mathbf{S}_{nk} - \bar{z}\mathbf{I}_{p-1})^{-2})\right. \\
&\quad \left. + \nu_4\mathrm{tr}((\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}\mathbf{S}_{nk}^2(\mathbf{S}_{nk} - \bar{z}\mathbf{I}_{p-1})^{-1})^2\right] \\
&\leq Cn^{-4}\mathrm{E}\left[\nu_8\mathrm{tr}(\mathbf{I} + |z|^4(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-2}(\mathbf{S}_{nk} - \bar{z}\mathbf{I}_{p-1})^{-2})\right. \\
&\quad \left. + \nu_4\mathrm{tr}(\mathbf{I} + |z|^2(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}(\mathbf{S}_{nk} - \bar{z}\mathbf{I}_{p-1})^{-1})^2\right] \\
&\leq Cn^{-2}\mathrm{E}\left[1 + n^{-1/2}|z|^4v^{-3}\Im(s_p(z)) + |z|^4v^{-2}(\Im(s_p(z)))^2\right] \\
&\leq Cn^{-2}[1 + |z|^4v^{-2}|\mathrm{E}s_p(z)| \\
&\qquad + |z|^4v^{-2}(\mathrm{E}|s_p(z) - \mathrm{E}s_p(z)|^2 + |\mathrm{E}s_p(z)|^2)] \\
&\leq Cn^{-2}[1 + |z|^4v^{-3}(\Delta + v/v_y) + |z|^4v^{-6}n^{-2}(\Delta + v/v_y) \\
&\quad + |z|^4v^{-4}(\Delta + v/v_y)^2] \\
&\leq Cn^{-2}[1 + |z|^4v^{-4}(\Delta + v/v_y)^2].
\end{aligned}$$

Applying Lemma 8.20, it follows that

$$\begin{aligned}
&= \frac{|z|^4}{n^4}\mathrm{E}\left|\mathrm{tr}(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1} - \mathrm{Etr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1}\right|^4 \\
&\leq \frac{C|z|^4}{n^4}\left[\mathrm{E}\left|\mathrm{tr}(\mathbf{S}_n - z\mathbf{I}_{p-1})^{-1} - \mathrm{Etr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1}\right|^4 + \frac{1}{v^4}\right]
\end{aligned}$$

$$
\begin{aligned}
&= C|z|^4 \left[\mathrm{E}\,|s_p(z) - \mathrm{E}s_p(z)|^4 + \frac{1}{n^4 v^4} \right] \\
&\le C|z|^4 \left[n^{-4} v^{-8} (\varDelta + v/v_y)^2 + \frac{1}{n^4 v^4} \right] \quad \text{(by Lemma 8.20)} \\
&\le \frac{C|z|^4}{n^4 v^8} (\varDelta + v/v_y)^2 \le \frac{C|z|^4}{n^2 v^4} (\varDelta + v/v_y)^2.
\end{aligned}
$$

Therefore, the three estimates above yield

$$
\mathrm{E}|\varepsilon_k|^4 \le \frac{C}{n^2} [1 + |z|^4 v^{-4} (\varDelta + v/v_y)^2].
$$

By Cauchy's inequality, we have

$$
\begin{aligned}
\mathrm{E}|\varepsilon_k|^3 &\le (\mathrm{E}|\varepsilon_k|^2 \mathrm{E}|\varepsilon_k|^4)^{1/2} \\
&\le C n^{-3/2} (1 + |z|^3 v^{-3} (\varDelta + v/v_y)^{3/2}).
\end{aligned}
$$

Substituting the estimates of the moments of ε_k above into (8.3.11) and noting that $|b_n| \le 1/\sqrt{y|z|}$ by Lemma 8.18, we obtain

$$
|\delta| \le C[|b_n|^2 n^{-1} v^{-1} + n^{-1} v^{-2} (\varDelta + v/v_y) + n^{-2} v^{-5} (\varDelta + v/v_y)^2].
$$

Define

$$
\mathcal{I} = \{M_0 v > n^{-1/2},\ |\delta| < v/[10(A+1)^2]\}. \tag{8.3.16}
$$

From (8.3.7), it follows that

$$
\begin{aligned}
|b_n|^2 &\le 2|\mathrm{E}s_p(z)|^2 + 2|\delta|^2 \\
&\le 2v^{-2} (\varDelta + v/v_y)^2 + 2|\delta|^2.
\end{aligned}
$$

Choose M_0 large. Then $C y^{-1} n^{-1} v^{-2} < \frac{1}{2}$. Consequently, $2Cn^{-1} v^{-1} |\delta_n| < \frac{1}{2}$, from which it follows that

$$
\begin{aligned}
|\delta| &\le C[n^{-1} v^{-3} (\varDelta + v/v_y)^2 + n^{-1} v^{-2} (\varDelta + v/v_y)] + \frac{1}{2}|\delta| \\
&\le C_0 n^{-1} v^{-3} (\varDelta + v/v_y)^2.
\end{aligned}
$$

By Lemma 8.21, if $v \in \mathcal{I}$, we have

$$
\varDelta \le C_1 v/v_y.
$$

Hence,

$$
|\delta_n| \le C_0 (C_1 + 1)^2 n^{-1} v_y^{-2}.
$$

At first, it is easy to verify that, for all large n, $v_0 = n^{-1/5} \in \mathcal{I}$.

We first consider the case where $a < n^{-1/3}$. If $v_1 = M_1 n^{-1/3+\eta} \in \mathcal{I}$, where $\eta > 0$ and $M_1 > \sqrt{10C_0}(A+1)(C_1+1)$, we have $\varDelta \le C_1 \sqrt{v_1}$. Choosing $v_2 = M_1 n^{-1/3+\eta/4}$, we then have

$$\begin{aligned}
|\delta_n| &\le C_0 n^{-1} v_2^{-3} \left(\Delta + \frac{v_2}{\sqrt{a} + \sqrt{v_2}} \right)^2 \\
&\le C_0 (C_1+1)^2 n^{-1} v_2^{-3} \left(\frac{v_1}{\sqrt{a} + \sqrt{v_1}} \right)^2 \\
&\le C_0 (C_1+1)^2 n^{-1} v_2^{-3} v_1 \\
&\le C_0 (C_1+1)^2 M_1^{-2} v_2 < v_2 / [10(A+1)^2].
\end{aligned}$$

This proves that $v_2 \in \mathcal{I}$. Starting from $\eta = \frac{1}{3} - \frac{1}{5} = 2/15$, recursively using the result above, we know that, for any m,

$$M_1 n^{-1/3+2/[15\times 4^m]} \in \mathcal{I}.$$

Making $m \to \infty$, we have shown that $M_1 n^{-1/3} \in \mathcal{I}$. Consequently,

$$\Delta \le O(n^{-1/6}).$$

Now, let us consider the case $a > n^{-1/3}$. If $v_3 = M_2 n^{-1/2+\eta} a^{-1/2} \in \mathcal{I}$, where $\eta > 0$ and $M_2 > \sqrt{10 C_0}(A+1)(C_1+1)$, then we have $\Delta \le C_1 v_3 / \sqrt{a}$. Choosing $v_4 = M_2 n^{-1/2+\eta/2} a^{-1/2}$, we then have

$$\begin{aligned}
|\delta_n| &\le C_0 n^{-1} v_4^{-3} \left(\Delta + \frac{v_4}{\sqrt{a} + \sqrt{v_4}} \right)^2 \\
&\le C_0 (C_1+1)^2 n^{-1} v_4^{-3} v_3^2 / a \\
&\le C_0 (C_1+1)^2 M_2^{-2} v_4 < v_4 / [10(A+1)^2].
\end{aligned}$$

This proves that $v_4 \in \mathcal{I}$. Starting from $\eta = \frac{1}{2} - \frac{1}{5} = 3/10$, recursively using the result above, we obtain, for any m,

$$M_2 n^{-1/2+3/[10\times 2^m]} a^{-1/2} \in \mathcal{I}.$$

Making $m \to \infty$, we have shown that $M_1 n^{-1/2} a^{-1/2} \in \mathcal{I}$. Consequently,

$$\Delta \le O(n^{-1/2} a^{-1}).$$

The proof of the theorem is complete.

8.4 Some Elementary Calculus

8.4.1 Increment of M-P Density

To apply Lemma B.19 to the truncation and centralization of the entries of $\mathbf{X}_n$, we need to estimate the incremental function g given in Lemma B.19. We have the following lemma.

Lemma 8.14. *For the M-P law with index $y \le 1$, the function g in Lemma B.19 can be taken as $g(v) = 2v/(y(\sqrt{a} + \sqrt{v}))$.*

Proof. Let $v > 0$ be a small number and

$$\Phi(x) = \int_x^{x+v} \frac{1}{2\pi t y} \sqrt{(b-t)(t-a)} I_{[a,b]}(t) dt.$$

To find the maximum value of $\Phi(x)$ for fixed v, we may assume that $a \le x \le b - v$ because $\Phi(x)$ is increasing for $x < a$ and decreasing for $x > b - v$. In this case,

$$\begin{aligned} \Phi(x) &\le \int_x^{x+v} \frac{1}{\pi y \sqrt{t}} dt \\ &= \frac{2}{\pi y}(\sqrt{x+v} - \sqrt{x}) \\ &= \frac{2v}{\pi y(\sqrt{x+v} + \sqrt{x})} \\ &\le \frac{2v}{\pi y(\sqrt{a} + \sqrt{v})}. \end{aligned}$$

To apply Corollary B.15, we need the following estimate.

Lemma 8.15. *For $v > n^{-1/2}$, we have*

$$\sup_x \int_{|u|<v} |F_y(x+u) - F_y(x)| du < \frac{11\sqrt{2(1+y)}}{3\pi y} v^2/v_y,$$

where F_y is the M-P *law with index $y \le 1$.*

Proof. Set

$$\Phi(\lambda) = \int_0^v |F_y(x+u) - F_y(x)| du,$$

where $\lambda = x - a$. It is obvious that

$$\sup_x \int_{|u|<v} |F_y(x+u) - F_y(x)|du \le 2 \sup_\lambda \varPhi(\lambda).$$

We now begin to estimate the maximum of $\varPhi(\lambda)$. To this end, we need only consider the case $\lambda \ge 0$. Then,

$$\begin{aligned}\varPhi(\lambda) &= \int_0^v \int_x^{x+u} \frac{1}{2\pi xy} \sqrt{(b-t)(t-a)} I_{[a,b]}(t) dt du \\ &= \int_{a+\lambda}^{a+\lambda+v} \frac{a+\lambda+v-t}{2\pi xy} \sqrt{(b-t)(t-a)} I_{[a,b]}(t) dt \\ &= \int_\lambda^{\lambda+v} \frac{\lambda+v-t}{2\pi y(t+a)} \sqrt{t(4\sqrt{y}-t)} I_{[0,4\sqrt{y}]}(t) dt.\end{aligned}$$

Write $\phi(t) = (a+t)^{-1}\sqrt{t(4\sqrt{y}-t)}$. Then

$$2\frac{d}{dt}\log\phi(t) = \frac{1}{t} - \frac{1}{4\sqrt{y}-t} - \frac{1}{t+a} = \frac{2(2\sqrt{y}a - (1+y)t)}{t(4\sqrt{y}-t)(t+a)}.$$

This shows that $\phi(t)$ is increasing in the interval $(0,\rho)$ and decreasing in $(\rho, 4\sqrt{y})$, where $\rho = 2a\sqrt{y}/(1+y)$. Since

$$\frac{d}{d\lambda}\varPhi(\lambda) = \frac{1}{2\pi y}\int_\lambda^{\lambda+v} [\phi(t) - \phi(\lambda)]dt,$$

it follows that $\varPhi(\lambda)$ is decreasing when $\lambda > \rho$ and increasing when $\lambda < \rho - v$. Therefore, the maximum of $\varPhi(\lambda)$ can only be reached when $\lambda \in [0 \vee (\rho - v), \rho]$. Suppose λ is in this interval. Then

$$\begin{aligned}\varPhi(\lambda) &\le \frac{2y^{1/4}}{2\pi y} \int_\lambda^{\lambda+v} \frac{\lambda+v-u}{u+a}\sqrt{u}du \\ &= 2(\pi y^{3/4})^{-1}\Bigg\{(\lambda+v+a)\Bigg[(\sqrt{\lambda+v}-\sqrt{\lambda}) \\ &\qquad -\sqrt{a}\left(\arctan\sqrt{\frac{\lambda+v}{a}} - \arctan\sqrt{\frac{\lambda}{a}}\right)\Bigg] \\ &\qquad -\frac{1}{3}\left[(\lambda+v)^{3/2} - \lambda^{3/2}\right]\Bigg\}.\end{aligned}$$

Since

$$\sqrt{a}\left(\arctan\sqrt{\frac{\lambda+v}{a}} - \arctan\sqrt{\frac{\lambda}{a}}\right) \ge \frac{a}{\lambda+v+a}\left(\sqrt{\lambda+v}-\sqrt{\lambda}\right),$$

we get, by setting $\lambda^* = \sqrt{\lambda+v} - \sqrt{\lambda}$,

$$\begin{aligned}&\Phi(\lambda)\\ &\leq \frac{2}{\pi y^{3/4}}\left\{(a+\lambda+v)\left(\lambda^* - \frac{a}{a+\lambda+v}\lambda^*\right) - \lambda^*\left(\lambda + \sqrt{\lambda}\lambda^* + \frac{1}{3}\lambda^{*2}\right)\right\}\\ &= \frac{2}{\pi y^{3/4}}\left[\sqrt{\lambda}\lambda^{*2} + \frac{2}{3}\lambda^{*3}\right].\end{aligned}$$

Let $c^2 = \frac{1+y}{2\sqrt{y}}$. Since $\lambda + v \geq c^{-2}a$ and

$$(\sqrt{\lambda+v}+\sqrt{\lambda})^2 \geq \lambda + v + 2\sqrt{\lambda(\lambda+v)} \geq 2\sqrt{\lambda v} + 2\sqrt{\lambda c^{-2}a},$$

we have

$$\begin{aligned}\frac{\sqrt{\lambda}}{(\sqrt{\lambda+v}+\sqrt{\lambda})^2} &\leq \frac{c}{2\sqrt{a}+2c\sqrt{v}} \leq \frac{c}{2\sqrt{a}+2\sqrt{v}}\ ,\\ \frac{1}{(\sqrt{\lambda+v}+\sqrt{\lambda})^3} &\leq \frac{2c}{(\sqrt{a}+\sqrt{v})v}\ ,\end{aligned}$$

where the last inequality follows from

$$\begin{aligned}(\sqrt{\lambda+v}+\sqrt{\lambda})^3 &\geq \sqrt{\lambda+v}v\\ &\geq \frac{1}{2}\left[\sqrt{c^{-2}a}+\sqrt{v}v\right]\\ &\geq \frac{1}{2c}\left[\sqrt{v}+\sqrt{a}\right]v.\end{aligned}$$

Hence

$$\Phi(\lambda) \leq \frac{2}{\pi y^{3/4}} \cdot \frac{11c}{6(\sqrt{a}+\sqrt{v})}v^2 = \frac{11\sqrt{2(1+y)}}{6\pi y}\frac{1}{\sqrt{v}+(1-\sqrt{y})}v^2.$$

This completes the proof of the lemma.

8.4.2 Integral of Tail Probability

This lemma estimates the integral of the tail probability, which is needed when applying Theorem B.14.

Lemma 8.16. *Under condition (8.3.1) and the additional assumption*

$$|x_{ij}| < n^{1/4}, \tag{8.4.1}$$

we have

$$\int_B^\infty |\mathrm{E}F_p(x) - F_y(x)|\mathrm{d}x = o(n^{-t}), \tag{8.4.2}$$

where $B = b+1$, $b = b(y)$ is defined below (8.3.2) and the constant $t > 0$ is fixed but can be arbitrarily large.

Proof. By Theorem 5.9, we have

$$\mathrm{P}(\lambda_{\max}(\mathbf{S}_n) \geq B + x) \leq Cn^{-t-1}(B + x - \varepsilon)^{-2}. \tag{8.4.3}$$

Note that

$$1 - F_p(x) \leq I_{[\lambda_{\max}(\mathbf{S}_n) \geq x]} \quad \text{for} \quad x \geq 0. \tag{8.4.4}$$

We have

$$\begin{aligned}
&\int_B^\infty |\mathrm{E}F_p(x) - F(x)|\mathrm{d}x \leq \int_B^\infty P(\lambda_p \geq x)dx \\
&\leq \int_B^\infty Cn^{-t-1}(B + x - \varepsilon)^{-2}dx \\
&= o(n^{-t}).
\end{aligned} \tag{8.4.5}$$

The proof is complete.

8.4.3 Bounds of Stieltjes Transforms of the M-P Law

We frequently need the bounds of the Stieltjes transform of the M-P law in the proof of Theorem 8.10.

Lemma 8.17. *For the Stieltjes transform of the M-P law, we have*

$$|s_y(z)| \leq \frac{\sqrt{2}}{\sqrt{y}v_y},$$

where $v_y = \sqrt{a} + \sqrt{v} = 1 - \sqrt{y} + \sqrt{v}$.

Proof. Recall that

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

which is a root of the quadratic equation

$$yzs^2 + (y + z - 1)s + 1 = 0.$$

Write the other root of the equation as

$$s_y^*(z) = -\frac{y + z - 1 + \sqrt{(1 + y - z)^2 - 4y}}{2yz}.$$

We claim that, for any $z \in \mathbb{C}^+$,

$$|s_y(z)| < |s_y^*(z)|. \tag{8.4.6}$$

Set

$$\alpha + i\beta = \sqrt{(1+y-z)^2 - 4y}, \quad \beta \geq 0.$$

Then,

$$\alpha\beta = v(u - y - 1), \tag{8.4.7}$$

$$\beta^2 - \alpha^2 = (b-u)(u-a) + v^2. \tag{8.4.8}$$

If $\beta = 0$, then $u = 1 + y$, which violates equation (8.4.8). Thus, $\beta > 0$ for all cases. Now,

$$\begin{aligned} |s_y(z)| < |s_y^*(z)| &\Leftrightarrow |y + z - 1 - \alpha - i\beta|^2 < |y + z - 1 + \alpha + i\beta|^2 \\ &\Leftrightarrow \alpha(y - 1 + u) + \beta v > 0. \end{aligned}$$

This is trivially true if

$$\alpha(y - 1 + u) \geq 0 \Leftrightarrow (y - 1 + u)(u - y - 1) \geq 0 \Leftrightarrow |1 - u| > y.$$

If (8.4.6) is not always true, then there is a $z \in \mathbb{C}^+$ such that

$$|s_y(z)| = |s_y^*(z)| \Rightarrow \alpha(y - 1 + u) + \beta v = 0,$$

which together with (8.4.7) implies that

$$\begin{aligned} \beta^2 &= v^{-1}\beta\alpha(1 - y - u) = (1 - y - u)(u - y - 1) \\ &= y^2 - (1 - u)^2 \leq (b - u)(u - a). \end{aligned}$$

The above leads to a contradiction to (8.4.8). Assertion (8.4.6) is proved.

With the observation that $s_y(z)s_y^*(z) = 1/yz$, we obtain

$$|s_y(z)| < 1/\sqrt{y|z|}. \tag{8.4.9}$$

Next, we shall get a better estimate for $|s_y(z)|$ when $|z|$ is small. When $u < a - v$, both the real and imaginary parts are positive and increasing. Thus, the maximum of $|s_y(z)|$ can only be reached when $u > a - v$. If $a < 2v$, then

$$|s_y(z)| \leq 1/\sqrt{yv} \leq \frac{\sqrt{2}}{\sqrt{y}(\sqrt{a} + \sqrt{v})}.$$

If $a > 2v$, then

$$\sqrt{|z|} \geq \sqrt[4]{a^2/4 + v^2} \geq \frac{1}{\sqrt{2}}(\sqrt{a} + \sqrt{v}).$$

We also have the same estimate. Thus, the proof of the lemma is complete.

8.4.4 Bounds for $\tilde{b}_n$

Similar to (8.3.7), we may expand $s_p(z)$ as

$$s_p(z) = \frac{1}{p}\sum_{k=1}^{p} \frac{1}{s_{kk} - z - n^{-2}\boldsymbol{\alpha}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}\boldsymbol{\alpha}_k}$$
$$= -\tilde{b}_n + \tilde{\delta}, \tag{8.4.10}$$

where

$$\tilde{\varepsilon}_k = (s_{kk} - 1) + y + yzs_p(z) - \frac{1}{n^2}\boldsymbol{\alpha}_k^*(\mathbf{S}_{nk} - zI_{p-1})^{-1}\boldsymbol{\alpha}_k,$$
$$\tilde{\delta} = \delta_p \; = \; -\frac{1}{p}\sum_{k=1}^{p} \tilde{b}_n \beta_k \varepsilon_k,$$
$$\tilde{b}_n = \tilde{b}_n(z) = \frac{1}{z + y - 1 + yzs_p(z)},$$
$$\beta_k = \beta_k(z) = \frac{1}{z + y - 1 + yzs_p(z) - \hat{\varepsilon}_k}. \tag{8.4.11}$$

Also, similar to the proof of (8.3.9), one may prove that, for all $z \in \mathbb{C}^+$,

$$s_p(z) = -\frac{z + y - 1 - yz\tilde{\delta} - \sqrt{(z + y - 1 + ys\tilde{\delta})^2 - 4yz}}{2yz}. \tag{8.4.12}$$

We will prove the following lemma.

Lemma 8.18. *For all* $z \in \mathbb{C}^+$,

$$|\tilde{b}_n| < \frac{1}{\sqrt{y|z|}}.$$

Similarly, we have

$$|b_n| < \frac{1}{\sqrt{y|z|}}.$$

Proof. By (8.4.10) and (8.4.12),

$$-\tilde{b}_n = s_p(z) - \tilde{\delta} = -\frac{z + y - 1 + yz\tilde{\delta} - \sqrt{(z + y - 1 + yz\tilde{\delta})^2 - 4yz}}{2yz}.$$

By the convention made to square roots of complex numbers (see (2.3.2)), we have

$$\sqrt{a} = \begin{cases} \sqrt{a/b}\sqrt{b}, & \text{if } \arg(a) > \arg(b), \\ -\sqrt{a/b}\sqrt{b}, & \text{if } \arg(a) < \arg(b), \end{cases}$$

where the angle $\arg(z)$ of the complex number z is defined to be the one in the range $[0, 2\pi)$. Therefore, we have

$$-\sqrt{yz}\tilde{b}_n = \begin{cases} s_{\rm semi}((z+y-1+yz\tilde{\delta})/\sqrt{yz}), & \text{if } z \in D_1, \\ -s_{\rm semi}(-(z+y-1+yz\tilde{\delta})/\sqrt{yz}), & \text{if } z \in D_2, \\ s_{\rm semi}^{-1}((z+y-1+yz\tilde{\delta})/\sqrt{yz}), & \text{if } z \in D_3, \\ -s_{\rm semi}^{-1}(-(z+y-1+yz\tilde{\delta})/\sqrt{yz}), & \text{if } z \in D_4, \end{cases}$$

where $s_{\rm semi}(\cdot)$ is the Stieltjes transform of the semicircular law defined in Section 2.3 (using different notation to avoid possible confusion with that of the M-P law) and

$$\begin{aligned} D_1 = \{z \in \mathbb{C}^+, \Im((z+y-1+yz\tilde{\delta})/\sqrt{yz}) > 0 \\ \text{and } \arg[(z+y-1+yz\tilde{\delta})^2 - 4yz] > \arg(yz)\}, \\ D_2 = \{z \in \mathbb{C}^+, \Im((z+y-1+yz\tilde{\delta})/\sqrt{yz}) < 0 \\ \text{and } \arg[(z+y-1+yz\tilde{\delta})^2 - 4yz] < \arg(yz)\}, \\ D_3 = \{z \in \mathbb{C}^+, \Im((z+y-1+yz\tilde{\delta})/\sqrt{yz}) > 0 \\ \text{and } \arg[(z+y-1+yz\tilde{\delta})^2 - 4yz] < \arg(yz)\}, \\ D_4 = \{z \in \mathbb{C}^+, \Im((z+y-1+\tilde{y}z\delta)/\sqrt{yz}) < 0 \\ \text{and } \arg[(z+y-1+yz\tilde{\delta})^2 - 4yz] > \arg(yz)\}. \end{aligned}$$

By (8.1.11), if $z \in D_1$ or D_2, then the conclusion of Lemma 8.18 holds. Then, we shall complete the proof of the first conclusion by showing that $D_3 = D_4 = \emptyset$.

We claim that $D_1 \neq \emptyset$. As $v = \Im(z)$ is large enough, we have the following estimates:

(i) $\Im\left(\frac{y-1}{\sqrt{yz}}\right) > 0$,
(ii) $\Im\left(\sqrt{z/y}\right) \sim \sqrt{v/y}$,
(iii) $|\sqrt{yz}\tilde{\delta}| = o(\sqrt{v})$ as $v \to \infty$.

These three estimates show that, for any fixed u and all large v,

$$\Im((z+y-1+yz\tilde{\delta})/\sqrt{yz}) > 0. \tag{8.4.13}$$

Also, by (8.4.10),

$$z+y-1+yz\tilde{\delta} = z+y-1+yzs_p(z) + \frac{1}{z+y-1+yzs_p(z)}.$$

Note that

$$z + y - 1 + yzs_p(z) = z - 1 + y\int_0^\infty \frac{x}{x-z}dF_p(x),$$

from which it is easy to deduce that, as $v \to \infty$,

$$\Im(z + y - 1 + yzs_p(z)) \geq v \to \infty,$$

and $\Re(z + y - 1 + yzs_p(z)) = u - 1 + o(1)$. Therefore, for any fixed $u < 1$ and all large v,

$$\begin{aligned} &\Im((z + y - 1 + yz\tilde{\delta})^2 - 4yz) \\ &= \Im[(z + y - 1 + yzs_p(z))^2 - 4yz] + o(1) \\ = &(u + y - 1 + \Re(yzs_p(z)))(v + \Im(yzs_p(z)) - 4yv) < 0. \end{aligned}$$

In this case, $\arg((z + y - 1 + yz\tilde{\delta})^2 - 4yz) > \pi > \arg(yz)$. Hence, D_1 is not empty.

Next, we claim that $D_3 = D_4 = \emptyset$. If not, then we may choose $z_1 \in D_1$ and $z_2 \in D_3 \cup D_4$. Draw a continuous curve from z_1 to z_2. Let z_0 be a point on the intersection of the curve and the boundary of $D_3 \cup D_4$. No matter whether $z_0 \in \partial D_1 \cap \partial D_3$ or the other three cases, by (8.1.11) we always have

$$|s_{semi}((z_0 + y - 1 + yz_0\tilde{\delta})/\sqrt{yz_0})| = 1,$$

which implies that

$$(z_0 + y - 1 + yz_0\tilde{\delta})/\sqrt{yz_0} = \pm 2$$

and

$$s_{semi}((z_0 + y - 1 + yz_0\tilde{\delta})/\sqrt{yz_0}) = \pm 1.$$

Recall that

$$\begin{aligned} \sqrt{yz_0}\tilde{b}_n(z_0) &= s_{semi}((z_0 + y - 1 + yz_0\tilde{\delta}(z_0))/\sqrt{yz_0}) \\ &= s_{semi}^{-1}((z_0 + y - 1 + yz_0\tilde{\delta}(z_0))/\sqrt{yz_0}) = \pm 1. \end{aligned}$$

These show that z_0 is a root of the equation $yz_0\tilde{\mathbf{b}}_n^2 = 1$, or equivalently

$$yz_0 = (y + z_0 - 1 + yz_0s_p(z_0))^2. \tag{8.4.14}$$

Note that $s_p(z)$ is analytic on $\mathbb{C}^+$. Because we can draw infinitely many disjoint curves from z_1 to z_2, we can find infinitely many distinct z_0's, roots of (8.4.14). By the unique extension theorem of analytic functions, we conclude that (8.4.14) holds for all $z \in \mathbb{C}^+$. This contradicts the fact that D_1 is not empty.

We conclude that $\mathbb{C}^+ \subset \overline{D}_1 \cup \overline{D}_2$. Hence, the first conclusion of the lemma follows from (8.1.11) and the continuity.

The second conclusion can be proved along the same lines, and thus the proof of the lemma is complete.

8.4.5 Integrals of Squared Absolute Values of Stieltjes Transforms

Applying Lemma B.20, we obtain

$$\int |s_y(z)|^2 du \leq \frac{2\pi}{\sqrt{y}(1-y)} \tag{8.4.15}$$

since the density function has an upper bound $1/(\pi\sqrt{y}(1-y))$.

It should be noted that this bound tends to infinity when y tends to zero or one. This is reasonable for the case where $y \to 0$ because the distribution tends to be degenerate. For the case $y \to 1$ or even $y = 1$, the distribution is still continuous, although the density for $y = 1$ is unbounded. One may want to have a finite bound (probably depending on v). We have the following inequality.

Lemma 8.19. *For any $y \leq 1$, we have*

$$\int |s_y(z)|^2 du \leq \frac{2\pi(1+\sqrt{y})}{yv_y}. \tag{8.4.16}$$

Proof. Using the notation and going through the same lines of the proof of Lemma B.20, we find that

$$\begin{aligned}
I =: & \int |s_y(z)|^2 \mathrm{d}u \\
= & \ 4\pi v \int_a^b \int_a^b \frac{1}{(u-x)^2 + 4v^2} \phi(x)\phi(u) \mathrm{d}x\mathrm{d}u \\
= & \ 8\pi v \int_a^b \int_x^b \frac{1}{(u-x)^2 + 4v^2} \phi(x)\phi(u) \mathrm{d}x\mathrm{d}u \ \ \text{(by symmetry)} \\
\leq & \ 4vy^{-1}\sqrt{b} \int_0^\infty \int_a^b \phi(x) \frac{1}{\sqrt{w+a}(w^2+4v^2)} \mathrm{d}w\mathrm{d}x \\
= & \ 2y^{-1}\sqrt{b} \int_0^\infty \frac{1}{\sqrt{2vw+a}(w^2+1)} \mathrm{d}w.
\end{aligned}$$

If $a \geq v$, we have

$$I \leq 2y^{-1}\sqrt{b} \int_0^\infty \frac{1}{\sqrt{a}(w^2+1)} dw$$

$$= \pi y^{-1}\sqrt{b/a} \le \frac{2\pi\sqrt{2b}}{y(\sqrt{a}+\sqrt{v})}.$$

If $a < v$, then

$$I \le 2y^{-1}\sqrt{b}\int_0^\infty \frac{1}{\sqrt{2wv}(w^2+1)}dw$$
$$= \pi y^{-1}\sqrt{b/v} \le \frac{2\pi\sqrt{b}}{y(\sqrt{a}+\sqrt{v})}.$$

Here we used the fact that

$$\int_0^\infty \frac{1}{\sqrt{w}(w^2+1)}\mathrm{d}w = \frac{\pi}{\sqrt{2}},$$

which may be computed by using the residue theorem and the equality

$$\int_0^\infty \frac{1}{\sqrt{w}(w^2+1)}\mathrm{d}w = \frac{1}{1-i}\int_{-\infty}^\infty \frac{1}{\sqrt{z}(z^2+1)}\mathrm{d}z.$$

This completes the proof of Lemma 8.19.

8.4.6 Higher Central Moments of Stieltjes Transforms

Lemma 8.20. *If* $|z| < A$, $v > n^{-1/2}$, *and* $\ell \ge 1$, *then*

$$\mathrm{E}|s_p(z) - \mathrm{E}s_p(z)|^{2\ell} \le \frac{C}{n^{2\ell}v^{-4\ell}y^{2\ell}}(\Delta + v/v_y)^\ell, \tag{8.4.17}$$

where A *is a positive constant and* $v_y = 1 - \sqrt{y} + \sqrt{v}$, *which is defined in Lemma 8.17.*

Proof. Write $\mathrm{E}_k(\cdot)$ as the conditional expectation given $\{x_{ij}; i \le k,\ j \le n\}$, with the convention that E_0 is the unconditional expectation. Then

$$s_p(z) - \mathrm{E}s_p(z) = \frac{1}{p}\sum_{k=1}^p \gamma_k,$$

where

$$\gamma_k = \mathrm{E}_k\mathrm{tr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1} - \mathrm{E}_{k-1}\mathrm{tr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1}$$
$$= \mathrm{E}_k\sigma_k - \mathrm{E}_{k-1}\sigma_k$$

and

$$\sigma_k = \beta_k(1 + n^{-2}\boldsymbol{\alpha}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-2}\boldsymbol{\alpha}_k).$$

By Lemma 2.13, we have

$$\begin{aligned} &\mathrm{E}|s_p(z) - \mathrm{E}s_p(z)|^{2\ell} \\ &\leq Cn^{-2\ell}\left\{\mathrm{E}\left(\sum_{k=1}^{p}\mathrm{E}_{k-1}|\gamma_k^2|\right)^{\ell} + \sum_{k=1}^{p}\mathrm{E}|\gamma_k|^{2\ell}\right\}. \end{aligned} \tag{8.4.18}$$

Define

$$\begin{aligned} b_{nk} &= \frac{1}{z+y-1+\frac{z}{n}\mathrm{tr}(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}}, \\ \tilde{\varepsilon}_k &= s_{kk} - 1 - \frac{1}{n^2}\boldsymbol{\alpha}_k^*(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}\boldsymbol{\alpha}_k + \frac{1}{n}\mathrm{tr}(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}\mathbf{S}_{nk}. \end{aligned}$$

Note that

$$|\tilde{b}_n^{-1} - b_{nk}^{-1}| = \left|\frac{z}{n}[\mathrm{tr}(\mathbf{S}_n - z\mathbf{I}_p)^{-1} - \mathrm{tr}(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}]\right| < \frac{|z|}{nv}.$$

By Lemma 8.18, when $v > 1/\sqrt{n}$, we have

$$|b_{nk}| = |\tilde{b}_n||1 + b_{nk}(\tilde{b}_n^{-1} - b_{nk}^{-1})| \leq \left(1 + \frac{|z|}{nv^2}\right)|\tilde{b}_n| \leq C/\sqrt{y|z|}. \tag{8.4.19}$$

Note that $\sigma_k = -b_{nk} - b_{nk}\sigma_k\tilde{\varepsilon}_k$ and $|\sigma_k| \leq 1/v$, which follows from the observation that

$$|1 + n^{-2}\boldsymbol{\alpha}_k^*(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}\boldsymbol{\alpha}_k| \leq v^{-1}\Im(\beta_k^{-1}).$$

We can rewrite γ_k as

$$\begin{aligned} \gamma_k = &-n^{-2}\mathrm{E}_k b_{nk}\mathbf{x}_k'\mathbf{X}_{nk}^*(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-2}\mathbf{X}_{nk}\bar{\mathbf{x}}_k \\ &-\mathrm{tr}(\mathbf{X}_{nk}^*(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-2}\mathbf{X}_{nk}) - (\mathrm{E}_k - \mathrm{E}_{k-1})b_{nk}\tilde{\varepsilon}_k\sigma_k. \end{aligned}$$

In what follows, we shall repeatedly use the identity

$$(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}\mathbf{S}_{nk} = \mathbf{I}_{p-1} + z(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}$$

and the inequality

$$(\mathbf{I}+\mathbf{B})^{\ell}[(\mathbf{I}+\mathbf{B})^*]^{\ell} \leq C[\mathbf{I} + \mathbf{B}^{\ell}[\mathbf{B}^*]^{\ell}],$$

for any normal matrix $\mathbf{B}$; that is, it satisfies $\mathbf{B}\mathbf{B}^* = \mathbf{B}^*\mathbf{B}$, where $\mathbf{A} \leq \mathbf{B}$ means that $\mathbf{B} - \mathbf{A}$ is a nonnegative definite matrix.

Applying Lemma B.26 and (8.4.19), similar to (8.1.34), we obtain

$$\begin{aligned} &n^{-4}\mathrm{E}_{k-1}|b_{nk}\mathbf{x}_k'\mathbf{X}_{nk}^*(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-2}\mathbf{X}_{nk}\bar{\mathbf{x}}_k \\ &\quad-\mathrm{tr}(\mathbf{X}_{nk}^*(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-2}\mathbf{X}_{nk})|^2 \end{aligned}$$

$$
\begin{aligned}
&\le \frac{1}{n^2}\mathrm{E}_{k-1}|b_{nk}|^2\mathrm{tr}((\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-2}\mathbf{S}_{nk}(\mathbf{S}_{nk}-\bar{z}\mathbf{I}_{p-1})^{-2}\mathbf{S}_{nk})\\
&\le \frac{C}{nv^2y|z|}\mathrm{tr}((\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}\mathbf{S}_{nk}(\mathbf{S}_{nk}-\bar{z}\mathbf{I}_{p-1})^{-1}\mathbf{S}_{nk})\\
&\le \frac{C}{nv^2y|z|}\left[n+|z|^2\mathrm{tr}((\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}(\mathbf{S}_{nk}-\bar{z}\mathbf{I}_{p-1})^{-1})\right]\\
&\le Cv^{-3}n^{-1}\mathrm{E}_{k-1}(1+\Im(s_p(z)))
\end{aligned}
\tag{8.4.20}
$$

and

$$
\begin{aligned}
&\mathrm{E}|(\mathrm{E}_k-\mathrm{E}_{k-1})b_{nk}\tilde{\varepsilon}_k\sigma_k|^2\\
&\le v^{-2}\mathrm{E}_{k-1}|b_{nk}|^2|\tilde{\varepsilon}_k|^2\\
&\le 2v^{-2}\mathrm{E}_{k-1}|b_{nk}|^2\left[|s_{kk}-1|^2+n^{-4}|\mathbf{x}_k'\mathbf{X}_{nk}^*(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}\mathbf{X}_{nk}\bar{\mathbf{x}}_k\right.\\
&\quad\left.-\mathrm{tr}(\mathbf{X}_{nk}^*(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}\mathbf{X}_{nk})|^2\right]\\
&\le Cv^{-2}\mathrm{E}_{k-1}|b_{nk}|^2\Big[n^{-1}\\
&\quad+n^{-2}|\mathrm{tr}((\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}\mathbf{S}_{nk}(\mathbf{S}_{nk}-\bar{z}\mathbf{I}_{p-1})^{-1}\mathbf{S}_{nk})\Big]\\
&\le Cv^{-3}n^{-1}\mathrm{E}_{k-1}\left(1+\Im(s_p(z))\right).
\end{aligned}
\tag{8.4.21}
$$

By (8.4.20) and (8.4.21), for large n,

$$
\mathrm{E}\left(\sum_{k=1}^{p}\mathrm{E}_{k-1}|\gamma_k^2|\right)^{\ell}\le Cn^{-2\ell}v^{-3\ell}\mathrm{E}(1+\Im(s_p(z)))^{\ell}.
$$

Furthermore, we have

$$
\begin{aligned}
&n^{-4\ell}\mathrm{E}|b_{nk}\mathbf{x}_k'\mathbf{X}_{nk}^*(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-2}\mathbf{X}_{nk}\bar{\mathbf{x}}_k\\
&\quad-\mathrm{tr}(\mathbf{X}_{nk}^*(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-2}\mathbf{X}_{nk})|^{2\ell}\\
&\le \frac{C}{n^{2\ell}}\mathrm{E}|b_{nk}|^{2\ell}\Big[\nu_{4\ell}\mathrm{tr}\big((\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-2}\mathbf{S}_{nk}(\mathbf{S}_{nk}-\bar{z}\mathbf{I}_{p-1})^{-2}\mathbf{S}_{nk}\big)^{\ell}\\
&\quad+\big[\nu_4\mathrm{tr}\big((\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-2}\mathbf{S}_{nk}(\mathbf{S}_{nk}-\bar{z}\mathbf{I}_{p-1})^{-2}\mathbf{S}_{nk}\big)\big]^{\ell}\Big]\\
&\le \frac{C}{n^{2\ell}y^{\ell}|z|^{\ell}}\mathrm{E}\Big[\nu_{4\ell}v^{-2\ell}\mathrm{tr}\big(\mathbf{I}_{p-1}+|z|^{2\ell}(\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-\ell}(\mathbf{S}_{nk}-\bar{z}\mathbf{I}_{p-1})^{-\ell}\big)\\
&\quad+\nu_4^{\ell}v^{-2\ell}\Big[n+|z|^2\mathrm{tr}\big((\mathbf{S}_{nk}-z\mathbf{I}_{p-1})^{-1}(\mathbf{S}_{nk}-\bar{z}\mathbf{I}_{p-1})^{-1}\big)^{\ell}\Big]\Big]\\
&\le \frac{C}{n^{2\ell}y^{\ell}|z|^{\ell}}\mathrm{E}\Big[n^{\ell}v^{-2\ell}\big(1+|z|^{2\ell}v^{-2\ell+1}\Im s_p(z)\big)\quad(\text{using }\nu_{4\ell}\le Cn^{\ell-1})\\
&\quad+n^{\ell}v^{-2\ell}\left(1+|z|^2v^{-1}\Im s_p(z)\right)^{\ell}\Big]\\
&\le \frac{C}{n^{\ell}y^{\ell}v^{3\ell}}\mathrm{E}\left[1+v^{-\ell+1}\Im s_p(z)+(\Im s_p(z))^{\ell}\right]
\end{aligned}
$$

$$\leq \frac{C}{n^\ell y^\ell v^{3\ell}} \mathrm{E}\left(1 + v^{-\ell+1}\Im s_p(z)\right) \tag{8.4.22}$$

and

$$\begin{aligned}
&\mathrm{E}|(\mathrm{E}_k - \mathrm{E}_{k-1}) b_{nk}\tilde{\varepsilon}_k \sigma_k|^{2\ell} \\
&\leq v^{-2\ell}\mathrm{E}|b_{nk}|^{2\ell}|\tilde{\varepsilon}_k|^{2\ell} \\
&\leq \frac{C}{v^{2\ell}y^\ell|z|^\ell}\mathrm{E}\left[|s_{kk}-1|^{2\ell} + n^{-4\ell}|\mathbf{x}_k'\mathbf{X}_{nk}^*(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}\mathbf{X}_{nk}\bar{\mathbf{x}}_k\right. \\
&\quad \left. -\mathrm{tr}(\mathbf{X}_{nk}^*(\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}\mathbf{X}_{nk})|^{2\ell}\right] \\
&\leq \frac{C}{v^{2\ell}y^\ell|z|^\ell}\mathrm{E}\Big[n^{-2\ell}(n\nu_{4\ell} + (\nu_4 n)^\ell) \\
&\quad + n^{-2\ell}\Big(\nu_{4\ell}\mathrm{tr}((\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-\ell}(\mathbf{S}_{nk} - \bar{z}\mathbf{I}_{p-1})^{-\ell}\mathbf{S}_{nk}^{2\ell}) \\
&\quad + \left(\nu_4\mathrm{tr}((\mathbf{S}_{nk} - z\mathbf{I}_{p-1})^{-1}(\mathbf{S}_{nk} - \bar{z}\mathbf{I}_{p-1})^{-1}\mathbf{S}_{nk}^2)\right)^\ell\Big)\Big] \\
&\leq \frac{C}{n^\ell v^{2\ell}y^\ell|z|^\ell}\mathrm{E}\left[1 + |z|^{2\ell}\Im s_p(z)v^{-2\ell+1} + v^{-\ell}|z|^{2\ell}\left(\Im(s_p(z))\right)^\ell\right] \\
&\leq \frac{C}{n^\ell v^{3\ell}y^\ell}\left(1 + \Im s_p(z)v^{-\ell+1}\right).
\end{aligned} \tag{8.4.23}$$

By (8.4.22) and (8.4.23), for large n,

$$\sum_{k=1}^p \mathrm{E}|\gamma_k|^{2\ell} \leq \frac{C}{n^{3\ell+1}y^\ell v^{3\ell}}(1 + \mathrm{E}\Im(s_p(z))v^{-\ell+1}).$$

Consequently,

$$\begin{aligned}
&|s_p(z) - \mathrm{E}s_p(z)|^{2\ell} \\
&\leq \frac{C}{n^{2\ell}v^{3\ell}y^\ell}(1 + \mathrm{E}(\Im(s_p(z)))^\ell + n^{-\ell+1}v^{-\ell+1}\mathrm{E}\Im(s_p(z))).
\end{aligned} \tag{8.4.24}$$

When $\ell = 1$, we have

$$\begin{aligned}
|s_p(z) - \mathrm{E}s_p(z)|^2 &\leq \frac{C}{n^2v^3y}(1 + \mathrm{E}(\Im(s_p(z))) \\
&\leq \frac{C}{n^2v^3y}(1 + |\mathrm{E}s_p(z) - s_y(z)| + |s_y(z)|) \\
&\leq \frac{C}{n^2v^3y}(1 + \varDelta/v + 1/\sqrt{y}v_y).
\end{aligned}$$

This shows that the lemma holds for $\ell = 1$. For $\ell \in (2^{t-1}, 2^t]$, the lemma will be proved by induction for $t = 0, 1, 2, \cdots$. To this end, we shall extend the lemma to the case $\ell \in (\frac{1}{2}, 1)$. We have, by Lemma 2.12 and the result for the case of $\ell = 1$,

$$\begin{aligned}
|s_p(z) - \mathrm{E}s_p(z)|^{2\ell} &\leq \frac{C}{n^{2\ell}}\mathrm{E}\left(\sum_{k=1}^{p}|\gamma_k|^2\right)^{\ell} \\
&\leq \frac{C}{n^{2\ell}}\left(\sum_{k=1}^{p}\mathrm{E}|\gamma_k|^2\right)^{\ell} \\
&= C\left(\mathrm{E}|s_p(z) - \mathrm{E}s_p(z)|^2\right)^{\ell} \\
&\leq \frac{C}{n^{2\ell}v^{4\ell}y^{2\ell}}(\Delta + v/v_y)^{\ell}.
\end{aligned}$$

This shows that the lemma holds when $\ell \in (2^{-1}, 2^0]$. Assume that the lemma holds for $\ell \in (\frac{1}{2}, 2^{t-1}]$. Consider the case where $\ell \in (2^{t-1}, 2^t]$. By (8.4.24), we have

$$\begin{aligned}
&|s_p(z) - \mathrm{E}s_p(z)|^{2\ell} \\
&\leq \frac{C}{n^{2\ell}v^{3\ell}y^{\ell}}(1 + \mathrm{E}|s_p(z) - \mathrm{E}s_p(y)|^{\ell} + |\mathrm{E}s_p(z)|^{\ell} \\
&\qquad + n^{-\ell+1}v^{-\ell+1}|\mathrm{E}s_p(z)|) \\
&\leq \frac{C}{n^{2\ell}v^{3\ell}y^{\ell}}\left(1 + \frac{(\Delta + v/v_y)^{\ell/2}}{n^{\ell}v^{2\ell}y^{\ell}} + |\Delta/v + 1/\sqrt{y}v_y|^{\ell}\right. \\
&\qquad \left. + n^{-\ell+1}v^{-\ell+1}(\Delta/v + 1/\sqrt{y}v_y)\right) \\
&\leq \frac{C}{n^{2\ell}v^{4\ell}y^{2\ell}}(\Delta + v/v_y)^{\ell}.
\end{aligned}$$

This completes the proof of the lemma.

8.4.7 Integral of δ

Lemma 8.21. *If $|\delta_n| < v/[10(A+1)^2]$ for all $|z| < A$, then there is a constant M such that*

$$\Delta \leq Mv/v_y,$$

where A is defined in Corollary B.15 for the M-P law with index $y \leq 1$.

Proof. By (3.3.1) and (8.3.9), we have

$$\begin{aligned}
&|s_p(z) - s_y(z)| \\
&\leq \left|\frac{\delta_n}{2}\right|\left[1 + \frac{|2(z+y-1) - yz\delta_n|}{|\sqrt{(z+y-1)^2 - 4yz} + \sqrt{(z+y-1+yz\delta_n)^2 - 4yz}|}\right].
\end{aligned} \tag{8.4.25}$$

By the convention on the sign of square roots of complex numbers (see (2.3.2)), the signs of the real parts of

$$\sqrt{(z+y-1)^2-4yz} \text{ and } \sqrt{(z+y-1+yz\delta_n)^2-4yz}$$

are $\text{sign}(u-y-1)$ and

$$\begin{aligned}&\text{sign}(u+y-1+\Re(yz\delta_n))(v+\Im(yz\delta_n))-2yv\\ &=\text{sign}(v(u-y-1)+(u+y-1)\Im(yz\delta_n)+\Re(yz\delta_n)(v+\Im(yz\delta_n))).\end{aligned}$$

By the condition on δ_n, we have

$$|(u+y-1)\Im(yz\delta_n)+\Re(yz\delta_n)(v+\Im(yz\delta_n))|<v/5(A+1).$$

If $|u-y-1|\ge 1/5(A+1)$, then the real parts of $\sqrt{(z+y-1)^2-4yz}$ and $\sqrt{(z+y-1+yz\delta_n)^2-4yz}$ have the same sign. Since they both have positive imaginary parts, it follows that

$$|s_p(z)-s_y(z)|\le\frac{1}{2}|\delta|\left[1+\frac{2A+2}{\sqrt{|(u-y-1)^2-v^2-4y|}}\right]. \tag{8.4.26}$$

If $|u-y-1|<1/[5(A+1)]$, then, for all large n, we have

$$\begin{aligned}|\sqrt{(z-y-1)^2-4y}-2i\sqrt{y}|&=\frac{|z-y-1|^2}{|\sqrt{(z-y-1)^2-4y}+2i\sqrt{y}|}\\ &\le\frac{1}{50\sqrt{y}(A+1)^2}.\end{aligned}$$

Therefore, for $|u-y-1|<1/[5(A+1)]$, we have

$$\begin{aligned}|s_p(z)-s_y(z)|&\le\frac{1}{2}|\delta|\left[1+\frac{2A+2}{\Im(\sqrt{|(u-y-1)^2-v^2-4y|})}\right]\\ &\le\frac{1}{2}|\delta|\left[1+\frac{2A+2}{\sqrt{y}}\right].\end{aligned} \tag{8.4.27}$$

Combining (8.4.26) and (8.4.27), for $|Z|\le A$, we have

$$|s_n(z)-s_y(z)|\le\begin{cases}C_1|\delta|, & \text{if } |u-y-1|<1/[5(A+1)],\\ \frac{1}{2}|\delta|\left[1+\frac{2A+2}{\sqrt{|(u+y-1)^2-v^2-4yu|}}\right], & \text{otherwise,}\end{cases} \tag{8.4.28}$$

where $C_1=C_1(y)$ is a positive constant depending upon y, say, and here it may take $\frac{A+2}{\sqrt{y}}$.

By (8.4.28), one finally gets

$$
\begin{aligned}
&\int_{-A}^{A} |s_n(z) - s_y(z)|\mathrm{d}u \\
&= \int_{[|u-y-1|\geq 1/[5(A+1)],|u|\leq A]} |s_n(z) - s_y(z)|du \\
&\quad + \int_{[|u-y-1|<1/[5(A+1)],|u|\leq A]} |s_n(z) - s_y(z)|du \\
&\leq Cv \int_{-A}^{A} \frac{1}{\sqrt{|(u+y-1)^2 - v^2 - 4yu|}}\mathrm{d}u + C\int_{-A}^{A} |\delta|\mathrm{d}u \\
&\leq \eta v + Cv = C_0 v, \qquad (8.4.29)
\end{aligned}
$$

where C_0 is some constant. The proof of the lemma is complete.

8.5 Rates of Convergence in Probability and Almost Surely

In this section, we further extend the results of the last section to the convergence in probability and almost surely. We have the following theorems.

Theorem 8.22. *Under the assumptions in (8.3.1), we have*

$$
\|F_p - F_{y_p}\| = \begin{cases} O_p(n^{-1/6}), & \text{if } a < n^{-1/3}, \\ O_p(n^{-2/5}a^{-2/5}), & \text{if } n^{-1/3} \leq a < 1. \end{cases} \qquad (8.5.30)
$$

Proof. We prove the two theorems simultaneously. By the arguments in Subsection 8.3.2, we can assume that the additional condition (8.4.1) holds. Then, by Theorem B.14, we have

$$
\begin{aligned}
\mathrm{E}(\|F_p - F_{y_p}\|) &\leq \Bigg(\int_{-A}^{A} \mathrm{E}(|s_p(z) - s_y(z)|)du \\
&\quad + v^{-1}\sup_x \int_{|y|\leq 2av} |F_y(x+y) - F_y(x)|dy + \int_B^\infty \mathrm{E}(F_p(x))dx \Bigg) \\
&\leq C\Bigg(\int_{-A}^{A} \mathrm{E}(|s_p(z) - \mathrm{E}(s_p(z))|)du \\
&\quad + \int_{-A}^{A} |\mathrm{E}(s_p(z)) - s_y(z)|du + v/v_y + o(n^{-2}) \Bigg). \qquad (8.5.31)
\end{aligned}
$$

By Lemma 8.20,

$$
\begin{aligned}
\mathrm{E}(|s_p(z) - \mathrm{E}(s_p(z))|) &\leq [\mathrm{E}(|s_p(z) - \mathrm{E}(s_p(z))|^2)]^{-1/2} \\
&\leq Cn^{-1}v^{-2}(\varDelta + v/v_y)^{1/2}.
\end{aligned}
$$

When $a < n^{-1/3}$, choosing $v = M_1 n^{-1/3}$ and noting that $\Delta = O(n^{-1/6})$, we obtain

$$\begin{aligned} \mathrm{E}(|s_p(z) - \mathrm{E}(s_p(z))|) &\le Cn^{-1}v^{-2}(\Delta + v/v_y)^{1/2} \\ &\le Cn^{-5/12}. \end{aligned} \tag{8.5.32}$$

In the proof of Theorem 8.10, it has been proved that the second integral on the right-hand side of (8.5.31) is $O(n^{-1/3})$. Substituting these into (8.5.31), we obtain

$$\mathrm{E}\|F_p - F_y\| = (n^{-1/6}).$$

This proves the conclusion for the case $a < n^{-1/3}$ of Theorem 8.22.

Now, assume that $a > n^{-1/3}$. In this case, we have $\Delta = O(n^{-1/2}a^{-1})$. Choose $v = M_2 n^{-2/5} a^{1/10}$, similar to (8.5.31), and we have

$$\begin{aligned} \mathrm{E}(|s_p(z) - \mathrm{E}(s_p(z))|) &\le Cn^{-1}v^{-2}(\Delta + v/v_y)^{1/2} \\ &\le Cn^{-1}v^{-1.5}a^{-1/4} \\ &= Cn^{-2/5}a^{-2/5} = Cva^{-1/2}. \end{aligned} \tag{8.5.33}$$

Then, by (8.5.31), we have

$$\mathrm{E}\|F_p - F_y\| = O(n^{-2/5}a^{-2/5}).$$

This proves the second case of Theorem 8.22.

Theorem 8.23. *Under the assumptions in (8.3.1), we have, for any $\eta > 0$,*

$$\|F_p - F_{y_p}\| = \begin{cases} O_{\text{a.s.}}(n^{-1/6}), & \text{if } a < n^{-1/3}, \\ O_{\text{a.s.}}(n^{-2/5+\eta}a^{-2/5}), & \text{if } n^{-1/3} \le a < 1. \end{cases} \tag{8.5.34}$$

Proof. The proof of this theorem is almost the same as that of Theorem 8.22. We only need to show that, for the case of $a < n^{-1/3}$ with $v = M_1 n^{-1/3}$,

$$\int_{-A}^{A} \mathrm{E}(|s_p(z) - s_y(z)|)du = O_{\text{a.s.}}(n^{-1/6}), \tag{8.5.35}$$

and for the case $a > n^{-1/3}$ with $v = M_2 n^{-2/5} a^{1/10}$,

$$\int_{-A}^{A} \mathrm{E}(|s_p(z) - s_y(z)|)du = O_{\text{a.s.}}(n^{-2/5+\eta}a^{-2/5}). \tag{8.5.36}$$

By Lemma 8.20, we have

$$n^{2\ell/6}\mathrm{E}(|s_p(z) - s_y(z)|^{2\ell}) \le Cn^{-2\ell}v^{-4\ell}(\Delta + v/v_y)^{\ell}.$$

When $a \le n^{-1/3}$, $\Delta = O(n^{-1/6})$. Thus, with $v = M_1 n^{-1/3}$,

$$n^{2\ell/6}\mathrm{E}(|s_p(z) - s_y(z)|^{2\ell}) \le Cn^{-\ell/2}.$$

Then (8.5.35) follows by choosing $\ell \ge 3$.

When $a > n^{-1/3}$, then $\Delta = O(n^{-1/2}a^{-1})$. Consequently, by choosing $v = M_2 n^{-2/5} a^{1/10}$, we have

$$n^{2\ell(2/5-\eta)} a^{4\ell/5}\mathrm{E}(|s_p(z) - s_y(z)|^{2\ell}) \le Cn^{-2\ell\eta}.$$

Then, (8.5.36) follows by choosing $\ell > 1/2\eta$. This completes the proof of Theorem 8.23.

Chapter 9
CLT for Linear Spectral Statistics

9.1 Motivation and Strategy

As mentioned in the introduction, many important statistics in multivariate analysis can be written as functionals of the ESD of some random matrices. The strong consistency of the ESD with LSD is not enough for more efficient statistical inferences, such as the test of hypotheses, confidence regions, etc. In this chapter, we shall introduce some results on deeper properties of the convergence of the ESD of large dimensional random matrices.

Let F_n be the ESD of a random matrix that has an LSD F. We shall call

$$\hat{\theta} = \int f(x) dF_n(x) = \frac{1}{n} \sum_{k=1}^{n} f(\lambda_k)$$

a linear spectral statistic (LSS), associated with the given random matrix, which can be considered as an estimator of $\theta = \int f(x) dF(x)$. To test hypotheses about θ, it is necessary to know the limiting distribution of

$$G_n(f) = \alpha_n(\hat{\theta} - \theta) = \int f(x) dX_n(x),$$

where $X_n(x) = \alpha_n(F_n(x) - F(x))$ and $\alpha_n \to \infty$ is a suitably chosen normalizer such that $G_n(f)$ tends to a nondegenerate distribution.

Ideally, if for some choice of α_n, $X_n(x)$ tends to a limiting process $X(x)$ in the C space or D space equipped with the Skorohod metric, then the limiting distribution of all LSS can be derived. Unfortunately, there is evidence indicating that $X_n(x)$ cannot tend to a limiting process in any metric space.

An example is given in Diaconis and Evans [94], in which it is shown that if F_n is the empirical distribution function of the angles of eigenvalues of a Haar matrix, then for $0 \le \alpha < \beta < 2\pi$, the finite-dimensional distributions of

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_9,

$$\frac{\pi n}{\sqrt{\log n}}(F_n(\beta) - F_n(\alpha) - \mathrm{E}[F_n(\beta) - F_n(\alpha)])$$

converge weakly to $Z_{\alpha,\beta}$, jointly normal variables, standardized, with covariances

$$\mathrm{Cov}(Z_{\alpha,\beta}, Z_{\alpha',\beta'}) = \begin{cases} 0.5, & \text{if } \alpha = \alpha' \text{ and } \beta \neq \beta', \\ 0.5, & \text{if } \alpha \neq \alpha' \text{ and } \beta = \beta', \\ -0.5, & \text{if } \alpha = \beta' \text{ or } \beta = \alpha', \\ 0, & \text{otherwise.} \end{cases}$$

This covariance structure cannot arise from a probability space on which $Z_{0,x}$ is defined as a stochastic process with measurable paths in $D[a,b]$ for any $0 < a < b < 2\pi$. Indeed, if so, then with probability 1, for x decreasing to a, $Z_{0,x} - Z_{0,a}$ would converge to 0, which implies its variance would approach 0. But its variance remains at 1. Furthermore, this result also shows that with any choice of α_n, $X_n(x)$ cannot tend to a nontrivial process in any metric space.

Therefore, we have to withdraw our attempts at looking for the limiting process of $X_n(x)$. Instead, we shall consider the convergence of $G_n(f)$ with $\alpha_n = n$. The earliest work dates back to Jonsson [169], in which he proved the CLT for the centralized sum of the r-th power of eigenvalues of a normalized Wishart matrix. Similar work for the Wigner matrix was obtained in Sinai and Soshnikov [269]. Later, Johansson [165] proved the CLT of linear spectral statistics of the Wigner matrix under density assumptions.

Because X_n tending to a weak limit implies the convergence of $G_n(f)$ for all continuous and bounded f, Diaconis and Evans' example shows that the convergence of $G_n(f)$ cannot be true for all f, at least for indicator functions. Thus, in this chapter, we shall confine ourselves to the convergence of $G_n(f)$ to a normal variable when f is analytic in a region containing the support of F for Wigner matrices and sample covariance matrices.

Our strategy will be as follows: Choose a contour $\mathcal{C}$ that encloses the support of F_n and F. Then, by the Cauchy integral formula, we have

$$f(x) = \frac{1}{2\pi i} \oint_{\mathcal{C}} \frac{f(z)}{z - x} dz. \tag{9.1.1}$$

By this formula, we can rewrite $G_n(f)$ as

$$G_n(f) = -\frac{1}{2\pi i} \oint_{\mathcal{C}} f(z)[n(s_n(z) - s(z)]dz, \tag{9.1.2}$$

where s_n and s are Stieltjes transforms of F_n and F, respectively. So, the problem of finding the limit distribution of $G_n(f)$ reduces to finding the limiting process of $M_n(z) = n(s_n(z) - s(z))$.

Before concluding this section, we present a lemma on estimation of moments of quadratic forms that is useful for the proofs of the CLT of LSS of both Wigner matrices and sample covariance matrices.

Lemma 9.1. *Suppose that* x_i, $i = 1, \cdots, n$, *are independent, with* $\mathrm{E}x_i = 0$, $\mathrm{E}|x_i|^2 = 1$, $\sup E|x_i|^4 = \nu < \infty$, *and* $|x_i| \le \eta\sqrt{n}$ *with* $\eta > 0$. *Assume that* $\mathbf{A}$ *is a complex matrix. Then, for any given* $2 \le p \le b\log(n\nu^{-1}\eta^4)$ *and* $b > 1$, *we have*

$$\mathrm{E}|\boldsymbol{\alpha}^*\mathbf{A}\boldsymbol{\alpha} - \mathrm{tr}(\mathbf{A})|^p \le \nu n^p (n\eta^4)^{-1}(40b^2\|\mathbf{A}\|\eta^2)^p,$$

where $\boldsymbol{\alpha} = (x_1, \cdots, x_n)^T$.

Proof. In the proof, we shall use the trivial inequality

$$a^t t^b \le \begin{cases} a^d c^b, & \text{if } d \le -b/\log a, \\ a^d d^b, & \text{if } -b/\log a < d \le t, \end{cases} \tag{9.1.3}$$

where $0 < a < 1$, b, d, c, t are positive, and $b \le -c\log a$.

Now, let us begin the proof of the lemma. Without loss of generality, we may assume that $p = 2s$ is an even integer. Write $\mathbf{A} = (a_{ij})$. We first consider

$$S_1 = \sum_{i=1} a_{ii}(|x_i|^2 - 1) := \sum_{i=1}^{n} a_{ii} M_i.$$

By noting $|a_{ii}| \le \|\mathbf{A}\|$ and $p \le 2b\log(n\nu^{-1}\eta^4)$, we apply (9.1.3) to get

$$\begin{aligned}
\mathrm{E}\,|S_1|^p &\le \sum_{\ell=1}^{s} \sum_{1\le j_1<\cdots<j_\ell\le n} \sum_{\substack{i_1+\cdots+i_\ell=p \\ i_1,\cdots,i_\ell\ge 2}} p! \prod_{t=1}^{\ell} \frac{|a_{j_t j_t}^{i_t}|\mathrm{E}|M_{j_t}^{i_t}|}{(i_t)!} \\
&\le \|\mathbf{A}\|^p \sum_{\ell=1}^{s} n^\ell \nu^\ell \sum_{\substack{i_1+\cdots+i_\ell=p \\ i_1,\cdots,i_\ell\ge 2}} (n\eta^2)^{p-2\ell} \frac{p!}{(i_1)!\cdots(i_\ell)!} \\
&\le (n\|\mathbf{A}\|\eta^2)^p \sum_{\ell=1}^{s} \left(\nu^{-1}n\eta^4\right)^{-\ell} \ell^p \\
&\le \begin{cases} \nu n^p s(\|\mathbf{A}\|\eta^2)^p (n\eta^4)^{-1}(2b)^p, & \text{if } p/\log(n\nu^{-1}\eta^4) \ge 1, \\ \nu n^p s(\|\mathbf{A}\|\eta^2)^p (n\eta^4)^{-1}, & \text{if } p/\log(n\nu^{-1}\eta^4) < 1, \end{cases} \\
&\le \nu n^p (2b\|\mathbf{A}\|s^{1/p}\eta^2)^p (n\eta^4)^{-1}.
\end{aligned}$$

Next, let us consider

$$S_2 = \sum_{1\le i\ne j\le n} a_{ij} x_i \bar{x}_j.$$

Then, we have

$$\mathrm{E}\,|S_2|^p = \sum a_{i_1 j_1}\bar{a}_{k_1,\ell_1}\cdots a_{i_s j_s}\bar{a}_{k_s,\ell_s} \mathrm{E}x_{i_1}\bar{x}_{k_1}\bar{x}_{j_1}x_{\ell_1}\cdots x_{i_s}\bar{x}_{k_s}\bar{x}_{j_s}x_{\ell_s}.$$

Draw a directional graph G of $p = 2s$ edges that link i_t to j_t and ℓ_t to k_t, $t = 1, \cdots, s$. Note that if G has a vertex whose degree is 1, then the graph corresponds to a term with expectation 0. That is, for any nonzero term, the vertex degrees of the graph are not less than 2. Write the noncoincident vertices as $v_1, \cdots, v_m$ with degrees $p_1, \cdots, p_m$ greater than 1. We have $m \le s$. By assumption, we have

$$|\mathrm{E} x_{i_1} \bar{x}_{k_1} \bar{x}_{j_1} x_{\ell_1} \cdots x_{i_s} \bar{x}_{k_s} \bar{x}_{j_s} x_{\ell_s}| \le (n\eta^2)^{p-m}.$$

Now, suppose that the graph consists of q connected components $G_1, \cdots, G_q$ with $m_1, \cdots, m_q$ noncoincident vertices, respectively. Let us consider the contribution by G_1 to $\mathrm{E}|S_2|^p$. Assume that G_1 has s_1 edges, $e_1, \cdots, e_{s_1}$. Choose a tree G_1' from G_1, and assume its edges are $e_1, \cdots, e_{m_1-1}$, without loss of generality. Note that

$$\sum_{v_1,\cdots,v_{m_1} \le n} \prod_{t=1}^{m_1-1} |a_{e_t}|^2 \le \|\mathbf{A}\|^{2m_1-2} n$$

and

$$\sum_{v_1,\cdots,v_{m_1} \le n} \prod_{t=m_1}^{s_1} |a_{e_t}|^2 \le \|\mathbf{A}\|^{2s_1-2m_1+2} n^{m_1-1}.$$

Here, the first inequality follows from the fact that $\sum_{v_1} |a_{v_1 v_2}|^2 \le \|\mathbf{A}\|^2$ since it is a diagonal element of $\mathbf{A}\mathbf{A}^*$. The second inequality follows from the fact that $\sum_{v_1} |a_{v_1 v_2}|^\ell \le \|\mathbf{A}\|^\ell$ for any $\ell \ge 2$ and that $s_1 \ge m_1$ since all vertices have degrees not less than 2. Therefore, the contribution of G_1 is bounded by

$$\begin{aligned}
&\sum_{v_1,\cdots,v_{m_1} \le n} \prod_{t=1}^{s_1} |a_{e_t}| \\
&\le \left(\sum_{v_1,\cdots,v_{m_1} \le n} \prod_{t=1}^{m_1-1} |a_{e_t}|^2 \sum_{v_1,\cdots,v_{m_1} \le n} \prod_{t=m_1}^{s_1} |a_{e_t}|^2 \right)^{1/2} \\
&\le \|\mathbf{A}\|^{s_1} n^{m_1/2}.
\end{aligned}$$

Noting that $m_1 + \cdots + m_q = m$ and $s_1 + \cdots + s_q = 2s$, eventually we obtain that the contribution of the isomorphic class for a given canonical graph is $\|\mathbf{A}\|^{2s} n^{m/2}$. Because the two vertices of each edge cannot coincide, we have $q \le m/2$. The number of canonical graphs is less than $\binom{m}{2}^p \le m^{2p}$. We finally obtain

$$\begin{aligned}
\mathrm{E}\,|S_2|^p &\le \|\mathbf{A}\|^{2s} \sum_{m=2}^{s} n^{m/2} (n\eta^2)^{p-m} m^{2p} \\
&\le n^p (\|\mathbf{A}\|\eta^2)^p \sum_{m=2}^{s} (n^{1/2}\eta^2)^{-m} m^{2p}
\end{aligned}$$

$$\leq \begin{cases} n^p s(\|\mathbf{A}\|\eta^2)^p (n^{1/2}\eta^2)^{-2}(4b)^{2p}, & \text{if } 2p/\log(n^{1/2}\eta^2) > 2, \\ n^p s(\|\mathbf{A}\|\eta^2)^p (n^{1/2}\eta^2)^{-2} 2^{2p}, & \text{otherwise} \end{cases}$$
$$= n^p(n\eta^4)^{-1}(16 s^{1/p} b^2 \|\mathbf{A}\|\eta^2)^p.$$

Since $\mathrm{E}|S_1 + S_2|^p \leq 2^{p-1}(\mathrm{E}|S_1|^p + \mathrm{E}|S_2|^p)$ and $16s^{1/p} \leq 16e^{1/2e} \leq 20$, the proof of the lemma is complete.

9.2 CLT of LSS for the Wigner Matrix

In this section, we shall consider the case where F_n is the ESD of the normalized Wigner matrix $\mathbf{W}_n$. More precisely, let $\mu(f)$ denote the integral of a function f with respect to a signed measure μ. Let $\mathcal{U}$ be an open set of the complex plane that contains the interval $[-2, 2]$; i.e., the support of the semicircular law F.

To facilitate the exposition, complex Wigner matrices will be called the *complex Wigner ensemble* (CWE) and real Wigner matrices will be called the *real Wigner ensemble* (RWE). In both cases, the entries are not necessarily identically distributed. If in addition the entries are Gaussian (with $\sigma^2 = 1$ and 2 for the CWE and RWE, respectively), the ensembles above are the classical Gaussian unitary ensemble (GUE) and the Gaussian orthogonal ensemble (GOE) of random matrices.

Next define $\mathcal{A}$ to be the set of analytic functions $f : \mathcal{U} \mapsto \mathbb{C}$. We then consider the empirical process $G_n := \{G_n(f)\}$ indexed by $\mathcal{A}$; i.e.,

$$G_n(f) := n \int_{-\infty}^{\infty} f(x)[F_n - F](dx) , \quad f \in \mathcal{A}. \tag{9.2.1}$$

To study the weak limit of G_n, we need conditions on the moments of the entries x_{ij} of the Wigner matrices $\sqrt{n}\mathbf{W}_n$. Note that the distributions of entries x_{ij} are allowed to depend on n, but the dependence on n is suppressed. Let:

[M1] For all i, $\mathrm{E}|x_{ii}|^2 = \sigma^2 > 0$, for all $i < j$, $\mathrm{E}|x_{ij}|^2 = 1$, and for CWE, $\mathrm{E}x_{ij}^2 = 0$.

[M2] (homogeneity of fourth moments) $M = \mathrm{E}|x_{ij}|^4$ for $i \neq j$.

[M3] (uniform tails) For any $\eta > 0$, as $n \to \infty$,

$$\frac{1}{\eta^4 n^2} \sum_{i,j} \mathrm{E}[|x_{ij}|^4 I(|x_{ij}| \geq \eta\sqrt{n})] = o(1).$$

Note that condition [M3] implies the existence of a sequence $\eta_n \downarrow 0$ such that

$$(\eta_n\sqrt{n})^{-4}\sum_{i,j}\mathrm{E}[|x_{ij}|^4 I(|x_{ij}|\geq \eta_n\sqrt{n})]=o(1). \tag{9.2.2}$$

Note that $\eta_n \to 0$ may be assumed to be as slow as desired. For definiteness, we assume that $\eta_n > 1/\log n$.

The main result of this section is the finite-dimensional convergence of the empirical process G_n to a Gaussian process. That is, for any k elements $f_1, \cdots, f_k$ of $\mathcal{A}$, the vector $(G_n(f_1), \cdots, G_n(f_k))$ converges weakly to a p-dimensional Gaussian distribution.

Let $\{T_k\}$ be the family of Tchebychev polynomials and define, for $f \in \mathcal{A}$ and any integer $\ell \geq 0$,

$$\begin{aligned}\tau_\ell(f) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(2\cos(\theta))e^{i\ell\theta}d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(2\cos(\theta))\cos(\ell\theta)d\theta \\ &= \frac{1}{\pi}\int_{-1}^{1} f(2t)T_\ell(t)\frac{1}{\sqrt{1-t^2}}dt. \end{aligned}\tag{9.2.3}$$

In order to give a unified statement for both ensembles, we introduce the parameter κ with values 1 and 2 for the complex and real Wigner ensembles, respectively. Moreover, set $\beta = \mathrm{E}(|x_{12}|^2-1)^2-\kappa$. In particular, for the GUE we have $\kappa = \sigma^2 = 1$ and for the GOE we have $\kappa = \sigma^2 = 2$, and in both cases $\beta = 0$.

We shall prove the following theorem that extends a result given in Bai and Yao [35].

Theorem 9.2. *Under conditions [M1]–[M3], the spectral empirical process $G_n = (G_n(f))$ indexed by the set of analytic functions $\mathcal{A}$ converges weakly in finite dimension to a Gaussian process $G := \{G(f) : f \in \mathcal{A}\}$ with mean function* $\mathrm{E}[G(f)]$ *given by*

$$\frac{\kappa-1}{4}\{f(2)+f(-2)\}-\frac{\kappa-1}{2}\tau_0(f)+(\sigma^2-\kappa)\tau_2(f)+\beta\tau_4(f) \tag{9.2.4}$$

and the covariance function $c(f,g) := \mathrm{E}[\{G(f)-\mathrm{E}G(f)\}\{G(g)-\mathrm{E}G(g)\}]$ given by

$$\sigma^2\tau_1(f)\tau_1(g)+2(\beta+1)\tau_2(f)\tau_2(g)+\kappa\sum_{\ell=3}^{\infty}\ell\tau_\ell(f)\tau_\ell(g) \tag{9.2.5}$$

$$=\frac{1}{4\pi^2}\int_{-2}^{2}\int_{-2}^{2} f'(t)g'(s)V(t,s)dtds, \tag{9.2.6}$$

where

$$V(t,s)=\left(\sigma^2-\kappa+\frac{1}{2}\beta ts\right)\sqrt{(4-t^2)(4-s^2)}$$

$$+\kappa \log\left(\frac{4-ts+\sqrt{(4-t^2)(4-s^2)}}{4-ts-\sqrt{(4-t^2)(4-s^2)}}\right). \tag{9.2.7}$$

Note that our definition implies that the variance of $G(f)$ equals $c(f,\bar{f})$. Let $\delta_a(dt)$ be the Dirac measure at a point a. The mean function can also be written as

$$\mathrm{E}[G(f)] = \int_{\mathbb{R}} f(2t)d\nu(t) \tag{9.2.8}$$

with signed measure

$$\begin{aligned} d\nu(t) &= \frac{\kappa-1}{4}[\delta_1(dt)+\delta_{-1}(dt)] \\ &+\frac{1}{\pi}\left[-\frac{\kappa-1}{2}+(\sigma^2-\kappa)T_2(t)+\beta T_4(t)\right]\frac{1}{\sqrt{1-t^2}}I([-1,1])(t)dt. \end{aligned} \tag{9.2.9}$$

In the cases of GUE and GOE, the covariance reduces to the third term in (9.2.5). The mean $\mathrm{E}[G(f)]$ is always zero for the GUE since in this case $\sigma^2=\kappa=1$ and $\beta=0$. As for the GOE, since $\beta=0$ and $\sigma^2=\kappa=2$, we have

$$\mathrm{E}[G(f)] = \frac{1}{4}\{f(2)+f(-2)\} - \frac{1}{2}\tau_0(f).$$

Therefore the limit process is not necessarily centered.

Example 9.3. Consider the case where $\mathcal{A}=\{f(x,t)\}$ and the stochastic process is

$$Z_n(t) = \sum_{k=1}^{n} f(\lambda_k,t) - n\int_{-2}^{2} f(x,t)F(dx).$$

If both f and $\partial f(x,t)/\partial t$ are analytic in x over a region containing $[-2,2]$, it follows easily from Theorem 9.2 that $Z_n(t)$ converges to a Gaussian process. Its finite-dimensional convergence is exactly the same as in Theorem 9.2, while its tightness can be obtained as a simple consequence of the same theorem.

9.2.1 Strategy of the Proof

Let $\mathcal{C}$ be the contour made by the boundary of the rectangle with vertices $(\pm a \pm iv_0)$, where $a>2$ and $1 \ge v_0 > 0$. We can always assume that the constants $a-2$ and v_0 are sufficiently small so that $\mathcal{C}\subset\mathcal{U}$.

Then, as mentioned in Section 9.1,

$$G_n(f) = -\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)n[s_n(z)-s(z)]dz, \tag{9.2.10}$$

where s_n and s are Stieltjes transforms of $\mathbf{W}_n$ and the semicircular law, respectively. The reader is reminded that the equality above may not be correct when some eigenvalues of $\mathbf{W}_n$ run outside the contour. A corrected version of (9.2.10) should be

$$G_n(f)I(B_n^c) = -\frac{1}{2\pi i} I(B_n^c) \oint_{\mathcal{C}} f(z) n[s_n(z) - s(z)]dz,$$

where $B_n = \{|\lambda_{ext}(\mathbf{W}_n)| \geq 1 + a/2\}$ and λ_{ext} denotes the smallest or largest eigenvalue of the matrix $\mathbf{W}_n$. But this difference will not matter in the proof because by Remark 5.7, after truncation and renormalization, for any $a > 2$ and $t > 0$,

$$\mathrm{P}(B_n) = o(n^{-t}). \tag{9.2.11}$$

This property will also be used in the proof of Corollary 9.8 later.

This representation reduces our problem to showing that the process $M_n := (M_n(z))$ indexed by $z \notin [-2, 2]$, where

$$M_n(z) = n[s_n(z) - s(z)], \tag{9.2.12}$$

converges to a Gaussian process $M(z)$, $z \notin [-2, 2]$. We will show this conclusion by the following theorem.

Throughout this section, we set $\mathbb{C}_0 = \{z = u + iv \ : \ |v| \geq v_0\}$.

Theorem 9.4. *Under conditions [M1]–[M3], the process $\{M_n(z);\ \mathbb{C}_0\}$ converges weakly to a Gaussian process $\{M(z);\ \mathbb{C}_0\}$ with the mean and covariance functions given in Lemma 9.5 and Lemma 9.6.*

Since the mean and covariance functions of $M(z)$ are independent of v_0, the process $\{M(z);\ \mathbb{C}_0\}$ in Theorem 9.4 can be taken as a restriction of a process $\{M(z)\}$ defined on the whole complex plane except the real axis. Further, by noting the symmetry, $M(\bar{z}) = \overline{M(z)}$, and the continuity of the mean and covariance functions of $M(z)$ on the real axis except for $z \in [-2, 2]$, we may extend the process to $\{M(z);\ \Re z \notin [-2, 2]\}$.

Split the contour $\mathcal{C}$ as the union $\mathcal{C}_u + \mathcal{C}_l + \mathcal{C}_r + \mathcal{C}_0$, where $\mathcal{C}_l = \{z = -a + iy,\ \zeta_n n^{-1} < |y| \leq v_1\}$, $\mathcal{C}_r = \{z = a + iy,\ \zeta_n n^{-1} < |y| \leq v_1\}$, and $\mathcal{C}_0 = \{z = \pm a + iy,\ |y| \leq n^{-1}\zeta_n\}$, where $\zeta_n \to 0$ is a slowly varying sequence of positive constants. By Theorem 9.4, we get the weak convergence

$$\int_{\mathcal{C}_u} M_n(z)dz \Rightarrow \int_{\mathcal{C}_u} M(z)dz.$$

To prove Theorem 9.2, we only need to show that, for $j = l, r, 0$,

$$\lim_{v_1 \downarrow 0} \limsup_{n \to \infty} \mathrm{E} \left| \int_{\mathcal{C}_j} M_n(z) I(B_n^c) dz \right|^2 = 0 \tag{9.2.13}$$

and

$$\lim_{v_1 \downarrow 0} \mathrm{E} \left| \int_{\mathcal{C}_j} M(z) dz \right|^2 = 0. \tag{9.2.14}$$

Estimate (9.2.14) can be verified directly by the mean and variance functions of $M(z)$. The proof of (9.2.13) for the case $j = 0$ will be given at the end of Subsection 9.2.2, and the proof of (9.2.13) for $j = l$ and r will be postponed until the proof of Theorem 9.4 is complete.

9.2.2 Truncation and Renormalization

Choose $\eta_n > 1/\log n$ according to (9.2.2), and we first truncate the variables as $\widehat{x}_{ij} = x_{ij} I(|x_{ij}| \le \eta_n \sqrt{n})$. We need to further normalize them by setting $\widetilde{x}_{ij} = (\widehat{x}_{ij} - \mathrm{E}\widehat{x}_{ij})/\sigma_{ij}$ for $i \ne j$ and $\widetilde{x}_{ii} = \sigma(\widehat{x}_{ii} - \mathrm{E}\widehat{x}_{ii})/\sigma_{ii}$, where σ_{ij} is the standard deviation of $\widehat{x}_{ij}$.

Let $\widehat{F}_n$ and $\widetilde{F}_n$ be the ESD of the random matrices $(\frac{1}{\sqrt{n}}\widehat{x}_{ij})$ and $(\frac{1}{\sqrt{n}}\widetilde{x}_{ij})$, respectively. According to (9.2.1), we similarly define $\widehat{G}_n$ and $\widetilde{G}_n$. First observe that

$$\mathrm{P}(G_n \ne \widehat{G}_n) \le \mathrm{P}(F_n \ne \widehat{F}_n) = o(1). \tag{9.2.15}$$

Indeed,

$$\begin{aligned} \mathrm{P}(F_n \ne \widehat{F}_n) &\le \mathrm{P}\{\text{ for some } i,j,\ \widehat{x}_{ij} \ne x_{ij}\} \\ &\le \sum_{i,j} \mathrm{P}\left\{|x_{ij}| \ge \eta_n \sqrt{n}\right\} \\ &\le (\eta_n \sqrt{n})^{-4} \sum_{i,j} \mathrm{E}[|x_{ij}|^4 I(|x_{ij}| \ge \eta_n \sqrt{n})] = o(1). \end{aligned}$$

Secondly, as f is analytic, by conditions [M2] and [M3] we have

$$\begin{aligned} &\mathrm{E}\left|\widetilde{G}_n(f) - \widehat{G}_n(f)\right|^2 \\ &\le C\mathrm{E}\left(\sum_{j=1}^n |\tilde{\lambda}_{nj} - \hat{\lambda}_{nj}|\right)^2 \\ &\le Cn\mathrm{E}\sum_{j=1}^n |\tilde{\lambda}_{nj} - \hat{\lambda}_{nj}|^2 \\ &\le Cn\mathrm{E}\sum_{ij} |n^{-1/2}(\tilde{x}_{ij} - \hat{x}_{ij})|^2 \end{aligned}$$

$$\leq C\Bigg[\sum_{i\neq j}[\mathrm{E}|x_{ij}|^2|1-\sigma_{ij}^{-1}|^2+|\mathrm{E}(\hat{x}_{ij})|^2\sigma_{ij}^{-2}]$$
$$+\sum_i[\mathrm{E}|x_{ii}|^2|1-\sigma\sigma_{ii}^{-1}|^2+|\mathrm{E}(\hat{x}_{ii})|^2\sigma_{ii}^{-2}]\Bigg]$$
$$\leq C\Bigg[\sum_{ij}[(n\eta_n^2)^{-2}+2(n\eta_n^2)^{-3}]\mathrm{E}^2|x_{ij}^4|I(|x_{ij}|\geq\sqrt{n}\eta_n)\Bigg]$$
$$=o_p(1),$$

where $\tilde{\lambda}_{nj}$ and $\hat{\lambda}_{nj}$ are the j-th largest eigenvalues of the Wigner matrices $n^{-1/2}(\tilde{x}_{ij})$ and $n^{-1/2}(\hat{x}_{ij})$, respectively. Therefore the weak limit of the variables $(G_n(f))$ is not affected if the original variables x_{ij} are replaced by the normalized truncated variables $\widetilde{x}_{ij}$.

From the normalization, the variables $\widetilde{x}_{ij}$ all have mean 0 and the same absolute second moments as the original variables. However, the fourth moments of the off-diagonal elements are no longer homogenous and, for the CWE, $\mathrm{E}x_{ij}^2$ is no longer 0. However, this does not matter because $\left|\sum_{i\neq j}[\mathrm{E}|x_{ij}^4|-\mathrm{E}|\widetilde{x}_{ij}|^4]\right|=o(n^{-2})$ and $\max_{i<j}|\mathrm{E}\widetilde{x}_{ij}^2|=O(1/n)$.

We now assume that the conditions above hold, and we still use x_{ij} to denote the truncated and normalized $\widetilde{x}_{ij}$ variables.

The proof of (9.2.13) for $j=0$. When B_n does not occur, for any $z\in\mathcal{C}_0$ we have $|s_n(z)|\leq 2/(a-2)$ and $|s(z)|\leq 1/(a-2)$. Hence,

$$\int_{\mathcal{C}_0}\mathrm{E}|M_n(z)I(B_n^c)|dz\leq 4n(2/(a-2))\|\mathcal{C}_0\|\to 0,$$

where $\|\mathcal{C}_0\|$ denotes the length of the segment $\mathcal{C}_0$.

9.2.3 Mean Function of M_n

Recalling (8.1.18), for $z\in\mathbb{C}_0$ we have

$$\mathrm{E}M_n(z)=n[\mathrm{E}s_n(z)-s(z)]=[1+s'(z)]n\mathrm{E}\delta(z)\{1+o(1)\}.$$

Lemma 9.5. *The mean function* $\mathrm{E}M_n(z)$ *uniformly tends to*

$$b(z)=[1+s'(z)]s(z)^3\Big[\sigma^2-1+(\kappa-1)s'(z)\beta s^2(z)\Big]$$

for $z\in\mathbb{C}_0$ and for both ensembles CWE and RWE.

Proof. For use in the proof of (9.2.13) with $j=l,r$, we show a stronger result that

$$\mathrm{E}M_n(z) - b(z) \to 0$$

uniformly in $z \in \mathcal{C}_n = \mathcal{C}_u + \mathcal{C}_l + \mathcal{C}_r$.

By (2.3.6), we have

$$n\delta_n(z) = -\sum_{k=1}^{n} b_n(z)\mathrm{E}\varepsilon_k\beta_k,$$

where $b_n(z)$ and β_k are defined above (8.1.19). By (8.1.18), to prove the lemma, it is enough to show that

$$n\delta_n - d(z) \to 0 \tag{9.2.16}$$

uniformly in $z \in \mathcal{C}_n$, where $d(z) = s(z)^3\Big[\sigma^2 - 1 + (\kappa - 1)s'(z)\beta s^2(z)\Big]$.

Using the identity for any integer p,

$$\frac{1}{u-\varepsilon} = \frac{1}{u}\left[1 + \frac{\varepsilon}{u} + \cdots + \frac{\varepsilon^p}{u^p} + \frac{\varepsilon^{p+1}}{u^p(u-\varepsilon)}\right],$$

we get

$$\begin{aligned} n\delta_n(z) &= -\sum_{k=1}^{n} b_n^2\mathrm{E}\varepsilon_k - \sum_{k=1}^{n} b_n^3\mathrm{E}\varepsilon_k^2 - \sum_{k=1}^{n} b_n^3\mathrm{E}\beta_k\varepsilon_k^3 \\ &= S_1 + S_2 + S_3. \end{aligned}$$

First we prove that $S_3 = o(1)$. To this end, we shall frequently use the fact that if $\sup_{z\in\mathcal{C}_n}|H(z)|I_{B_{nk}^c} < K$ and $\sup_{z\in\mathcal{C}_n}|H(z)| < Kn^{\iota}$ for some constant $\iota > 0$, where $B_{nk} = \{|\lambda_{ext}(\mathbf{W}_k)| \ge 1 + a/2\}$, then, for any t,

$$\sup_{z\in\mathcal{C}_n}\mathrm{E}|H(z)| < K + o(n^{-t}). \tag{9.2.17}$$

This inequality is an easy consequence of (9.2.11) with the fact that $B_{nk} \subset B_n$. Further, we claim that if $|H(z)| \le n^{\iota}$ uniformly in $z \in \mathcal{C}_n$ for some $\iota > 0$, then

$$\mathrm{E}|\beta_k H(z)| \le 2\mathrm{E}|H(z)| + o(n^{-t}) \tag{9.2.18}$$

uniformly in $z \in \mathcal{C}_n$.

Examining the proof of (8.1.19), one can prove that

$$\left|\tilde{b}_n\right| < 1 \tag{9.2.19}$$

along the same lines, where $\tilde{b}_n = \frac{1}{z+s_n(z)}$.

Then $|\beta_k| > 2$ implies that $|\tilde{\varepsilon}_k| = |\beta_k^{-1} - \tilde{b}_n^{-1}| > \frac{1}{2}$, where

$$\tilde{\varepsilon}_k = \frac{1}{\sqrt{n}} x_{kk} - \frac{1}{n}\Big(\boldsymbol{\alpha}_k^* \mathbf{D}_k^{-1} \boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{D}^{-1}\Big), \tag{9.2.20}$$

$\mathbf{D}_k = \mathbf{W}_k - z\mathbf{I}$, and $\mathbf{D} = \mathbf{W} - z\mathbf{I}$.

Note that

$$\left|\frac{1}{\sqrt{n}} x_{kk}\right| \le \eta_n \to 0$$

and

$$\begin{aligned}
&\Big|\mathrm{tr}\mathbf{D}_k^{-1} - \mathrm{tr}(\mathbf{D}^{-1})\Big| I_{B_n^c} \\
&\le \left[\sum_{j=1}^{n-1} \frac{|\lambda_j - \lambda_{kj}|}{|(\lambda_j - z)(\lambda_{kj} - z)|} + \frac{1}{|\lambda_n - z|}\right] I_{B_n^c} \\
&\le K \left[\sum_{j=1}^{n-1} (\lambda_j - \lambda_{kj}) + 1\right] I_{B_n^c} \\
&\le K[\lambda_1 - \lambda_n + 1] I_{B_n^c} \quad \text{(by the interlacing theorem)} \\
&\le K(2a+1),
\end{aligned} \tag{9.2.21}$$

where λ_j and λ_{kj} are the eigenvalues of $\mathbf{W}$ and $\mathbf{W}_k$ in decreasing order, respectively.

Therefore, for all large n, by Lemma 9.1 we have

$$\begin{aligned}
&\mathrm{E}|\beta_k H(z)| \le 2\mathrm{E}|H(z)| + n^\iota \mathrm{P}\left(|\tilde{\varepsilon}_k| \ge \frac{1}{2}\right) \\
&\le 2\mathrm{E}|H(z)| + n^\iota \mathrm{P}\left(|\boldsymbol{\alpha}_k^* \mathbf{D}_k^{-1} \boldsymbol{\alpha}_x - \mathrm{tr}\mathbf{D}_k^{-1}| \ge \frac{n}{4},\ B_{nk}^c\right) + n^\iota \mathrm{P}(B_n) \\
&\le 2\mathrm{E}|H(z)| + o(n^{-t}) + K n^\iota \mathrm{E}\left|\frac{1}{n}(\boldsymbol{\alpha}_k^* \mathbf{D}_k^{-1} \boldsymbol{\alpha}_x - \mathrm{tr}\mathbf{D}_k^{-1})\right|^\ell I_{B_{nk}^c} \\
&\le 2\mathrm{E}|H(z)| + o(n^{-t}) + K n^{\iota - 1} (K\eta_n)^{\ell - 4} = 2\mathrm{E}|H(z)| + o(n^{-t})
\end{aligned}$$

uniformly in $z \in \mathcal{C}_n$, provided that ℓ is chosen as $\eta_n^{-1/2} \log(n\nu^{-1}\eta_n^4)$.

Now, let us apply (9.2.18) to prove $S_3 = o(1)$ uniformly in $z \in \mathcal{C}_n$. Choose $H = |\varepsilon_k|^3$. By noting $|\beta_k| \le 1/v$ and $|\varepsilon_k| \le Knv^{-1}$, one needs only to verify that

$$\mathrm{E}|\varepsilon_k|^3 \le (n^{-3/2})$$

uniformly in $z \in \mathcal{C}_n$ and $k \le n$.

By estimation from (9.2.11) and Lemma 9.1, we have

$$\begin{aligned}
&\mathrm{E}\left|\frac{1}{n}(\boldsymbol{\alpha}_k^* \mathbf{D}_k^{-1} \boldsymbol{\alpha}_x - \mathrm{tr}\mathbf{D}_k^{-1})\right|^3 \\
&\le K n^{-1} \eta_n^2 \mathrm{E}\|\mathbf{D}_k\|^3 \\
&\le K n^{-1} \eta_n^2 + K v^{-3} P(B_n)
\end{aligned}$$

$$\leq o(n^{-1})$$

uniformly in $z \in \mathcal{C}_n$ and $k \leq n$.

By the martingale decomposition (see Lemma 8.7) and Burkholder inequality (Lemma 2.12), for any fixed $t \geq 2$,

$$\begin{aligned} \mathrm{E}|s_n - \mathrm{E}s_n(z)|^t &= n^{-3}\mathrm{E}\left|\sum_{k=1}^n \gamma_k\right|^t \\ &\leq Kn^{-t}\mathrm{E}\left(\sum_{k=1}^n |\gamma_k|^2\right)^{t/2} \\ &\leq Kn^{-t/2-1}\sum_{k=1}^n \mathrm{E}|\gamma_k|^t. \end{aligned} \tag{9.2.22}$$

Recall that

$$\begin{aligned} \gamma_k &= -(\mathrm{E}_k - \mathrm{E}_{k-1})\beta_k\left(1 + \frac{1}{n}\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k\right) \\ &= -\frac{1}{n}\mathrm{E}_k\tilde{b}_n(\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{D}_k^{-2}) \\ &\quad -(\mathrm{E}_k - \mathrm{E}_{k-1})\tilde{b}_n\beta_k\tilde{\varepsilon}_k(1 + \boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k). \end{aligned}$$

Applying Lemma 9.1 and using (9.2.17), it follows that

$$\begin{aligned} &\mathrm{E}\left|\frac{1}{n}\mathrm{E}_k\tilde{b}_n(\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{D}_k^{-2})\right|^t \\ &\leq Kn^{-1}\eta_n^{2t-4}\mathrm{E}\|\mathbf{D}_k^{-1}\|^{2t} \\ &\leq Kn^{-1}\eta_n^{2t-4}. \end{aligned}$$

Also, applying (9.2.18) twice and Lemma 9.1, we obtain

$$\begin{aligned} &\mathrm{E}|(\mathrm{E}_k - \mathrm{E}_{k-1})\tilde{b}_n\beta_k\tilde{\varepsilon}_k(1 + \boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k)|^t \\ &\leq 4\mathrm{E}|\tilde{\varepsilon}_k(1 + \boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k)|^t + o(n^{-t}) \\ &\leq K\left(\mathrm{E}|\tilde{\varepsilon}_k|^{2t}\mathrm{E}|1 + \boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k|^{2t}\right)^{1/2} + o(n^{-t}) \\ &\leq o(n^{-1/2}), \end{aligned}$$

so that from (9.2.22) it follows that

$$\mathrm{E}|s_n - \mathrm{E}s_n(z)|^t \leq o(n^{-(t+1)/2}) \tag{9.2.23}$$

uniformly in $z \in \mathcal{C}_n$.

Finally, taking $t = 3$ in (9.2.23) and using (9.2.21) and the fact that $\mathrm{E}\left|\frac{1}{\sqrt{n}}x_{kk}\right|^3 \le Kn^{-3/2}$, we conclude that

$$\mathrm{E}|\varepsilon_k|^3 = o(n^{-1}),$$

which completes the proof that

$$S_3 = o(1)$$

uniformly for $z \in \mathcal{C}_n$.

Next, we find the limit of $\mathrm{E}\varepsilon_k$. We have

$$\begin{aligned}\mathrm{E}\varepsilon_k &= \mathrm{E}\Big(\frac{x_{k,k}}{\sqrt{n}} - n^{-1}\alpha_k^*\mathbf{D}_n^{-1}\alpha_k\Big) + \mathrm{E}s_n(z)\\ &= n^{-1}[\mathrm{E}\mathrm{tr}\mathbf{D}^{-1} - \mathrm{E}\mathrm{tr}\mathbf{D}_n^{-1}]\\ &= -\frac{1}{n}\mathrm{E}\beta_k(1 + n^{-1}\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k).\end{aligned}$$

By (9.2.17) and Lemma 9.1, we have

$$\begin{aligned}&\mathrm{E}|n^{-1}[\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{D}_k^{-2}]|\\ &\le n^{-1}(\mathrm{E}|\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{D}_k^{-2}|^2)^{1/2}\\ &\le Kn^{-1/2}(\mathrm{E}\|\mathbf{D}_k^{-1}\|^4)^{1/2}\\ &\le o(1).\end{aligned}$$

If F_{nk} denotes the ESD of $\mathbf{W}_k$, then by the interlacing theorem, we have

$$\|F_n - F_{nk}\| \le \frac{1}{n}.$$

Since $F_n \to F$, by the semicircular law with probability 1 and $\|F_{nk} - F_n\| \le 1/n$ by the interlacing theorem, we have

$$\max_{k\le n}\|F_{nk} - F\| \to 0, \text{ a.s.}$$

Again, by (9.2.17) we have

$$\begin{aligned}&\sup_{z\in\mathcal{C}_n}\mathrm{E}\left|\frac{1}{n}\mathrm{tr}\mathbf{D}_k^{-2} - s'(z)\right\|\\ &= \sup_{z\in\mathcal{C}_n}\left|\frac{n-1}{n}\int\frac{dF_{nk}(x)}{(x-z)^2} - \int\frac{dF(x)}{(x-z)^2}\right|\\ &= o(1). \qquad (9.2.24)\end{aligned}$$

Applying (9.2.18), we obtain

$$\sum_{k=1}^{n} \mathrm{E}\varepsilon_k = -\frac{1+s'(z)}{n}\sum_{k=1}^{n} \mathrm{E}\beta_k + o(1) = (1+s'(z))\mathrm{E}s_n(z) + o(1),$$

where $o(1)$ is uniform for $z \in \mathcal{C}_n$.

Applying (9.2.17) and $\|F_n - F\| \to 0$ a.s., we have

$$\sup_{z\in\mathcal{C}_n} |\mathrm{E}s_n(z) - s(z)| \le o(1),$$

which implies that

$$\sum_{k=1}^{n} \mathrm{E}\varepsilon_k = s(z)(1+s'(z)) + o(1). \tag{9.2.25}$$

Now, let us find the approximation of $\mathrm{E}\varepsilon_k^2$. By the previous estimation for $\mathrm{E}\varepsilon_k$, we have

$$\sum_{k=1}^{n} \mathrm{E}\varepsilon_k^2 = \sum_{k=1}^{n} \mathrm{E}(\varepsilon_k - \mathrm{E}\varepsilon_k)^2 + O(n^{-1}),$$

where $O(n^{-1})$ is uniform in $z \in \mathcal{C}_n$.

Furthermore, by the definition of ε_k, we have

$$\begin{aligned}\varepsilon_k - \mathrm{E}\varepsilon_k &= \frac{1}{\sqrt{n}}x_{kk} - n^{-1}[\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-1}\alpha_k - \mathrm{E}\mathbf{D}_k^{-1}] \\ &= \frac{1}{\sqrt{n}}x_{kk} - n^{-1}[\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-1}\boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{D}_k^{-1}] + n^{-1}[\mathrm{tr}\mathbf{D}_k^{-1} - \mathrm{E}\mathrm{tr}\mathbf{D}_k^{-1}]\ .\end{aligned}$$

Therefore

$$\begin{aligned}\mathrm{E}[\varepsilon_k - \mathrm{E}\varepsilon_k]^2 &= \frac{\sigma^2}{n} + \frac{1}{n^2}\mathrm{E}[\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-1}\alpha_k - \mathrm{tr}\mathbf{D}_k^{-1}]^2 \\ &\quad + \frac{1}{n^2}\mathrm{E}[\mathrm{tr}\mathbf{D}_k^{-1} - \mathrm{E}\mathrm{tr}\mathbf{D}_k^{-1}]^2.\end{aligned} \tag{9.2.26}$$

From (9.2.23) with $t = 2$, we have

$$n^{-2}\mathrm{E}\left[\mathrm{tr}\mathbf{D}_k^{-1} - \mathrm{E}\mathrm{tr}\mathbf{D}_k^{-1}\right]^2 = o(n^{-3/2}).$$

By simple calculation, for matrices $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$, we have the identity

$$\begin{aligned}&\mathrm{E}(\boldsymbol{\alpha}_k^*\mathbf{A}\boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{A})(\boldsymbol{\alpha}_k^*\mathbf{B}\boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{B}) \\ &= \mathrm{tr}\mathbf{AB} + \sum_{i,j} a_{ij}b_{ji}\mathrm{E}x_{ik}^2\mathrm{E}\bar{x}_{jk}^2 + \sum_{i=1}^{n} a_{ii}b_{ii}(\mathrm{E}|x_{ik}|^4 - 2 - |\mathrm{E}x_{ik}^2|^2).\end{aligned}$$

Combining this identity and the assumption that $\mathrm{E}x_{ij}^2 = o(1)$ for the CWE and $\mathrm{E}x_{ij}^2 = 1$ for the RWE, we have

$$\mathrm{E}[\boldsymbol{\alpha}_k^* \mathbf{D}_k^{-1} \boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{D}_k^{-1}]^2 = \kappa \mathrm{E}\left[\mathrm{tr}\mathbf{D}_k^{-2}\right] + \beta \mathrm{E}\left[\sum_i d_{ii}^2\right] + o(n), \tag{9.2.27}$$

where d_{ii} are the diagonal entries of the matrix $\mathbf{D}_k^{-1}$.

Furthermore, by Lemma 9.9 to be given later,

$$\limsup_n \max_{i,k} \sum_{z \in \mathcal{C}_n} \mathrm{E}|d_{ii} - s(z)|^2 = 0.$$

Since $|d_{ii}| \le \max\{\frac{2}{a-2}, 1/v_0\}$ when B_{nk}^c occurs, then by (9.2.17)

$$\begin{aligned}
&\lim_n \max_{i,k} \sum_{z \in \mathcal{C}_n} \mathrm{E}|d_{ii}^2 - s^2(z)| \\
&\le \limsup_n \max_{i,k} \sum_{z \in \mathcal{C}_n} [\mathrm{E}|d_{ii} - s(z)|^2 + 2\mathrm{E}|(d_{ii} - s(z))s(z)|] = 0.
\end{aligned}$$

Hence, by (9.2.24), we obtain

$$\sum_{k=1}^n \mathrm{E}\varepsilon_k^2 = \sigma^2 + \kappa s'(z) + \beta s(z)^2 + o_{L_1}(1),$$

where $o_{L_1}(1)$ is uniform for $z \in \mathcal{C}_n$ in the sense of L_1-convergence.

Summarizing the three terms and noting $|b_n| < 1$, we get

$$n\delta \le K,$$

which implies that $b_n(z) = -s(z+\delta) = -s(z) + o(1)$ and thus

$$n\delta(z) = s^3\left(\sigma^2 - 1 + (\kappa - 1)s' + \beta s^2\right) + o(1).$$

The lemma is proved.

9.2.4 Proof of the Nonrandom Part of (9.2.13) for $j = l, r$

Using the notation defined and results obtained in the last section, it follows that, for $j = l$ or r,

$$\lim_{v_1 \downarrow 0} \limsup_{n \to \infty} \int_{\mathcal{C}_j} |\mathrm{E}M_n(z) - b(z)|\, dz \tag{9.2.28}$$

$$\leq \lim_{v_1 \downarrow 0} \limsup_{n\to\infty} \int_{\mathcal{C}_j} |\mathrm{E}M_n(z)|\, dz + \int_{\mathcal{C}_j} |b(z)|\, dz = 0, \tag{9.2.29}$$

where the first limit follows from the fact that $\sup_{z\in\mathcal{C}_n} |\mathrm{E}M_n(z) - b(z)| \to 0$ and the second follows from the fact that $b(z)$ is continuous, and hence bounded.

9.3 Convergence of the Process $M_n - \mathrm{E}M_n$

9.3.1 Finite-Dimensional Convergence of $M_n - \mathrm{E}M_n$

Following the martingale decomposition given in Section 2.3, we may rewrite

$$M_n(z) - \mathrm{E}M_n(z) = \sum_{k=1}^{n} \gamma_k,$$

where

$$\begin{aligned}
\gamma_k &= (\mathrm{E}_{k-1} - \mathrm{E}_k)\mathrm{tr}\mathbf{D}^{-1} \\
&= (\mathrm{E}_{k-1} - \mathrm{E}_k)(\mathrm{tr}\mathbf{D}^{-1} - \mathrm{tr}\mathbf{D}_k^{-1}) \\
&= (\mathrm{E}_{k-1} - \mathrm{E}_k)a_k - \mathrm{E}_{k-1}d_k\ , \\
a_k &= -\beta_k\tilde{b}_k g_k(1 + n^{-1}\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k), \\
\tilde{b}_k &= \left(z + \frac{1}{n}\mathrm{tr}\mathbf{D}_k\right)^{-1}, \qquad (9.3.1) \\
d_k &= h_k\tilde{b}_k(z), \\
g_k &:= n^{-1/2}x_{kk} - n^{-1}(\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-1}\boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{D}_k^{-1}), \\
h_k &:= n^{-1}(\boldsymbol{\alpha}_k^*\mathrm{tr}\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{D}_k^{-2}). \qquad (9.3.2)
\end{aligned}$$

We have

$$\begin{aligned}
a_k &= -\beta_k\tilde{b}_k(z)g_k(1 + n^{-1}\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k) \\
&= -\tilde{b}_k^2(z)g_k(1 + n^{-1}\mathrm{tr}\mathbf{D}_k^{-2}) - h_k g_k\tilde{b}_k^2(z) \\
&\qquad -\beta_k\tilde{b}_k^2(z)(1 + n^{-1}\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k)g_k^2 \\
&:= a_{k1} + a_{k2} + a_{k3}.
\end{aligned}$$

By noting $|\tilde{b}_n| < 1$, (9.2.18), (9.2.17), and using Lemma 9.1, we obtain

$$\mathrm{E}\left|\sum_{k=1}^{n}(\mathrm{E}_{k-1} - \mathrm{E}_k)a_{k3}\right|^2 = \sum_{k=1}^{n}\mathrm{E}\left|(\mathrm{E}_{k-1} - \mathrm{E}_k)a_{k3}\right|^2$$

$$\leq 4\sum_{k=1}^{n}\mathrm{E}\Big|(1+n^{-1}\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k)^2 g_k^4\Big|$$
$$\leq K\sum_{k=1}^{n}\Big[\mathrm{E}\Big|(1+n^{-1}\mathrm{tr}\mathbf{D}_k^{-2})^2 g_k^4\Big| + \mathrm{E}|h_k^2 g_k^4|\Big]$$
$$\leq K\sum_{k=1}^{n}\Big[\mathrm{E}|(1+n^{-1}\mathrm{tr}\mathbf{D}_k^{-2})^2|[n^{-2}+n^{-1}\eta_n^4\|\mathbf{D}_k\|^4$$
$$+(n^{-1}\eta_n^4\|\mathbf{D}_k\|^4)(n^{-1}\eta_n^{12}\|\mathbf{D}_k\|^4)]^{1/2}\Big]$$
$$= o(1), \tag{9.3.3}$$

where $o(1)$ is uniform in $z \in \mathcal{C}_n$.

For the same reason, we have

$$\mathrm{E}\left|\sum_{k=1}^{n}(\mathrm{E}_{k-1}-\mathrm{E}_k)a_{k2}\right|^2 = \sum_{k=1}^{n}\mathrm{E}\left|(\mathrm{E}_{k-1}-\mathrm{E}_k)a_{k2}\right|^2$$
$$\leq \sum_{k=1}^{n}\mathrm{E}\left|h_k g_k\right|^2$$
$$\leq \frac{1}{v_0^4}\sum_{k=1}^{n}(\mathrm{E}|h_k|^4\mathrm{E}|g_k|^4)^{1/2} = o(1). \tag{9.3.4}$$

Hence, we have

$$M_n(z) - \mathrm{E}M_n(z)$$
$$= \sum_{k=1}^{n}\mathrm{E}_{k-1}\Big[-\tilde{b}_n^2(1+n^{-1}\mathrm{tr}\mathbf{D}_k^{-2})g_k - d_k\Big] + o_{L_2}(1)$$
$$= \sum_{k=1}^{n}\mathrm{E}_{k-1}\psi_k(z) + o_{L_2}(1),$$

where $\psi_k(z) = \dfrac{d}{dz}\phi_k(z)$, $\phi_k(z) = \tilde{b}_n g_k$, and $o_{L_2}(1)$ is uniform in $z \in \mathcal{C}_n$.

Let $\{z_t, t = 1,\cdots,m\}$ be m different points belonging to $\mathbb{C}_0$ (now, we return to assuming $z \in \mathbb{C}_0$). The problem is then reduced to determining the weak convergence of the vector martingale

$$\mathbf{Z}_n := \sum_{k=1}^{n}\mathrm{E}_{k-1}(\psi_k(z_1),\cdots,\psi_k(z_m)) =: \sum_{k=1}^{n}\mathrm{E}_{k-1}\boldsymbol{\Psi}_k. \tag{9.3.5}$$

Lemma 9.6. *Assume conditions [M1]–[M3] are satisfied. For any set of m points $\{z_s, s = 1,\cdots,m\}$ of $\mathbb{C}_0$, the random vector $\mathbf{Z}_n$ converges weakly to an m-dimensional zero-mean Gaussian distribution with covariance matrix* *given, with* $s_j = s(z_j)$, *by*

$$\begin{aligned}\Gamma(z_i, z_j) &= \frac{\partial^2}{\partial z_i \partial z_j}\left[(\sigma^2-\kappa)s_i s_j + \frac{1}{2}\beta(s_i s_j)^2 - \kappa \log(1-s_i s_j)\right]\\ &= s_i' s_j' \left[\sigma^2 - \kappa + 2\beta s_i s_j + \frac{\kappa}{(1-s_i s_j)^2}\right]. \qquad (9.3.6)\end{aligned}$$

Proof. We apply the CLT to the martingale $\mathbf{Z}_n$ defined in (9.3.5). Consider its *hook* process:

$$\Gamma_n(z_i, z_j) := \sum_{k=1}^n \mathrm{E}_k \left[\mathrm{E}_{k-1}\frac{d}{dz_i}\phi_k(z_i)\mathrm{E}_{k-1}\frac{d}{dz_j}\phi_k(z_j)\right].$$

Then we have to check the following two conditions:

[C.1] Lyapounov's condition: for some $a > 2$,

$$\sum_{k=1}^n \mathrm{E}\left\|\mathrm{E}_{k-1}\boldsymbol{\Psi}_k\right\|^a \to 0.$$

[C.2] Γ_n converges in probability to the matrix Γ.

Indeed, the assertion [C.1] follows from Lemma 9.1 with $p = 4$. Now, we begin to derive the limit Γ.

For any $z_1, z_2 \in \mathbb{C}_0$,

$$\Gamma_n(z_1, z_2) = \frac{\partial^2}{\partial z_1 \partial z_2}\sum_{k=1}^n \mathrm{E}_k \phi_k(z_1)\mathrm{E}_{k-1}\phi_k(z_2).$$

Applying Vitali's lemma (see Lemma 2.14), we only need to find the limit of

$$\begin{aligned}&\sum_{k=1}^n \mathrm{E}_k \phi_k(z_1)\mathrm{E}_{k-1}\phi_k(z_2)\\ &= \sum_{k=1}^n \mathrm{E}_k \tilde{b}_k(z_1) g_k(z_1)\mathrm{E}_{k-1}\tilde{b}_k(z_2) g_k(z_2).\end{aligned}$$

Recalling that $\frac{1}{n}\mathrm{tr}\mathbf{D}^{-1} = s_n(z) \xrightarrow{L_2} s(z)$, we obtain

$$\begin{aligned}&\sum_{k=1}^n \mathrm{E}_k \phi_k(z_1)\mathrm{E}_{k-1}\phi_k(z_2)\\ &= s(z_1)s(z_2)\sum_{k=1}^n \mathrm{E}_k\left[\mathrm{E}_{k-1}g_k(z_1)\mathrm{E}_{k-1}g_k(z_2)\right] + o_{L_2}(1)\end{aligned}$$

$$:= s(z_1)s(z_2)\widetilde{\Gamma}_n(z_1,z_2) + o_{L_2}(1).$$

By the definition of g_k, we have

$$\begin{aligned}
&\mathrm{E}_k\left[g_k(z_1)\mathrm{E}_{k-1}g_k(z_2)\right] \\
&= \frac{\sigma^2}{n} + \frac{1}{n^2}\mathrm{E}_k(\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z_1\mathbf{I})^{-1}\boldsymbol{\alpha}_k - \mathrm{tr}(\mathbf{W}_k - z_1\mathbf{I})^{-1}) \\
&\quad \times\mathrm{E}_{k-1}(\boldsymbol{\alpha}_k^*(\mathbf{W}_k - z_2\mathbf{I})^{-1}\boldsymbol{\alpha}_k - \mathrm{tr}(\mathbf{W}_k - z_2\mathbf{I})^{-1}).
\end{aligned}$$

To evaluate the second term, write $\mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_\ell) = (b_{ijk}^{(\ell)})$, $\ell = 1,2$. By a computation similar to that leading to (9.2.27), we get

$$\begin{aligned}
&\mathrm{E}_k\Big[\mathrm{E}_{k-1}(\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-1}(z_1)\boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{D}_k^{-1}(z_1)) \\
&\quad \times\mathrm{E}_{k-1}(\boldsymbol{\alpha}_k^*\mathbf{D}_k^{-1}(z_2)\boldsymbol{\alpha}_k - \mathrm{tr}\mathbf{D}_k^{-1}(z_2))\Big] \\
&= \kappa\sum_{ij>k} b_{ijk}^{(1)}b_{jik}^{(2)} + \beta\sum_{i>k} b_{iik}^{(1)}b_{iik}^{(2)} + o_{L_2}(1).
\end{aligned}$$

Therefore

$$\begin{aligned}
\widetilde{\Gamma}_n(z_1,z_2) &= \sigma^2 + \frac{\kappa}{n^2}\sum_{k=1}^{n}\sum_{ij>k} b_{ijk}^{(1)}b_{jik}^{(2)} + \frac{\beta}{n^2}\sum_{k=1}^{n}\sum_{i>k} b_{iik}^{(1)}b_{iik}^{(2)} + o_{L_2}(1) \\
&= \sigma^2 + \kappa S_1 + S_2 + o_{L_2}(1). \qquad (9.3.7)
\end{aligned}$$

By Lemma 9.9 to be given later, we find that

$$S_2 \to \frac{1}{2}\beta s(z_1)s(z_2) \text{ in } L_2.$$

In the following, let us find the limit of S_1.

9.3.2 Limit of S_1

To evaluate the sum S_1 in (9.3.7), we need the following decomposition. Let $\mathbf{e}_j$ $(j = 1,\cdots,k-1,k+1,\cdots,n)$ be the $(n-1)$-vectors whose j-th (or $(j-1)$-th) element is 1 and others are 0 if $j < k$ (or $j > k$ correspondingly). By definition,

$$\mathbf{D}_k = \sum_{i,j\neq k} n^{-1/2}x_{ij}\mathbf{e}_i\mathbf{e}_j' - z\mathbf{I}_{n-1}.$$

Multiplying both sides by $\mathbf{D}_k^{-1}$ gives the identity

$$z\mathbf{D}_k^{-1} + \mathbf{I}_{n-1} = \sum_{i,j\neq k} n^{-1/2}x_{ij}\mathbf{e}_i\mathbf{e}_j'\mathbf{D}_k^{-1}. \qquad (9.3.8)$$

Let (i, j) be two different indices of k. Define

$$\mathbf{W}_{kij} = \mathbf{W}_n(k) - \frac{1}{\sqrt{n}}\delta_{ij}(x_{ij}\mathbf{e}_i\mathbf{e}_j' + x_{ji}\mathbf{e}_j\mathbf{e}_i'), \tag{9.3.9}$$

where $\mathbf{D}_{kij} = \mathbf{W}_{kij} - z\mathbf{I}_{n-1}$, $\delta_{ij} = 1$ for $i \neq j$, and $\delta_{ii} = \frac{1}{2}$. It is easy to verify that

$$\mathbf{D}_k^{-1} - \mathbf{D}_{kij}^{-1} = -\frac{1}{\sqrt{n}}\mathbf{D}_{kij}^{-1}\delta_{ij}(x_{ij}\mathbf{e}_i\mathbf{e}_j' + x_{ji}\mathbf{e}_j\mathbf{e}_i')\mathbf{D}_k^{-1}. \tag{9.3.10}$$

From (9.3.8) and (9.3.10), we get

$$\begin{aligned}
&z\mathrm{E}_{k-1}\mathbf{D}_k^{-1} \\
&= -\mathbf{I}_{n-1} + \sum_{i,j>k} n^{-1/2}x_{ij}\mathbf{e}_i\mathbf{e}_j'\mathrm{E}_{k-1}\mathbf{D}_{kij}^{-1}, \\
&\quad - \sum_{i,j\neq k} n^{-1}\mathrm{E}_{k-1}x_{ij}\mathbf{e}_i\mathbf{e}_j'\mathbf{D}_{kij}^{-1} \qquad (9.3.11)\\
&\quad \delta_{ij}(x_{ij}\mathbf{e}_i\mathbf{e}_j' + x_{ji}\mathbf{e}_j\mathbf{e}_i')\mathbf{D}_k^{-1} \\
&= -\mathbf{I}_{n-1} + \mathbf{A}_k(z) + \mathbf{B}_k(z) + \mathbf{C}_k(z) + \mathbf{E}_k(z) + \mathbf{F}_k(z), \qquad (9.3.12)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{A}_k(z) &= \frac{1}{\sqrt{n}}\sum_{i,j>k} x_{ij}\mathbf{e}_i\mathbf{e}_j'\mathrm{E}_{k-1}\mathbf{D}_{kij}^{-1}, \\
\mathbf{B}_k(z) &= -s(z)\frac{n-3/2}{n}\sum_{i\neq k}\mathbf{e}_i\mathbf{e}_i'\mathrm{E}_{k-1}\mathbf{D}_k^{-1}, \\
\mathbf{C}_k(z) &= -\frac{1}{n}\sum_{i,j\neq k}\delta_{ij}\mathrm{E}_{k-1}\left(|x_{ij}|^2[\mathbf{e}_j'\mathbf{D}_{kij}^{-1}\mathbf{e}_j - s(z)]\right) \\
&\qquad \mathbf{e}_i\mathbf{e}_i'\mathbf{D}_k^{-1}, \\
\mathbf{E}_k(z) &= -\frac{1}{n}\sum_{i,j\neq k}\delta_{ij}\mathrm{E}_{k-1}(|x_{ij}|^2 - 1)s(z)\mathbf{e}_i\mathbf{e}_i'\mathbf{D}_k^{-1}, \\
\mathbf{F}_k(z) &= -\frac{1}{n}\sum_{i,j\neq k}\delta_{ij}\mathrm{E}_{k-1}x_{ij}^2\mathbf{D}_{kij}^{-1}\mathbf{e}_i\mathbf{e}_j'\mathbf{D}_k^{-1}.
\end{aligned}$$

By (A.2.2), it is easy to see that the norm of a matrix is not less than that of its submatrices. Therefore, we have

$$\left|\sum_{\ell_2>k}\mathbf{e}_{\ell_1}'\mathbf{D}_k^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_2}'\mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_{\ell_1}\right| \leq v_0^{-2}. \tag{9.3.13}$$

Then, for any $k < \ell_1, \ell_2 \leq n$, by applying Lemma 9.9,

$$
\begin{aligned}
&\mathrm{E}\left|\sum_{\ell_2>k}\mathbf{e}'_{\ell_1}\mathbf{C}_k(z_1)\mathbf{e}_{\ell_2}\mathbf{e}'_{\ell_2}\mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_{\ell_1}\right| \\
&=\frac{1}{n}\mathrm{E}\left|\sum_{j,\ell_2>k}\delta_{\ell_1 j}\mathrm{E}_{k-1}|x_{\ell_1 j}\right|^2[\mathbf{e}_j\mathbf{D}_{k\ell_1 j}^{-1}(z_1)\mathbf{e}_j-s(z_1)] \\
&\quad\times\mathbf{e}'_{\ell_1}\mathbf{D}_k^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}'_{\ell_2}\mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_{\ell_1}| \\
&\le\frac{1}{nv_0^2}\sum_{j>k}\mathrm{E}|x_{\ell j}|^2|\mathbf{e}'_j\mathbf{D}_{k\ell_1 j}^{-1}(z_1)\mathbf{e}_j-s(z_1)| \\
&=o(1).
\end{aligned}
\tag{9.3.14}
$$

Again, using (9.3.13) and employing the Cauchy-Schwarz inequality, we obtain

$$
\begin{aligned}
&\mathrm{E}\left|\sum_{\ell_2>k}\mathbf{e}'_{\ell_1}\mathbf{E}_k(z_1)\mathbf{e}_{\ell_2}\mathbf{e}'_{\ell_2}\mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_{\ell_1}\right| \\
&=\frac{1}{n}\mathrm{E}\left|\sum_{j,\ell_2>k}\delta_{\ell_1 j}\mathrm{E}_{k-1}(|x_{\ell j}|^2-1)s(z_1)\mathbf{e}'_{\ell_1}\mathbf{D}_k^{-1}(z_1)\mathbf{e}_{\ell_2}\right. \\
&\qquad\left.\times\mathbf{e}'_{\ell_2}\mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_{\ell_1}\right| \\
&\le\frac{1}{nv_0^2}\mathrm{E}\left|\sum_{j>k}\delta_{\ell_1 j}(|x_{\ell_1 j}|^2-1)\right|=O(n^{-1/2}).
\end{aligned}
\tag{9.3.15}
$$

Next, we estimate $\mathbf{F}_k(z_1)$. Let $\mathbf{H}$ be the matrix whose (i,j)-th entry is

$$
\sum_{\ell>k}\mathbf{e}'_i\mathbf{D}_k^{-1}(z_1)\mathbf{e}_\ell\mathbf{e}'_\ell\mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_j.
$$

Obviously, $\mathbf{H}$ is the product of the submatrices of the last $n-k$ rows of $\mathbf{D}_k^{-1}(z_1)$ and the last $n-k$ columns of $\mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)$. Hence, $\|\mathbf{H}\|\le v_0^{-2}$. Using these, we have

$$
\begin{aligned}
&\sum_{ij\ne k}\mathrm{E}\left|\sum_{\ell>k}\mathbf{e}'_\ell\mathbf{D}_{kij}^{-1}(z_1)\mathbf{e}_i\mathbf{e}'_j\mathbf{H}\mathbf{e}_\ell\right|^2 \\
&\le 2\sum_{ij\ne k}\mathrm{E}\left|\sum_{\ell>k}\mathbf{e}'_\ell\mathbf{D}_k^{-1}(z_1)\mathbf{e}_i\mathbf{e}'_j\mathbf{H}\mathbf{e}_\ell\right|^2 \\
&\quad+\frac{2}{\sqrt{n}}\sum_{ij\ne k}\mathrm{E}\left|\sum_{\ell>k}\mathbf{e}'_\ell\mathbf{D}_k^{-1}(z_1)\delta_{ij}(x_{ij}\mathbf{e}_i\mathbf{e}'_j+x_{ji}\mathbf{e}_j\mathbf{e}'_i)\mathbf{D}_{kij}^{-1}(z_1)\mathbf{e}_i\mathbf{e}'_j\mathbf{H}\mathbf{e}_\ell\right|^2
\end{aligned}
$$

$$\leq 2nv_0^{-6} + \frac{4}{v_0^2\sqrt{n}} \left(\sum_{ij\neq k} \mathrm{E}|x_{ij}|^4 \sum_{ij\neq k} \mathrm{E}|\sum_{\ell>k} \mathbf{e}_\ell' \mathbf{D}_k^{-1}(z_1)\mathbf{e}_i\mathbf{e}_j'\mathbf{H}\mathbf{e}_\ell|^2 \right.$$
$$\left. +|\sum_{\ell>k} \mathbf{e}_\ell' \mathbf{D}_k^{-1}(z_1)\mathbf{e}_j\mathbf{e}_j'\mathbf{H}\mathbf{e}_\ell|^2 \right)^{1/2}$$
$$= O(n^{3/2}),$$

where the last step follows from the fact that, for any $i, j \neq k$,

$$\left| \sum_{\ell>k} \mathbf{e}_\ell' \mathbf{D}_k^{-1}(z_1)\mathbf{e}_i\mathbf{e}_j'\mathbf{H}\mathbf{e}_\ell \right| \leq v_0^{-3}.$$

Applying this inequality, we obtain

$$\frac{1}{n}\mathrm{E}\left| \sum_{\ell_1,\ell_2>k} \mathbf{e}_{\ell_1}' \mathbf{E}_k(z_1)\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_2}' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_{\ell_1} \right|$$
$$\leq \frac{1}{n^2} \sum_{ij\neq k} \delta_{ij} \mathrm{E}\left| x_{ij}^2 \sum_{\ell_1,\ell_2>k} \mathbf{e}_{\ell_1}' \mathbf{D}_{kij}^{-1}(z_1)\mathbf{e}_i \right.$$
$$\left. \times \mathbf{e}_j' \mathbf{D}_k^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_2}' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_{\ell_1} \right|$$
$$\leq \frac{1}{n^2} \left(\sum_{ij\neq k} \delta_{ij} \mathrm{E}|x_{ij}^4| \sum_{ij\neq k} \mathrm{E}| \sum_{\ell_1,\ell_2>k} \mathbf{e}_{\ell_1}' \mathbf{D}_{kij}^{-1}(z_1)\mathbf{e}_i \right.$$
$$\left. \mathbf{e}_j' \mathbf{D}_k^{-1}(z_1)\mathbf{e}_{\ell_2} \mathrm{E}_{k-1}\mathbf{e}_{\ell_2}' \mathbf{D}_k^{-1}(z_2)\mathbf{e}_{\ell_1}|^2 \right)^{1/2}$$
$$= O(n^{-1/4}). \tag{9.3.16}$$

The inequalities above show that the matrices $\mathbf{C}_k$, $\mathbf{E}_k$, and $\mathbf{F}_k$ are negligible. Now, let us evaluate the contributive components. First, for any $k < \ell \leq n$, by Lemma 9.9, we have

$$\mathbf{e}_\ell'(-\mathbf{I}_{n-1})\mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_\ell \xrightarrow{L_2} -s_2. \tag{9.3.17}$$

Next, let us estimate $\sum_{\ell_2>k} \mathbf{e}_\ell' \mathbf{A}_k(z_1)\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_2}' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_\ell$. We claim that

$$\frac{1}{\sqrt{n}} \sum_{j,\ell_2>k} x_{\ell j}\mathbf{e}_j' \mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}$$
$$\times \mathbf{e}_{\ell_2}' \mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_\ell \xrightarrow{L_2} 0. \tag{9.3.18}$$

To prove this, we consider its squared terms first. We have

$$\frac{1}{n}\sum_{j>k}\mathrm{E}\left|x_{\ell j}\sum_{\ell_2>k}\mathbf{e}_j'\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_2}'\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_\ell\right|^2$$
$$=\frac{1}{n}\sum_{j>k}\mathrm{E}\left|\sum_{\ell_2>k}\mathbf{e}_j'\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_2}'\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_\ell\right|^2.$$

If we replace $\mathbf{W}_{k\ell j}$ by $\mathbf{W}_k$ on the right-hand side of the equality above, then it becomes

$$\frac{1}{n}\mathrm{E}\mathbf{e}_\ell'\mathrm{E}_{k-1}\mathbf{H}^*\mathrm{E}_{k-1}\mathbf{H}\mathbf{e}_\ell=O(n^{-1}).$$

To consider the difference caused by this replacement, we apply (9.3.10) to both $\mathbf{D}_{k\ell j}^{-1}(z_1)$ and $\mathbf{D}_{k\ell j}^{-1}(z_2)$. The difference will also be of the order $O(n^{-1})$. As an illustration, we give the estimation of the difference caused by the replacement in $\mathbf{D}_{k\ell j}^{-1}(z_2)$, which is

$$\frac{1}{n^2}\sum_{j>k}\mathrm{E}\Bigg|\sum_{\ell_2>k}\mathbf{e}_j'\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_2}'\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_2)$$
$$\times\delta_{\ell,j}(x_{j,\ell}\mathbf{e}_j\mathbf{e}_\ell+x_{\ell,j}\mathbf{e}_\ell\mathbf{e}_j')\mathbf{D}_k^{-1}(z_2)\mathbf{e}_\ell\Bigg|^2$$
$$\le\frac{1}{n^2v_0^6}\sum_{j>k}\mathrm{E}|x_{\ell,j}^2|=O(n^{-1}),$$

where we have used the fact that, for $t=j$ or ℓ, $|\mathbf{e}_t'\mathbf{D}_k^{-1}(z_2)\mathbf{e}_\ell|\le v_0^{-1}$ and

$$\left|\sum_{\ell_2>k}\mathbf{e}_j'\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_2}'\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_t\right|\le v_0^{-2}$$

by noting that the left-hand side of the inequality above is the (ℓ,t)-th element of the product of the matrices of the last $n-k$ rows of $\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)$ and the $n-k$ columns of $\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_2)$.

Next, let us consider the sum of cross terms, which is

$$\frac{1}{n}\sum_{j\neq j'>k}\mathrm{E}x_{\ell j}\bar{x}_{\ell j'}\sum_{\ell_2>k}\mathbf{e}_j'\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_2}'\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_\ell$$
$$\times\sum_{\ell_3>k}\mathbf{e}_\ell'\mathrm{E}_{k-1}\mathbf{D}_{k\ell j'}^{-1}(\bar{z}_2)\mathbf{e}_{\ell_3}\mathbf{e}_{\ell_3}'\mathrm{E}_{k-1}\mathbf{D}_{k\ell j'}^{-1}(\bar{z}_1)\mathbf{e}_{j'}.$$

To estimate it, we define $\mathbf{W}_{ki'j'}^{ij}=\mathbf{W}_{ki'j'}-\frac{1}{\sqrt{n}}\delta_{ij}(x_{ij}\mathbf{e}_i\mathbf{e}_j'+x_{ji}\mathbf{e}_j\mathbf{e}_i')$ and $\mathbf{D}_{ki'j'}^{ij}=(\mathbf{W}_{ki'j'}^{ij}-z\mathbf{I}_{n-1})$ for $\{i,j\}\neq\{i',j'\}$. By independence, the quantity above will be 0 if the matrix $\mathbf{W}_{k\ell j'}$ is replaced by $\mathbf{W}_{k\ell j'}^{\ell j}$. Then, by a formula similar to (9.3.10), the difference caused by this replacement of the first $\mathbf{W}_{k\ell j'}$

is controlled by

$$\begin{aligned}
&\frac{1}{n^{3/2}} \sum_{j \neq j' > k} \mathrm{E}|x_{\ell j}|^2 |x_{\ell j'}| \\
&\left| \sum_{\ell_2 > k} \mathbf{e}_j' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j}^{-1}(z_1) \mathbf{e}_{\ell_2} \mathbf{e}_{\ell_2}' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j}^{-1}(z_2) \mathbf{e}_\ell \right| \\
&\times \left[\left| \sum_{\ell_3 > k} \mathbf{e}_\ell' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j'}^{\ell j}(\bar{z}_2)^{-1} \mathbf{e}_\ell \mathbf{e}_j \mathbf{D}_{k\ell j'}^{-1}(\bar{z}_2) \mathbf{e}_{\ell_3} \mathbf{e}_{\ell_3}' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j'}^{-1}(\bar{z}_1) \mathbf{e}_{j'} \right| \right. \\
&\left. + \left| \sum_{\ell_3 > k} \mathbf{e}_\ell' (\mathbf{D}_{k\ell j'}^{\ell j}(\bar{z}_2)^{-1} \mathbf{e}_j \mathbf{e}_\ell \mathbf{D}_{k\ell j'}^{-1}(\bar{z}_2) \mathbf{e}_{\ell_3} \mathbf{e}_{\ell_3}' \mathbf{D}_{k\ell j'}^{-1}(\bar{z}_1) \mathbf{e}_{j'} \right| \right] \\
= {} & O(n^{-1/2}).
\end{aligned}$$

Here, the last estimation follows from Hölder's inequality. The mathematical treatment for the two terms is similar. As an illustration of their treatment, the first term is bounded by

$$\begin{aligned}
&\frac{1}{n^{3/2}} \left(\sum_{j \neq j' > k} \mathrm{E}|x_{\ell j}|^4 \left| \sum_{\ell_2 > k} \mathbf{e}_j' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j}^{-1}(z_1) \mathbf{e}_{\ell_2} \mathbf{e}_{\ell_2}' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j}^{-1}(z_2) \mathbf{e}_\ell \right|^2 \right. \\
&\times \sum_{j \neq j' > k} \mathrm{E}|x_{\ell j'}|^2 \left| \sum_{\ell_3 > k} \mathbf{e}_\ell' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j'}^{\ell j}(\bar{z}_2)^{-1} \mathbf{e}_\ell \mathbf{e}_j \mathbf{D}_{k\ell j'}^{-1}(\bar{z}_2) \mathbf{e}_{\ell_3} \right. \\
&\left. \left. \mathbf{e}_{\ell_3}' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j'}^{-1}(\bar{z}_1) \mathbf{e}_{j'} \right|^2 \right)^{1/2} \\
\leq {} & \frac{C}{n^{3/2} v_0} \left(\sum_{j \neq j' > k} \mathrm{E} \left| \sum_{\ell_2 > k} \mathbf{e}_j' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j}^{-1}(z_1) \mathbf{e}_{\ell_2} \mathbf{e}_{\ell_2}' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j}^{-1}(z_2) \mathbf{e}_\ell \right|^2 \right. \\
&\left. \times \sum_{j \neq j' > k} \mathrm{E} \left| \sum_{\ell_3 > k} \mathbf{e}_j \mathbf{D}_{k\ell j'}^{-1}(\bar{z}_2) \mathbf{e}_{\ell_3} \mathbf{e}_{\ell_3}' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j'}^{-1}(\bar{z}_1) \mathbf{e}_{j'} \right|^2 \right)^{1/2} \\
= {} & O(n^{-1/2}).
\end{aligned}$$

Here, for the first factor in the brackets, note that by (9.3.10) we have

$$\begin{aligned}
&\sum_{\substack{j \neq j' > k \\ j' \text{ fixed}}} \mathrm{E} \left| \sum_{\ell_2 > k} \mathbf{e}_j' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j}^{-1}(z_1) \mathbf{e}_{\ell_2} \mathbf{e}_{\ell_2}' \mathrm{E}_{k-1} \mathbf{D}_{k\ell j}^{-1}(z_2) \mathbf{e}_\ell \right|^2 \\
= {} & \sum_{\substack{j \neq j' > k \\ j' \text{ fixed}}} \mathrm{E} \left| \sum_{\ell_2 > k} \mathbf{e}_j' \mathrm{E}_{k-1} \mathbf{D}_k^{-1}(z_1) \mathbf{e}_{\ell_2} \mathbf{e}_{\ell_2}' \mathrm{E}_{k-1} \mathbf{D}_k^{-1}(z_2) \mathbf{e}_\ell \right|^2 + O(1)
\end{aligned}$$

$$= O(1).$$

Hence the order of the first factor is $O(n)$. The second factor can be shown to have the order $O(n)$ by a similar approach.

Now, by (9.3.18), we have

$$\begin{aligned}
&\sum_{\ell_2>k} \mathbf{e}'_\ell \mathbf{A}_k(z_1)\mathbf{e}_{\ell_2}\mathbf{e}'_{\ell_2}\mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_\ell \\
&= \frac{1}{n^{1/2}} \sum_{j,\ell_2>k} x_{\ell j}\mathbf{e}'_j \mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}'_{\ell_2} \\
&\quad \times \mathrm{E}_{k-1}[\mathbf{D}_k^{-1}(z_2) - \mathbf{D}_{k\ell j}^{-1}(z_2)]\mathbf{e}_\ell + o_{L_2}(1) \\
&= -\frac{1}{n} \sum_{j,\ell_2>k} \delta_{\ell j} x_{\ell j}^2 \mathbf{e}'_j \mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2} \\
&\quad \times \mathbf{e}'_{\ell_2}\mathrm{E}_{k-1}[\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_\ell\mathbf{e}'_j\mathbf{D}_k^{-1}(z_2)]\mathbf{e}_\ell \\
&\quad -\frac{1}{n} \sum_{j,\ell_2>k} \delta_{\ell j}|x_{\ell j}^2|\mathbf{e}'_j \mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}'_{\ell_2} \\
&\quad \times \mathrm{E}_{k-1}[\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_j\mathbf{e}'_\ell\mathbf{D}_k^{-1}(z_2)]\mathbf{e}_\ell + o_{L_2}(1). \qquad (9.3.19)
\end{aligned}$$

Furthermore, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\mathrm{E}\left|\sum_{j,\ell_2>k} \delta_{\ell j} x_{\ell j}^2 \mathbf{e}'_j \mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}'_{\ell_2}\mathrm{E}_{k-1}[\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_\ell\mathbf{e}'_j\mathbf{D}_k^{-1}(z_2)]\mathbf{e}_\ell\right|^2 \\
&\le \left(\sum_{j>k}\mathrm{E}\left|x_{\ell j}^4\right|\left|\sum_{\ell_2>k}\mathbf{e}'_j\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}'_{\ell_2}\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_\ell\right|^2\right. \\
&\quad \left.\times \sum_{j>k}\mathrm{E}|\mathbf{e}'_j\mathbf{D}_k^{-1}(z_2)\mathbf{e}_\ell|^2\right)^{1/2} \\
&\le \left(v_0^{-2}\sum_{j>k}\mathrm{E}|\mathbf{e}'_j\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_\ell|^2\right)^{1/2} \\
&= O(1).
\end{aligned}$$

Therefore, we only need to consider the second term in (9.3.19). By Lemma 9.9, it follows that

$$\begin{aligned}
&\sum_{\ell_2>k} \mathbf{e}'_\ell \mathbf{A}_k(z_1)\mathbf{e}_{\ell_2}\mathbf{e}'_{\ell_2}\mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_\ell \\
&= -\frac{s_2}{n} \sum_{j,\ell_2>k} \delta_{\ell j}|x_{\ell j}^2|\mathbf{e}'_j \mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}'_{\ell_2}
\end{aligned}$$

$$\times \mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_j + o_{L_2}(1).$$

We claim that

$$\sum_{\ell_2>k} \mathbf{e}_\ell' \mathbf{A}_k(z_1)\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_2}' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_\ell$$
$$= -\frac{s_2}{n} \sum_{j,\ell_2>k} \mathbf{e}_j' \mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_2}' \mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_j + o_{L_2}(1). \tag{9.3.20}$$

Obviously, (9.3.20) is a consequence of

$$\frac{1}{n}\sum_{j,\ell_2>k} \mathrm{E}|\delta_{\ell j}[|x_{\ell j}^2| - 1]\mathbf{e}_j' \mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_1)\mathbf{e}_{\ell_2}\mathbf{e}_{\ell_2}' \mathrm{E}_{k-1}\mathbf{D}_{k\ell j}^{-1}(z_2)\mathbf{e}_j|^2$$
$$= o_{L_2}(1).$$

Noticing that the mean of the left-hand side is 0, it then can be proven in the same way as in the proof of (9.3.18). The details are left to the reader as an exercise.

We trivially have, for any $k < \ell \le n$,

$$\sum_{\ell,\ell_1>k} \mathbf{e}_\ell' \mathbf{B}_k(z_1)\mathbf{e}_{\ell_1}\mathbf{e}_{\ell_1}' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_\ell$$
$$= -s_1 \sum_{\ell,\ell_1>k} \mathbf{e}_\ell' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_1)\mathbf{e}_{\ell_1}\mathbf{e}_{\ell_1}' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_\ell + o_{L_2}(1). \tag{9.3.21}$$

Collecting the estimates above from (9.3.14) to (9.3.21), we find that

$$\frac{z_1+s_1}{n}\sum_{\ell,\ell_1>k} \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_1)\mathbf{e}_{\ell_1}\mathbf{e}_{\ell_1}' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)$$
$$= -\left(1-\frac{k}{n}\right)s_2 - \frac{1}{n}\left(1-\frac{k}{n}\right)s_2 \sum_{j,\ell_1>k} \mathbf{e}_j' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_1)\mathbf{e}_{\ell_1}$$
$$\times \mathbf{e}_{\ell_1}' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_j + o_p(1),$$

which, together with the fact that $z_1 + s_1 = -1/s_1$, implies that

$$\frac{1}{n}\sum_{j,\ell>k} \mathbf{e}_\ell' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_1)\mathbf{e}_j\mathbf{e}_j' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_\ell$$
$$= \left(1-\frac{k}{n}\right)s_1s_2 + \frac{1}{n}\left(1-\frac{k}{n}\right)s_1s_2 \sum_{j,\ell>k} \mathbf{e}_j' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_1)\mathbf{e}_\ell$$
$$\times \mathbf{e}_\ell' \mathrm{E}_{k-1}\mathbf{D}_k^{-1}(z_2)\mathbf{e}_j + o_p(1)$$

$$= \frac{(1-\frac{k}{n})s_1 s_2}{1-(1-\frac{k}{n})s_1 s_2} + o_p(1). \tag{9.3.22}$$

Therefore,

$$\begin{aligned}\lim_n S_1 &= \lim_n \frac{1}{n^2}\sum_{k=1}^n \sum_{i,j>k} b_{ijk}^{(1)} b_{jik}^{(2)} \\ &= \lim_n \frac{1}{n}\sum_{k=1}^n \frac{(1-\frac{k}{n})s_1 s_2}{1-(1-\frac{k}{n})s_1 s_2} \\ &= \int_0^1 \frac{t s_1 s_2}{1-t s_1 s_2} dt \\ &= -1 - \frac{1}{s_1 s_2}\log(1-s_1 s_2).\end{aligned}$$

Finally, $\widetilde{\Gamma}_n(z_1, z_2)$ converges in probability to

$$\widetilde{\Gamma}(z_1, z_2) = \sigma^2 - \kappa + \frac{1}{2}\beta s_1 s_2 - \kappa (s_1 s_2)^{-1}\log(1-s_1 s_2).$$

The proof of Lemma 9.6 is then complete.

9.3.3 Completion of the Proof of (9.2.13) for $j = l, r$

Since we have proved (9.2.29), to complete the proof of (9.2.13) we only need to show that

$$\lim_{v_1 \downarrow 0} \limsup_{n\to\infty} \int_{\mathcal{C}_j} \mathrm{E}\left|M_n(z) - \mathrm{E}M_n(z)\right|^2 dz = 0. \tag{9.3.23}$$

Using the notation defined in the last section, we have

$$\mathrm{E}|M_n - \mathrm{E}M_n|^2 \le K \sum_{k=1}^n [\mathrm{E}|a_{k1}|^2 + \mathrm{E}|a_{k2}|^2 + \mathrm{E}|a_{k3}|^2 + \mathrm{E}|d_k|^2].$$

By Lemma 9.1 and (9.2.17),

$$\sup_{z\in\mathcal{C}_n} \mathrm{E}|d_k|^2 \le \sup_{z\in\mathcal{C}_n} K n^{-1}\mathrm{E}|\tilde{b}_k|^2 \|\mathbf{D}_k^{-1}\|^2 \le K/n,$$

where we have used the fact that $|\tilde{b}_k| < 1$, which can be proven along the same lines as for (8.1.19).

Similarly,

$$\sup_{z\in\mathcal{C}_n} \mathrm{E}|a_{k1}|^2 \le \sup_{z\in\mathcal{C}_n} Kn^{-1}\mathrm{E}|\tilde{b}_k|^2\|\mathbf{D}_k^{-1}\|^4\left|1+\frac{1}{n}\mathrm{tr}\mathbf{D}_k^{-1}\right|^2$$
$$\le K/n.$$

By (9.3.3) and (9.3.4), we obtain

$$\begin{aligned}
&\lim_{v_1\downarrow 0}\limsup_{n\to\infty}\int_{\mathcal{C}_j}\mathrm{E}\left|M_n(z)-\mathrm{E}M_n(z)\right|^2 dz\\
&\le \lim_{v_1\downarrow 0}\limsup_{n\to\infty} K\int_{\mathcal{C}_j}\sum_{k=1}^n[\mathrm{E}|a_{k1}|^2+\mathrm{E}|d_k|^2]dz\\
&\le \lim_{v_1\downarrow 0} Kv_1 = 0.
\end{aligned}$$

The proof is complete.

9.3.4 Tightness of the Process $M_n(z) - \mathrm{E}M_n(z)$

It is enough to establish the following Hölder condition: for some positive constant K and $z_1, z_2 \in \mathbb{C}_0$,

$$\mathrm{E}|M_n(z_1)-M_n(z_2)-\mathrm{E}(M_n(z_1)-M_n(z_2))|^2 \le K|z_1-z_2|^2. \tag{9.3.24}$$

Recalling the martingale decomposition given in Section 2.3, we have

$$\begin{aligned}
&\mathrm{E}|M_n(z_1)-M_n(z_2)-\mathrm{E}(M_n(z_1)-M_n(z_2))|^2\\
&=\sum_{k=1}^n \mathrm{E}|\gamma_k(z_1)-\gamma_k(z_2)|^2,
\end{aligned}$$

where

$$\begin{aligned}
\gamma_k(z) &= (\mathrm{E}_k-\mathrm{E}_{k-1})\sigma_k(z),\\
\sigma_k(z) &= \beta_k(z)\left(1+\frac{1}{n}\gamma_k^*\mathbf{D}_k^{-2}\boldsymbol{\alpha}_k\right),\\
\beta_k(z) &= -\frac{1}{\frac{1}{\sqrt{n}}x_{kk}-z-\frac{1}{n}\boldsymbol{\alpha}_k\mathbf{D}_k^{-1}\boldsymbol{\alpha}_k}.
\end{aligned}$$

Using the notation defined in (9.3.2), we decompose $\gamma_k(z_1)-\gamma_k(z_2)$ as

$$\begin{aligned}
&(\mathrm{E}_k-\mathrm{E}_{k-1})\Big[\beta_k(z_1)(h_k(z_1)-h_k(z_2))\\
&+\beta_k(z_1)\sigma_k(z_2)(g_k(z_1)-g_k(z_2))\\
&+\frac{1}{n}\beta_k(z_1)\tilde{b}_n(z_1)[\mathrm{tr}\mathbf{D}_k^{-2}(z_1)-\mathbf{D}_k^{-2}(z_2)]g_k(z_1)
\end{aligned}$$

$$+\frac{1}{n}\tilde{b}_k(z_1)\tilde{b}_k(z_2)\mathrm{tr}[\mathbf{D}_k^{-1}(z_1)-\mathbf{D}_k^{-1}(z_2)]h_k(z_2)$$
$$+\frac{1}{n}\beta_k(z_1)\tilde{b}_k(z_1)\sigma_k(z_2)\mathrm{tr}[\mathbf{D}_k^{-1}(z_1)-\mathbf{D}_k^{-1}(z_2)]g_k(z_1)$$
$$\left.+\frac{1}{n}\tilde{b}_k(z_1)\tilde{b}_k(z_2)\sigma_k(z_2)\mathrm{tr}[\mathbf{D}_k^{-1}(z_1)-\mathbf{D}_k^{-1}(z_2)]g_k(z_2)\right].$$

Since $\beta_k^{-1}(z)\geq v_0$ and $|\gamma_k(z)|\leq v_0^{-1}$, we have

$$\begin{aligned}
&\sum_{k=1}^{n}\mathrm{E}|\beta_k(z_1)(h_k(z_1)-h_k(z_2))|^2\\
&\leq v_0^{-2}\sum_{k=1}^{n}\mathrm{E}|h_k(z_1)-h_k(z_2)|^2\\
&\leq \frac{C}{n^2v_0^2}\sum_{k=1}^{n}\mathrm{E}\mathrm{tr}[\mathbf{D}_k^{-2}(z_1)-\mathbf{D}_k^{-2}(z_2)][\mathbf{D}_k^{-2}(\bar{z}_1)-\mathbf{D}_k^{-2}(\bar{z}_2)]\\
&\leq \frac{4C|z_1-z_2|^2}{v_0^8}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\sum_{k=1}^{n}\mathrm{E}|\beta_k(z_1)\sigma_k(z_2)(g_k(z_1)-g_k(z_2))|^2\\
&\leq v_0^{-4}\sum_{k=1}^{n}\mathrm{E}|g_k(z_1)-g_k(z_2)|^2\\
&\leq \frac{4C|z_1-z_2|^2}{v_0^8}.
\end{aligned}$$

For the other four terms, the similar estimates follow trivially from the fact that

$$\mathrm{E}|g_k(z)|^2\leq C/n \quad \text{and} \quad \mathrm{E}|h_k(z)|^2\leq C/n.$$

Hence (9.3.24) is proved and the tightness of the process $M_n-\mathrm{E}M_n$ holds.

9.4 Computation of the Mean and Covariance Function of $G(f)$

9.4.1 Mean Function

Let $\mathcal{C}$ be a contour as defined in Subsection 9.2.1. By (9.2.10) and Lemma 9.5, we have

$$\begin{aligned}
&\mathrm{E}(G_n(f)) = -\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\mathrm{E}M_n(z)dz \\
\to\ &\mathrm{E}(G(f)) = -\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\mathrm{E}M(z)dz \\
&= -\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)[1+s'(z)]s^3(z)\left[\sigma^2-1+(\kappa-1)s'(z)+\beta s^2(z)\right]dz.
\end{aligned}$$

Select $\rho < 1$ but so close to 1 that the contour

$$\mathcal{C}' = \{z = -(\rho e^{i\theta} + \rho^{-1}e^{-i\theta}) : 0 \le \theta < 2\pi\}$$

is completely contained in the analytic region of f. Note that when z runs a cycle along $\mathcal{C}'$ anticlockwise, s runs a cycle along the circle $|s| = \rho$ anticlockwise because $z = -(s + s^{-1})$[1] By Cauchy's theorem, the integral along $\mathcal{C}$ above equals the integral along $\mathcal{C}'$. Thus, by changing variable z to s and noting that $s' = s^2/(1-s^2)$, we obtain

$$\begin{aligned}
&\mathrm{E}(G(f)) \\
&= -\frac{1}{2\pi i}\oint_{|s|=\rho} f(-s-s^{-1})s\left[\sigma^2-1+(\kappa-1)\frac{s^2}{1-s^2}+\beta s^2\right]ds.
\end{aligned}$$

By setting $s = -e^{i\theta}$ and then $t = \cos\theta$, using $T_k(\cos\theta) = \cos(k\theta)$,

$$\begin{aligned}
&-\frac{1}{2\pi i}\oint_{|s|=1} f(-s-s^{-1})s\left[\sigma^2-1+(\kappa-1)\frac{s^2}{1-s^2}+\beta s^2\right]ds \\
&= -\frac{1}{2\pi}\int_{-\pi}^{\pi} f(2\cos\theta)\left[(\sigma^2-1)e^{2i\theta}+(\kappa-1)\frac{e^{4i\theta}}{1-e^{2i\theta}}+\beta e^{4i\theta}\right]d\theta \\
&= -\frac{1}{\pi}\int_0^{\pi} f(2\cos\theta)\Big[(\sigma^2-1)\cos 2\theta \\
&\qquad -\frac{1}{2}(\kappa-1)(1+2\cos 2\theta)+\beta\cos 4\theta\Big]d\theta \\
&= \frac{1}{\pi}\int_{-1}^{1} f(2t)\left[-\frac{1}{2}(\kappa-1)+(\sigma^2-\kappa)T_2(t)+\beta T_4(t)\right]\frac{1}{\sqrt{1-t^2}}dt \\
&= -\frac{1}{2}(\kappa-1)\tau_0(f)+(\sigma^2-\kappa)\tau_2(f)+\beta\tau_4(f).
\end{aligned}$$

Let us evaluate the difference

$$\frac{1}{2\pi i}\left[\oint_{|s|=1} - \oint_{|s|=\rho}\right] f(-s-s^{-1})s\left[\sigma^2-1+(\kappa-1)\frac{s^2}{1-s^2}+\beta s^2\right]ds.$$

[1] The reason for choosing $|s| = \rho < 1$ is due to the fact that the mode of the Stieltjes transform of the semicircular law is less than 1; see (8.1.11).

Note that the integrand has two poles on the circle $|s| = 1$ with residuals $-\frac{1}{2}f(\pm 2)$ at points $s = \mp 1$. By contour integration, we have

$$\frac{1}{2\pi i}\left[\oint_{|s|=1} - \oint_{|s|=\rho}\right] f(-s-s^{-1})s\left[\sigma^2 - 1 + (\kappa-1)\frac{s^2}{1-s^2} + \beta s^2\right]ds$$
$$= \frac{\kappa-1}{4}(f(2)+f(-2)).$$

Putting together these two results gives the formula (9.2.4) for $\mathrm{E}[G(f)]$.

9.4.2 Covariance Function

Let $\mathcal{C}_j$, $j = 1, 2$, be two disjoint contours with vertices $\pm(2+\varepsilon_j) \pm iv_j$. The positive values of ε_j and v_j are chosen sufficiently small so that the two contours are contained in $\mathcal{U}$. By (9.2.10) and Theorem 9.4, we have

$$\begin{aligned}
&\mathrm{Cov}(G_n(f), G_n(g)) \\
&= -\frac{1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)g(z_2)\mathrm{Cov}(M_n(z_1), M_n(z_2))dz_1dz_2 \\
&= -\frac{1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)g(z_2)\Gamma_n(z_1,z_2)dz_1dz_2 + o(1) \\
&\longrightarrow c(f,g) = -\frac{1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)g(z_2)\Gamma(z_1,z_2)dz_1dz_2,
\end{aligned}$$

where $\Gamma(z_1, z_2)$ is given in (9.3.6).

By the proof of Lemma 9.6, we have

$$\Gamma(z_1,z_2) = \frac{\partial^2}{\partial z_1 \partial z_2} s(z_1)s(z_2)\widetilde{\Gamma}(z_1,z_2).$$

Integrating by parts, we obtain

$$\begin{aligned}
c(f,g) &= -\frac{1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f'(z_1)g'(z_2)s(z_1)s(z_2)\widetilde{\Gamma}(z_1,z_2)dz_1dz_2 \\
&= -\frac{1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} A(z_1,z_2)dz_1dz_2,
\end{aligned}$$

where

$$A(z_1,z_2) = f'(z_1)g'(z_2)\Big[s(z_1)s(z_2)(\sigma^2-\kappa) + \frac{1}{2}\beta s^2(z_1)s^2(z_2)$$

$$-\kappa\log(1-s(z_1)s(z_2))\Bigg].$$

Let $v_j \to 0$ first and then $\varepsilon_j \to 0$. It is easy to show that the integral along the vertical edges of the two contours tends to 0 when $v_j \to 0$. Therefore, it follows that

$$\begin{aligned}&c(f,g)\\ &= -\frac{1}{4\pi^2}\int_{-2}^{2}\int_{-2}^{2}[A(t_1^-,t_2^-)-A(t_1^-,t_2^+)-A(t_1^+,t_2^-)+A(t_1^+,t_2^+)]dt_1dt_2,\end{aligned}$$

where $t_j^{\pm} := t_j \pm i0$. Since f' and g' are continuous in $\mathcal{U}$, we have $f'(t_1^{\pm}) = f'(t_1)$ and $g'(t_2^{\pm}) = g'(t_2)$. Recalling that $s(t \pm i0) = \frac{1}{2}(-t \pm i\sqrt{4-t^2})$, we have

$$\begin{aligned}&f'(t_1)g'(t_2)[s(t_1^-)s(t_2^-)-s(t_1^+)s(t_2^-)-s(t_1^-)s(t_2^+)+s(t_1^+)s(t_2^+)]\\ &= -f'(t_1)g'(t_2)\sqrt{4-t_1^2}\sqrt{4-t_2^2},\\ &f'(t_1)g'(t_2)[s^2(t_1^-)s^2(t_2^-)-s^2(t_1^+)s^2(t_2^-)\\ &\qquad -s^2(t_1^-)s^2(t_2^+)+s^2(t_1^+)s^2(t_2^+)]\\ &= -f'(t_1)g'(t_2)t_1t_2\sqrt{4-t_1^2}\sqrt{4-t_2^2},\\ &f'(t_1)g'(t_2)[\log(1-s(t_1^-)s(t_2^-))-\log(1-s(t_1^+)s(t_2^-))\\ &\qquad -\log(1-s(t_1^-)s(t_2^+))+\log(1-s(t_1^+)s(t_2^+))]\\ &= f'(t_1)g'(t_2)\log\left|\frac{1-s(t_1^-)s(t_2^-)}{1-s(t_1^-)s(t_2^+)}\right|^2\\ &= -f'(t_1)g'(t_2)\log\left(\frac{4-t_1t_2-\sqrt{(4-t_1^2)(4-t_2^2)}}{4-t_1t_2+\sqrt{(4-t_1^2)(4-t_2^2)}}\right).\end{aligned}$$

Therefore, we have formula (9.2.6).

To derive the first representation of the covariance (i.e., formula (9.2.5)), let $\rho_1 < \rho_2 < 1$ and define contours $\mathcal{C}_j'$ as in the last subsection. Then,

$$\begin{aligned}c(f,g) &= -\frac{1}{4\pi^2}\oint_{\mathcal{C}_1'}\oint_{\mathcal{C}_2'} f(z_1)g(z_2)\Gamma(z_1,z_2)dz_1dz_2\\ &= -\frac{1}{4\pi^2}\oint_{|s_1|=\rho_1}\oint_{|s_2|=\rho_2} f(-s_1-s_1^{-1})g(-s_2-s_2^{-1})\\ &\qquad\times\left(\sigma^2-\kappa+2\beta s_1s_2+\frac{\kappa}{(1-s_1s_2)^2}\right)ds_1ds_2.\end{aligned}$$

By Cauchy's theorem, we may change $\rho_2 = 1$ without affecting the value of the integral. Rewriting $\rho_1 = \rho$, expanding the fraction as a Taylor series, and then making variable changes $s_1 = -\rho e^{i\theta_1}$ and $s_2 = -e^{i\theta_2}$, we obtain

$$\begin{aligned} &c(f,g) \\ &= \frac{1}{4\pi^2} \int_{[-\pi,\pi]^2} f(\rho e^{i\theta_1} + \rho^{-1} e^{-i\theta_1}) g(2\cos\theta_2) \Big[\sigma^2 \rho e^{i(\theta_1+\theta_2)} \\ &\quad +2(\beta+1)\rho^2 e^{i2(\theta_1+\theta_2)} + \kappa \sum_{k=3}^{\infty} k\rho^k e^{ik(\theta_1+\theta_2)} \Big] d\theta_1 d\theta_2 \\ &= \sigma^2 \rho \tau_1(f,\rho)\tau_1(g) + 2(\beta+1)\rho^2 \tau_2(f,\rho)\tau_2(g) \\ &\quad + \kappa \sum_{k=3}^{\infty} k\rho^k \tau_k(f,\rho)\tau_k(g), \end{aligned}$$

where $\tau_k(f,\rho) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\rho e^{i\theta} + \rho^{-1}e^{-i\theta}) e^{ik\theta} d\theta$. By integration by parts, for $k \geq 3$ we have

$$\begin{aligned} \tau_k(f,\rho) &= \frac{\rho^{-1}}{k}\tau_{k-1}(f',\rho) - \frac{\rho}{k}\tau_{k+1}(f',\rho) \\ &= \frac{\rho^2}{k(k+1)}\tau_{k+2}(f'',\rho) - \frac{2}{k^2-1}\tau_k(f'',\rho) + \frac{\rho^{-2}}{k(k-1)}\tau_{k-2}(f'',\rho). \end{aligned}$$

Since f'' is uniformly bounded in $\mathcal{U}$, we have $|\tau_k(f,\rho)| \leq K/k(k-1)$ uniformly for all ρ close to 1. Then (9.2.5) follows from the dominated convergence theorem and letting $\rho \to 1$ under the summation.

9.5 Application to Linear Spectral Statistics and Related Results

First note that $W_n/(2\sqrt{n})$ is a scaled Wigner matrix in the sense that the limit law is the scaled Wigner semicircular law $\frac{2}{\pi}\sqrt{1-x^2}dx$ on the interval $[-1,1]$. To deal with this scaling, we define, for any function f, its scaled copy $\tilde{f}$ by the relation $f(2x) = \tilde{f}(x)$ for all x.

9.5.1 Tchebychev Polynomials

Consider first a Tchebychev polynomial T_k with $k \geq 1$, and define ϕ_k such that $\tilde{\phi}_k = T_k$. Set $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ elsewhere. Using the orthogonality property

$$\frac{1}{\pi}\int_{-1}^{1} T_i(t)T_j(t)\frac{1}{\sqrt{1-t^2}}dt = \begin{cases} \delta_{ij}, & \text{if } i=0 \\ \frac{1}{2}\delta_{ij}, & \text{elsewhere,} \end{cases}$$

it is easily seen that $\tau_\ell(\phi_k) = \frac{1}{2}\delta_{k\ell}$ for any integer $\ell \geq 0$. Thus, by (9.2.4), we have for the mean

$$\begin{aligned} m_k := \mathrm{E}[G(\phi_k)] &= \frac{\kappa-1}{4}(T_k(1)+T_k(-1)) + \frac{1}{2}(\sigma^2-\kappa)\delta_{k2} + \frac{1}{2}\beta\delta_{k4} \\ &= \frac{1}{2}\left[(\kappa-1)e(k) + (\sigma^2-\kappa)\delta_{k2} + \beta\delta_{k4}\right], \end{aligned} \tag{9.5.1}$$

with $e(k) = 1$ if k is even and $e(k) = 0$ elsewhere.

For two Tchebychev polynomials T_k and T_ℓ, by (9.2.5) the asymptotic covariance between $G_n(\phi_k)$ and $G_n(\phi_\ell)$ equals 0 for $k \neq \ell$, and for $k = \ell$,

$$\Sigma_{\ell\ell} = \left(\frac{1}{2}\right)^2 \left[(\sigma^2-\kappa)\delta_{\ell 1} + (4\beta+2\kappa)\delta_{\ell 2} + \kappa\ell\right]. \tag{9.5.2}$$

An application of Theorem 9.2 readily yields the following corollary.

Corollary 9.7. *Assume conditions [M1]–[M3] hold. Let $T_1, \cdots, T_p$ be p first Tchebychev polynomials and define the ϕ_k's such that $\tilde{\phi}_k = T_k$. Then the vector $[G_n(\phi_1), \cdots, G_n(\phi_p)]$ converges in distribution to a Gaussian vector with mean $w_p = (m_k)$ and a diagonal covariance matrix $D_p = (\Sigma_{kk})$ with their elements defined in equations (9.5.1) and (9.5.2), respectively.*

In particular, these Tchebychev polynomial statistics are asymptotically independent. Consider now the Gaussian case. For the GUE ensemble, we have $\kappa = \sigma^2 = 1$ and $\beta = 0$. Then $m_k = 0$ and $\Sigma_{kk} = k(\frac{1}{2})^2$. As for the GOE ensemble, since $\kappa = \sigma^2 = 2$ and $\beta = 0$, we get $m_k = \frac{1}{2}e(k)$ and $\Sigma_{kk} = 2k(\frac{1}{2})^2$. Therefore, with Corollary 9.7 we have recovered the CLT established by Johansson [165] for linear spectral statistics of Gaussian ensembles (see Theorem 2.4 and Corollary 2.8 there).

9.6 Technical Lemmas

With the notation defined in the previous sections, we prove the following two lemmas that were used in the proofs in previous sections.

Lemma 9.8. *For any positive constants ν and t, when $z \in \mathbb{C}_0$, all of the following probabilities have order $o(n^{-t})$:*

$$P(|\varepsilon_k| \geq \nu), \quad P(|g_k| \geq \nu), \quad P(|h_k| \geq \nu).$$

When $z \notin \mathbb{C}_0$ but $|\Re(z)| \geq a$, the same estimates remain true.

Proof. The estimates for $P(|g_k| \geq \nu)$ and $P(|h_k| \geq \nu)$ directly follow from Lemma 9.1 and the Chebyshev inequality.

Recalling the definition of ε_k, we have

$$\begin{aligned}|\varepsilon_k| &= \left| n^{-1/2} x_{kk} - \frac{1}{n} \boldsymbol{\alpha}_k^* (\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1} \boldsymbol{\alpha}_k + \mathrm{E}s_n(z) \right| \\ &\leq |g_k(z)| + n^{-1} |\mathrm{tr}(\mathbf{W}_k - z\mathbf{I}_{n-1})^{-1} - (\mathbf{W} - z\mathbf{I}_n)^{-1}| \\ &\qquad + |s_n(z) - \mathrm{E}s_n(z)|.\end{aligned}$$

Noting that the second term is less than $1/nv_0$, the estimate for $P(|\varepsilon_k| \geq \nu)$ follows from Lemmas 9.1 and 8.7.

The proof of the lemma is complete.

Lemma 9.9. *Suppose $v_0 > 0$ is a fixed constant. Then, for any $z \in \mathbb{C}_0$, we have*

$$\sup_{z \in \mathcal{C}_n} \max_{i,j,k,\ell} \mathrm{E}|\mathrm{E}_k \mathbf{e}_\ell' \mathbf{D}_{kij}^{-1} \mathbf{e}_\ell - s(z)|^2 \to 0,$$

where the maximum is taken over all k, $i, j \neq k$, and all ℓ.

Proof. Recall identity (9.3.10). Since $|x_{ij}| \leq \eta_n \sqrt{n}$, by (9.2.17) we have

$$\sup_{z \in \mathcal{C}_n} \mathrm{E}|\mathbf{e}_\ell'(\mathbf{D}_{kij} - \mathbf{D}_k)\mathbf{e}_\ell|^2 \leq K\eta_n \sup_{z \in \mathcal{C}_n} \mathrm{E}\|\mathbf{D}_{kij}\|^{-1}\|^2 \mathbf{D}_k^{-1}\|^2 \to 0.$$

Again, by (9.2.17),

$$\mathrm{E}|\mathbf{e}_\ell'(\mathbf{D}_k^{-1} - \mathbf{D}^{-1})\mathbf{e}_\ell|^2 \to 0.$$

Moreover, by definition,

$$\begin{aligned}\mathbf{e}_\ell' \mathbf{D}^{-1} \mathbf{e}_\ell &= \frac{1}{n^{-1/2} x_{\ell\ell} - z - n^{-1} \boldsymbol{\alpha}_\ell^* \mathbf{D}_\ell^{-1} \boldsymbol{\alpha}_\ell} \\ &= \frac{1}{-z - s(z)} + \frac{-s(z) - [n^{-1/2} x_{\ell\ell} - n^{-1} \boldsymbol{\alpha}_\ell^* \mathbf{D}_\ell^{-1} \boldsymbol{\alpha}_\ell]}{(n^{-1/2} x_{\ell\ell} - z - n^{-1} \boldsymbol{\alpha}_\ell^* \mathbf{D}_\ell^{-1} \boldsymbol{\alpha}_\ell)[-z - s(z)]} \\ &= s(z) + s(z)\beta_\ell [s(z) + n^{-1/2} x_{\ell\ell} - n^{-1} \boldsymbol{\alpha}_\ell^* \mathbf{D}_\ell^{-1} \boldsymbol{\alpha}_\ell].\end{aligned}$$

By (9.2.18), (9.2.17), and Lemma 9.1, it follows that

$$\begin{aligned}&\mathrm{E}|\mathbf{e}_\ell' \mathbf{D}^{-1} \mathbf{e}_\ell - s(z)|^2 \\ &\leq \mathrm{E}|s(z)\beta_\ell [s(z) + n^{-1/2} x_{\ell\ell} - n^{-1} \boldsymbol{\alpha}_\ell^* \mathbf{D}_\ell^{-1} \boldsymbol{\alpha}_\ell]|^2 \\ &\leq 2\mathrm{E}|s(z) + n^{-1/2} x_{\ell\ell} - n^{-1} \boldsymbol{\alpha}_\ell^* \mathbf{D}_\ell^{-1} \boldsymbol{\alpha}_\ell|^2 + o(1) \\ &\leq K\mathrm{E}|s(z) - n^{-1} \mathrm{tr}\mathbf{D}_\ell^{-1}|^2 + n^{-2} \mathrm{E}|\boldsymbol{\alpha}_\ell^* \mathbf{D}_\ell^{-1} \boldsymbol{\alpha}_\ell - \mathrm{tr}\mathbf{D}_\ell^{-1}|^2 + o(1) \\ &\leq o(1)\end{aligned}$$

uniformly for any $z \in \mathcal{C}_n$.

The lemma follows.

9.7 CLT of the LSS for Sample Covariance Matrices

In this section, we shall consider the CLT for the LSS associated with the general form of a sample covariance matrix considered in Chapter 6,

$$\mathbf{B}_n = \frac{1}{n}\mathbf{T}^{1/2}\mathbf{X}_n\mathbf{X}_n^*\mathbf{T}^{1/2}.$$

Some limiting theorems on the ESD and the spectrum separation of $\mathbf{B}_n$ have been discussed in Chapter 6. In this section, we shall consider more special properties of the LSS constructed using eigenvalues of $\mathbf{B}_n$.

It has been proven that, under certain conditions, with probability 1, the ESD of $\mathbf{B}_n$ tends to a limit $F^{y,H}$ whose Stieltjes transform is the unique solution to

$$s = \int \frac{1}{\lambda(1-y-yzs)-z}dH(\lambda)$$

in the set $\{s \in \mathbb{C}^+ : -\frac{1-y}{z} + ys \in \mathbb{C}^+\}$.

Define $\underline{\mathbf{B}}_n \equiv (1/n)\mathbf{X}_n^*\mathbf{T}_n\mathbf{X}_n$, and denote its LSD and limiting Stieltjes transform as $\underline{F}^{y,H}$ and $\underline{s} = \underline{s}(z)$. Then the equation takes on a simpler form when $F^{y,H}$ is replaced by

$$\underline{F}^{y,H} \equiv (1-y)I_{[0,\infty)} + yF^{y,H};$$

namely

$$\underline{s}(z) \equiv s_{\underline{F}^{y,H}}(z) = -\frac{1-y}{z} + ys(z)$$

has inverse

$$z = z(\underline{s}) = -\frac{1}{\underline{s}} + y\int \frac{t}{1+t\underline{s}}dH(t). \tag{9.7.1}$$

Now, let us consider the linear spectral statistics defined as

$$\mu_n(f) = \int f(x)dF^{\mathbf{B}_n}(x).$$

Theorem 9.10, presented below, shows that $\mu_n(f) - \int f(x)dF^{y_n,H_n}(x)$ has convergence rate $1/p$. Since the convergence of $y_n \to y$ and $H_n \to H$ may be very slow, the difference $p(\mu_n(f) - \int f(x)dF^{y,H}(x))$ may not have a limiting distribution. More importantly, from the point of view of statistical inference, H_n can be viewed as a description of the current population and y_n is the ratio of dimension to sample size for the current sample. The limit $\int f(x)dF^{y,H}(x)$ should be viewed as merely a mathematical convenience allowing the result to be expressed as a limit theorem. Thus we consider $p(\mu_n(f) - \int f(x)dF^{y_n,H_n}(x))$.

For notational purposes, write

$$X_n(f) = \int f(x) dG_n(x),$$

where $G_n(x) = p(F^{\mathbf{B}_n}(x) - F^{y_n, H_n}(x))$.

The main result is stated in the following theorem, which extends a result presented in Bai and Silverstein [30].

Theorem 9.10. *Assume that the X-variables satisfy the condition*

$$\frac{1}{np} \sum_{ij} \mathrm{E}|x_{ij}^4| I(|x_{ij}| \geq \sqrt{n}\eta) \to 0 \tag{9.7.2}$$

for any fixed $\eta > 0$ and that the following additional conditions hold:

(a) *For each n, $x_{ij} = x_{ij}^{(n)}$, $i \leq p$, $j \leq n$ are independent. $\mathrm{E}x_{ij} = 0$, $\mathrm{E}|x_{ij}|^2 = 1$, $\max_{i,j,n} \mathrm{E}|x_{ij}|^4 < \infty$, $p/n \to y$.*

(b) *$\mathbf{T}_n$ is $p \times p$ nonrandom Hermitian nonnegative definite with spectral norm bounded in p, with $F^{\mathbf{T}_n} \xrightarrow{\mathscr{D}} H$ a proper c.d.f.*

Let $f_1, \cdots, f_k$ be functions analytic on an open region containing the interval

$$\left[\liminf_n \lambda_{\min}^{\mathbf{T}_n} I_{(0,1)}(y)(1-\sqrt{y})^2, \limsup_n \lambda_{\max}^{\mathbf{T}_n}(1+\sqrt{y})^2\right]. \tag{9.7.3}$$

Then

(1) *the random vector*

$$(X_n(f_1), \cdots, X_n(f_k)) \tag{9.7.4}$$

forms a tight sequence in n.

(2) *If x_{ij} and $\mathbf{T}_n$ are real and $\mathrm{E}(x_{ij}^4) = 3$, then (9.7.4) converges weakly to a Gaussian vector $(X_{f_1}, \cdots, X_{f_k})$ with means*

$$\mathrm{E}X_f = -\frac{1}{2\pi i} \int_{\mathcal{C}} f(z) \frac{y \int \frac{\underline{s}(z)^3 t^2 dH(t)}{(1+t\underline{s}(z))^3}}{\left(1 - y \int \frac{\underline{s}(z)^2 t^2 dH(t)}{(1+t\underline{s}(z))^2}\right)^2} dz \tag{9.7.5}$$

and covariance function

$$\begin{aligned} &\mathrm{Cov}(X_f, X_g) \\ &= -\frac{1}{2\pi^2} \int_{\mathcal{C}_1} \int_{\mathcal{C}_2} \frac{f(z_1) g(z_2)}{(\underline{s}(z_1) - \underline{s}(z_2))^2} \underline{s}'(z_1) \underline{s}'(z_2) dz_1 dz_2 \end{aligned} \tag{9.7.6}$$

($f, g \in \{f_1, \cdots, f_k\}$). The contours in (9.7.5) and (9.7.6) (two in (9.7.6), which may be assumed to be nonoverlapping) are closed and are taken in

the positive direction in the complex plane, each enclosing the support of $F^{y,H}$.

(3) *If* x_{ij} *is complex with* $\mathrm{E}(x_{ij}^2) = 0$ *and* $\mathrm{E}(|x_{ij}|^4) = 2$, *then (2) also holds, except the means are zero and the covariance function is* $1/2$ *times the function given in (9.7.6).*

This theorem can be viewed as an extension of results obtained in Jonsson [169], where the entries of $\mathbf{X}_n$ are Gaussian, $\mathbf{T}_n = \mathbf{I}$, and $f_k = x^k$.

9.7.1 Truncation

We begin the proof of Theorem 9.10 here with the replacement of the entries of $\mathbf{X}_n$ with truncated and centralized variables. By condition (9.7.2), we may select $\eta_n \downarrow 0$ and such that

$$\frac{1}{np\eta_n^4} \sum_{ij} \mathrm{E}|x_{ij}|^4 I(|x_{ij}| \geq \eta_n \sqrt{n}) \to 0. \tag{9.7.7}$$

The convergence rate of the constants η_n can be arbitrarily slow and hence we may assume that $\eta_n n^{1/5} \to \infty$. Let $\widehat{\mathbf{B}}_n = (1/n)\mathbf{T}^{1/2}\widehat{\mathbf{X}}_n\widehat{\mathbf{X}}_n^*\mathbf{T}^{1/2}$ with $\widehat{\mathbf{X}}_n$ $p \times n$ having (i,j)-th entry $\hat{x}_{ij} = x_{ij} I_{\{|x_{ij}|<\eta_n\sqrt{n}\}}$.

We have then

$$\begin{aligned} \mathrm{P}(\mathbf{B}_n \neq \widehat{\mathbf{B}}_n) &\leq \sum_{ij} \mathrm{P}(|x_{ij}| \geq \eta_n \sqrt{n}) \\ &\leq \frac{1}{np\eta_n^{-4}} \sum_{ij} \mathrm{E}|x_{ij}|^4 I(|x_{ij}| \geq \eta_n \sqrt{n}) = o(1). \end{aligned}$$

Define $\widetilde{\mathbf{B}}_n = (1/n)\mathbf{T}^{1/2}\widetilde{\mathbf{X}}_n\widetilde{\mathbf{X}}_n^*\mathbf{T}^{1/2}$ with $\widetilde{\mathbf{X}}_n$ $p \times n$ having (i,j)-th entry $\tilde{x}_{ij} = (\widehat{x}_{ij} - \mathrm{E}\widehat{x}_{ij})/\sigma_{ij}$, where $\sigma_{ij}^2 = \mathrm{E}|\widehat{x}_{ij} - \mathrm{E}\widehat{x}_{ij}|^2$. From Theorem 5.11, we know that both $\limsup_n \lambda_{\max}^{\widehat{\mathbf{B}}_n}$ and $\limsup_n \lambda_{\max}^{\widetilde{\mathbf{B}}_n}$ are almost surely bounded by $\limsup_n \|\mathbf{T}_n\|(1+\sqrt{y})^2$. We use $\widehat{G}_n(x)$ and $\widetilde{G}_n(x)$ to denote the analogues of $G_n(x)$ with the matrix $\mathbf{B}_n$ replaced by $\widehat{\mathbf{B}}_n$ and $\widetilde{\mathbf{B}}_n$, respectively. Let $\lambda_i^{\mathbf{A}}$ denote the i-th smallest eigenvalue of Hermitian $\mathbf{A}$. Using the approach and bounds that are used in the proof of Corollary A.42, we have, for each $j = 1, 2, \cdots, k$,

$$\begin{aligned} \mathrm{E}\left|\int f_j(x) d\widehat{G}_n(x) - \int f_j(x) d\widetilde{G}_n(x)\right| &\leq K_j \sum_{k=1}^{n} \mathrm{E}|\lambda_k^{\widehat{\mathbf{B}}_n} - \lambda_k^{\widetilde{\mathbf{B}}_n}| \\ \leq 2K_j \Big(\mathrm{E}\mathrm{tr}\mathbf{T}^{1/2}(\widehat{\mathbf{X}}_n - \widetilde{\mathbf{X}}_n)(\widehat{\mathbf{X}}_n - \widetilde{\mathbf{X}}_n)^*\mathbf{T}^{1/2}\Big)^{1/2} \end{aligned}$$

$$\times\left(n^{-2}\mathrm{E}\mathrm{tr}\mathbf{T}^{1/2}(\widehat{\mathbf{X}}_n\widehat{\mathbf{X}}_n^* + \widetilde{\mathbf{X}}_n\widetilde{\mathbf{X}}_n^*)\mathbf{T}^{1/2}\right)^{1/2},$$

where K_j is a bound on $|f_j'(z)|$. From (9.7.7), $\max_{ij}|\sigma_{ij}-1| \to 0$. Thus, we have

$$\begin{aligned}\sum_{ij}(\sigma_{ij}^{-1}-1)^2 &\le K\sum_{ij}(\mathrm{E}|x_{ij}|^2 I(|x_{ij}|\ge \eta_n\sqrt{n}))^2\\ &\le \frac{1}{n^2\eta_n^4}\sum_{ij}(\mathrm{E}|x_{ij}|^4 I(|x_{ij}|\ge \eta_n\sqrt{n}))^2 = o(1).\end{aligned}$$

Moreover,

$$\sum_{ij}|\mathrm{E}\widehat{x}_{ij}|^2 \le \frac{1}{n^3\eta_n^6}\sum_{ij}(\mathrm{E}|x_{ij}^4|I(|x_{ij}|\ge\eta_n\sqrt{n}))^2 = o(1).$$

These give us

$$\begin{aligned}&\mathrm{E}\mathrm{tr}\mathbf{T}^{1/2}(\widehat{\mathbf{X}}_n-\widetilde{\mathbf{X}}_n)(\widehat{\mathbf{X}}_n-\widetilde{\mathbf{X}}_n)^*\mathbf{T}^{1/2}\\ &\le 2\sum_{ij}[(1-1/\sigma_{ij})^2\mathrm{E}|\hat{x}_{ij}|^2+\sigma_{ij}^{-2}|\mathrm{E}\hat{x}_{ij}|^2]\\ &\le 2\sum_{ij}[(1-1/\sigma_{ij})^2+|\mathrm{E}\hat{x}_{ij}|^2]\\ &= o(1).\end{aligned}$$

Similarly,

$$\begin{aligned}&n^{-2}\mathrm{E}\mathrm{tr}\mathbf{T}^{1/2}(\widehat{\mathbf{X}}_n\widehat{\mathbf{X}}_n^*+\widetilde{\mathbf{X}}_n\widetilde{\mathbf{X}}_n^*)\mathbf{T}^{1/2}\\ &\le \frac{1}{n^2}\sum_{ij}\mathrm{E}[|\hat{x}_{ij}|^2+|\tilde{x}_{ij}|^2]\\ &\le \frac{3}{n^2}\sum_{ij}\mathrm{E}|x_{ij}|^2 \le K.\end{aligned}$$

From the estimates above, we obtain

$$\int f_j(x)dG_n(x) = \int f_j(x)d\widetilde{G}_n(x) + o_p(1).$$

Therefore, we only need to find the limiting distribution of $\{\int f_j(x)d\widetilde{G}_n(x), j=1,\cdots,k\}$. Hence, in what follows, we shall assume the underlying variables are truncated at $\eta_n\sqrt{n}$, centralized, and renormalized. For simplicity, we shall suppress all sub- or superscripts on the variables and assume that $|x_{ij}|<\eta_n\sqrt{n}$, $\mathrm{E}x_{ij}=0$, $\mathrm{E}|x_{ij}|^2=1$, $\mathrm{E}|x_{ij}|^4<\infty$, and for the assumption made in Part (2) of Theorem 9.10, $\mathrm{E}|x_{ij}|^4=3+o(1)$, while for the assumption in (3), $\mathrm{E}x_{ij}^2=o(1/n)$ and $\mathrm{E}|x_{ij}|^4=2+o(1)$.

After truncation, centralization, and renormalization, with modifications in the proof of Theorem 5.10, for any $\mu_1 > \limsup \|\mathbf{T}_n\|(1+\sqrt{y})^2$ and $0 < \mu_2 < \liminf_n \lambda_{\min}^{\mathbf{T}_n} I_{(0,1)}(y)(1-\sqrt{y})^2$, we have

$$P(\|\mathbf{B}_n\| \geq \mu_1) = o(n^{-\ell}) \tag{9.7.8}$$

and

$$P(\lambda_{\min}^{\mathbf{B}_n} \leq \mu_2) = o(n^{-\ell}). \tag{9.7.9}$$

The modifications are given in Subsection 9.12.5. The main proof of Theorem 9.10 will be given in the following sections.

9.8 Convergence of Stieltjes Transforms

After truncation and centralization, our proof of the main theorem relies on establishing limiting results on

$$M_n(z) = p[s_{F^{\mathbf{B}_n}}(z) - s_{F^{y_n,H_n}}(z)] = n[s_{F^{\underline{\mathbf{B}}_n}}(z) - s_{\underline{F}^{y_n,H_n}}(z)],$$

or more precisely on $\widehat{M}_n(\cdot)$, a truncated version of $M_n(\cdot)$ when viewed as a random two-dimensional process defined on a contour $\mathcal{C}$ of the complex plane, described as follows. Let $v_0 > 0$ be arbitrary. Let x_r be any number greater than the right endpoint of interval (9.7.3). Let x_l be any negative number if the left endpoint of (9.7.3) is zero. Otherwise choose $x_l \in (0, \liminf_n \lambda_{\min}^{\mathbf{T}_n} I_{(0,1)}(y)(1-\sqrt{y})^2)$. Let

$$\mathcal{C}_u = \{x + iv_0 : x \in [x_l, x_r]\}.$$

Then define

$$\mathcal{C}^+ \equiv \{x_l + iv : v \in [0, v_0]\} \cup \mathcal{C}_u \cup \{x_r + iv : v \in [0, v_0]\}$$

and $\mathcal{C} = \mathcal{C}^+ \cup \overline{\mathcal{C}^+}$. Further, we now define the subsets $\mathcal{C}_n$ of $\mathcal{C}^+$ on which $M_n(\cdot)$ agrees with $\widehat{M}_n(\cdot)$. Choose sequence $\{\varepsilon_n\}$ decreasing to zero satisfying, for some $\alpha \in (0,1)$,

$$\varepsilon_n \geq n^{-\alpha}. \tag{9.8.1}$$

Let

$$\mathcal{C}_l = \begin{cases} \{x_l + iv : v \in [n^{-1}\varepsilon_n, v_0]\}, & \text{if } x_l > 0, \\ \{x_l + iv : v \in [0, v_0]\}, & \text{if } x_l < 0, \end{cases}$$

and

$$C_r = \{x_r + iv : v \in [n^{-1}\varepsilon_n, v_0]\}.$$

Then $\mathcal{C}_n = \mathcal{C}_l \cup \mathcal{C}_u \cup \mathcal{C}_r$. The process $\widehat{M}_n(\cdot)$ can now be defined. For $z = x + iv$, we have

$$\widehat{M}_n(z) = \begin{cases} M_n(z), & \text{for } z \in \mathcal{C}_n, \\ M_n(x_r + in^{-1}\varepsilon_n), & \text{for } x = x_r, v \in [0, n^{-1}\varepsilon_n], \\ M_n(x_l + in^{-1}\varepsilon_n), & \text{for } x = x_l, v \in [0, n^{-1}\varepsilon_n]. \end{cases} \tag{9.8.2}$$

$\widehat{M}_n(\cdot)$ is viewed as a random element in the metric space $C(\mathcal{C}^+, \mathbb{R}^2)$ of continuous functions from $\mathcal{C}^+$ to $\mathbb{R}^2$. All of Chapter 2 of Billingsley [57] applies to continuous functions from a set such as $\mathcal{C}^+$ (homeomorphic to $[0, 1]$) to finite-dimensional Euclidean space, with $|\cdot|$ interpreted as Euclidean distance.

We first prove the following lemma.

Lemma 9.11. *Under conditions (a) and (b) of Theorem 9.10, $\{\widehat{M}_n(\cdot)\}$ forms a tight sequence on $\mathcal{C}^+$. Moreover, if assumptions in (2) or (3) of Theorem 9.10 on x_{ij} hold, then $\widehat{M}_n(\cdot)$ converges weakly to a two-dimensional Gaussian process $M(\cdot)$ satisfying for $z \in \mathcal{C}^+$ under the assumptions in (2)*

$$\mathrm{E}M(z) = \frac{y \int \frac{\underline{s}(z)^3 t^2 dH(t)}{(1+t\underline{s}(z))^3}}{\left(1 - y \int \frac{\underline{s}(z)^2 t^2 dH(t)}{(1+t\underline{s}(z))^2}\right)^2} \tag{9.8.3}$$

and, for $z_1, z_2 \in \mathcal{C}$,

$$\begin{aligned} \mathrm{Cov}(M(z_1), M(z_2)) &\equiv \mathrm{E}[(M(z_1) - \mathrm{E}M(z_1))(M(z_2) - \mathrm{E}M(z_2)) \\ &= 2\left(\frac{\underline{s}'(z_1)\underline{s}'(z_2)}{(\underline{s}(z_1) - \underline{s}(z_2))^2} - \frac{1}{(z_1 - z_2)^2}\right), \end{aligned} \tag{9.8.4}$$

while under the assumptions in (3) $\mathrm{E}M(z) = 0$ and the "covariance" function analogous to (9.8.4) is $1/2$ the right-hand side of (9.8.4).

We now show how Theorem 9.10 follows from the lemma above. We use the identity

$$\int f(x) dG(x) = -\frac{1}{2\pi i} \int_{\mathcal{C}} f(z) s_G(z) dz, \tag{9.8.5}$$

valid for any c.d.f. G and f analytic on an open set containing the support of G. The complex integral on the right is over any positively oriented contour enclosing the support of G and on which f is analytic. Choose v_0, x_r, and x_l so that $f_1, \cdots, f_k$ are all analytic on and inside the resulting $\mathcal{C}$.

Due to the a.s. convergence of the extreme eigenvalues of $(1/n)\mathbf{X}_n\mathbf{X}_n^*$ and the bounds

$$\lambda_{\max}^{\mathbf{AB}} \le \lambda_{\max}^{\mathbf{A}} \lambda_{\max}^{\mathbf{B}}, \qquad \lambda_{\min}^{\mathbf{AB}} \ge \lambda_{\min}^{\mathbf{A}} \lambda_{\min}^{\mathbf{B}},$$

valid for $n \times n$ Hermitian nonnegative definite $\mathbf{A}$ and $\mathbf{B}$, we have with probability 1

$$\liminf_{n\to\infty} \min\left(x_r - \lambda_{\max}^{\mathbf{B}_n}, \lambda_{\min}^{\mathbf{B}_n} - x_l\right) > 0.$$

It also follows that the support of F^{y_n,H_n} is contained in

$$\left[\lambda_{\min}^{\mathbf{T}_n} I_{(0,1)}(y_n)(1-\sqrt{y_n})^2, \lambda_{\max}^{\mathbf{T}_n}(1+\sqrt{y_n})^2\right].$$

Therefore, for any $f \in \{f_1, \cdots, f_k\}$, with probability 1

$$\int f(x) dG_n(x) = -\frac{1}{2\pi i}\int_{\mathcal{C}} f(z) M_n(z) dz$$

for all large n, where the complex integral is over $\mathcal{C}$. Moreover, with $\widehat{M}_n(z) = \overline{\widehat{M}_n(\bar{z})}$ for $z \in \overline{\mathcal{C}^+}$, we have with probability 1, for all large n,

$$\begin{aligned}
&\left|\int_{\mathcal{C}} f(z)(M_n(z) - \widehat{M}_n(z))dz\right| \\
&\le 4K\varepsilon_n(|\max(\lambda_{\max}^{\mathbf{T}_n}(1+\sqrt{y_n})^2, \lambda_{\max}^{\mathbf{B}_n}) - x_r|^{-1} \\
&\quad +|\min(\lambda_{\min}^{\mathbf{T}_n} I_{(0,1)}(y_n)(1-\sqrt{y_n})^2, \lambda_{\min}^{\mathbf{B}_n}) - x_l|^{-1}),
\end{aligned}$$

which converges to zero as $n \to \infty$. Here K is a bound on f over $\mathcal{C}$.

Since

$$\widehat{M}_n(\cdot) \longrightarrow \left(-\frac{1}{2\pi i}\int f_1(z)\,\widehat{M}_n(z)dz, \cdots, -\frac{1}{2\pi i}\int f_k(z)\,\widehat{M}_n(z)dz\right)$$

is a continuous mapping of $C(\mathcal{C}^+, \mathbb{R}^2)$ into $\mathbb{R}^{2k}$, it follows that the vector above and subsequently (9.7.4) form tight sequences. Letting $M(\cdot)$ denote the limit of any weakly converging subsequence of $\{\widehat{M}_n(\cdot)\}$, we have the weak limit of (9.7.4) equal in distribution to

$$\left(-\frac{1}{2\pi i}\int_{\mathcal{C}} f_1(z)\,M(z)dz, \cdots, -\frac{1}{2\pi i}\int_{\mathcal{C}} f_k(z)\,M(z)dz\right).$$

The fact that this vector, under the assumptions in (2) or (3), is multivariate Gaussian follows from the fact that Riemann sums corresponding to these integrals are multivariate Gaussian and that weak limits of Gaussian vectors can only be Gaussian. The limiting expressions for the mean and covariance follow immediately.

The interval (9.7.3) in Theorem 9.10, on which the functions f_i are assumed to be analytic, can be reduced to a smaller one, due to the results in Chapter 6, relaxing the assumptions on the f_i's. Indeed, the f_i's need only be defined on an open interval I containing the closure of

$$\limsup_n S_{F^{y_n,H_n}} = \bigcap_{m=1}^{\infty} \bigcup_{n \ge m} S_{F^{y_n,H_n}}$$

since closed intervals in the complement of I will satisfy (f) of Theorem 6.3, so with probability 1, all eigenvalues will stay within I for all large n. Moreover, when $y[1-H(0)]>0$, which implies $p>n$, we have, by (2) of Theorem 6.3, the existence of $x_0>0$ for which $\lambda_n^{\mathbf{B}_n}$, the n-th largest eigenvalue of $\mathbf{B}_n$, which is $\lambda_{\min}^{\mathbf{B}_n}$, converges almost surely to x_0. Therefore, in this case, for all sufficiently small $\epsilon>0$, with probability 1,

$$X_n(f)=\int_{x_0-\epsilon}^{\infty} f(x)dG_n(x)$$

for all large n. Therefore the left endpoint of I can be taken to be $x_0-\epsilon$. When $\liminf_n \lambda_{\min}^{\mathbf{T}_n}>0$, a lower bound for x_0 can be taken to be any number less than $\liminf_n \lambda_{\min}^{\mathbf{T}_n}(1-\sqrt{y})^2$. For the proof of Theorem 9.10, the contour $\mathcal{C}$ could be adjusted accordingly.

Notice the assumptions in (2) and (3) require x_{ij} to have the same first, second, and fourth moments of either a real or complex Gaussian variable, the latter having real and imaginary parts i.i.d. $N(0,1/2)$. We will use the terms “RSE” and “CSE” to refer to the real and complex sample covariance matrices with these moment conditions.

The reason why concrete results are at present only obtained for the assumptions in (2) and (3) is mainly due to the identity

$$\begin{aligned}&\mathrm{E}(\mathbf{x}_t^*\mathbf{A}\mathbf{x}_t-\mathrm{tr}\mathbf{A})(\mathbf{x}_t^*\mathbf{B}\mathbf{x}_t-\mathrm{tr}\mathbf{B})\\ &=\sum_{i=1}^{p}(\mathrm{E}|x_{it}|^4-|\mathrm{E}x_{it}^2|^2-2)a_{ii}b_{ii}\\ &+\mathrm{tr}\mathbf{A}_x\mathbf{B}_x^T+\mathrm{tr}AB \qquad (9.8.6)\end{aligned}$$

valid for $p\times p$ $\mathbf{A}=(a_{ij})$ and $\mathbf{B}=(b_{ij})$, where $\mathbf{x}_t$ is the t-th column of $\mathbf{X}_n$, $\mathbf{A}_x=(\mathrm{E}x_{it}^2a_{ij})$, and $\mathbf{B}_x=(\mathrm{E}x_{it}^2b_{ij})$ (note t is fixed). This formula will be needed in several places in the proof of Lemma 9.11. The assumptions in (3) leave only the last term on the right, whereas those in (2) leave the last two, but in this case the matrix $\mathbf{B}$ will always be symmetric. This also accounts for the relation between the two covariance functions and the difficulty in obtaining explicit results more generally. As will be seen in the proof, whenever (9.8.6) is used, little is known about the limiting behavior of $\sum a_{ii}b_{ii}$ even when we assume the underlying distributions are identical.

Simple substitution reveals

$$\text{RHS of (9.7.6)}=-\frac{1}{2\pi^2}\int_{\mathcal{C}_1}\int_{\mathcal{C}_2}\frac{f(z(s_1))g(z(s_2))}{(s_1-s_2)^2}d(s_1)d(s_2). \qquad (9.8.7)$$

However, the contours depend on the z_1, z_2 contours and cannot be arbitrarily chosen. It is also true that

$$\text{RHS of (9.7.6)} = \frac{1}{\pi^2} \iint f'(x)g'(y) \log \left| \frac{\underline{s}(x) - \overline{\underline{s}}(y)}{\underline{s}(x) - \underline{s}(y)} \right| dxdy$$

$$= \frac{1}{2\pi^2} \iint f'(x)g'(y) \log \left(1 + 4 \frac{\underline{s}_i(x)\underline{s}_i(y)}{|\underline{s}(x) - \underline{s}(y)|^2} \right) dxdy \qquad (9.8.8)$$

and

$$\mathrm{E}X_f = \frac{1}{2\pi} \int f'(x) \arg \left(1 - y \int \frac{t^2 \underline{s}^2(x)}{(1 + t\underline{s}(x))^2} dH(t) \right) dx. \qquad (9.8.9)$$

Here, for $0 \neq x \in \mathbb{R}$,

$$\underline{s}(x) = \lim_{z \to x} \underline{s}(z), \qquad z \in \mathbb{C}^+, \qquad (9.8.10)$$

known to exist and satisfying (9.7.1), and $\underline{s}_i(x) = \Im\, \underline{s}(x)$. The term

$$j(x) = \arg \left(1 - y \int \frac{t^2 \underline{s}^2(x)}{(1 + t\underline{s}(x))^2} dH(t) \right)$$

in (9.8.9) is well defined for almost every x and takes values in $(-\pi/2, \pi/2)$. Section 9.12 contains proofs of all the expressions above. Subsections 9.12.1 and 9.12.2 contain the proof of (9.8.8) and (9.8.9) along with showing

$$k(x, y) \equiv \log \left(1 + 4 \frac{\underline{s}_i(x)\underline{s}_i(y)}{|\underline{s}(x) - \underline{s}(y)|^2} \right) \qquad (9.8.11)$$

to be Lebesgue integrable on $\mathbb{R}^2$. It is interesting to note that the support of $k(x, y)$ matches the support of $f^{y,H}$ on $\mathbb{R} - \{0\}$:

$$k(x, y) = 0 \Longleftrightarrow \min(f^{y,H}(x), f^{y,H}(y)) = 0.$$

We also have $f^{y,H}(x) = 0 \Longrightarrow j(x) = 0$.

Subsection 9.12.3 contains derivations of the relevant quantities associated with Example 1.1. The linear spectral statistic $(1/p)T_n$ has a.s. limit $d(y)$ as stated in Example 1.1. The quantity $T_n - pd(p/n)$ converges weakly to a Gaussian random variable $X_{\log}$ with

$$\mathrm{E}X_{\log} = \frac{1}{2} \log(1 - y) \qquad (9.8.12)$$

and

$$\mathrm{Var}\, X_{\log} = -2 \log(1 - y). \qquad (9.8.13)$$

Jonsson [169] derived the limiting distribution of $\mathrm{tr}\mathbf{S}_n^r - \mathrm{E}\mathrm{tr}\mathbf{S}_n^r$ when $n\mathbf{S}_n$ is a standard Wishart matrix. As a generalization to this work, results on both $\mathrm{tr}\mathbf{S}_n^r - \mathrm{E}\mathbf{S}_n^r$ and $p[\int x^r dF^{S_n}(x) - \mathrm{E} \int x^r dF^{S_n}(x)]$ for positive integer r are derived in Section 9.12.4, where the following expressions are presented for means and covariances in this case ($H = I_{[1,\infty)}$). We have

$$\mathrm{E}X_{x^r} = \frac{1}{4}((1-\sqrt{y})^{2r} + (1+\sqrt{y})^{2r}) - \frac{1}{2}\sum_{j=0}^{r}\binom{r}{j}^2 y^j \tag{9.8.14}$$

and

$$\begin{aligned}
&\mathrm{Cov}(X_{x^{r_1}}, X_{x^{r_2}}) \\
&= 2y^{r_1+r_2}\sum_{k_1=0}^{r_1-1}\sum_{k_2=0}^{r_2}\binom{r_1}{k_1}\binom{r_2}{k_2}\left(\frac{1-y}{y}\right)^{k_1+k_2}\sum_{\ell=1}^{r_1-k_1}\ell\binom{2r_1-1-(k_1+\ell)}{r_1-1} \\
&\quad\times\binom{2r_2-1-k_2+\ell}{r_2-1}.
\end{aligned} \tag{9.8.15}$$

It is worth mentioning here a consequence of (9.8.8), namely that if the assumptions in (2) or (3) of Theorem 9.10 were to hold, then G_n, considered as a random element in $D[0,\infty)$ (the space of functions on $[0,\infty)$ that are right-continuous with left-hand limits, together with the Skorohod metric), cannot form a tight sequence in $D[0,\infty)$. Indeed, under either assumption, if $G(x)$ is a weak limit of a subsequence, then, because of Theorem 9.10, it is straightforward to conclude that for any x_0 in the interior of the support of F and positive ε,

$$\int_{x_0}^{x_0+\varepsilon} G(x)dx$$

would be Gaussian and therefore so would

$$G(x_0) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{x_0}^{x_0+\varepsilon} G(x)dx.$$

However, the variance would necessarily be

$$\lim_{\varepsilon\to 0}\frac{1}{2\pi^2}\frac{1}{\varepsilon^2}\int_{x_0}^{x_0+\varepsilon}\int_{x_0}^{x_0+\varepsilon} k(x,y)dxdy = \infty.$$

Still, under the assumptions in (2) or (3), a limit may exist for $\{G_n\}$ when G_n is viewed as a linear functional,

$$f \longrightarrow \int f(x)dG_n(x);$$

that is, a limit expressed in terms of a measure in a space of generalized functions. The characterization of the limiting measure of course depends on the space, which in turn relies on the set of test functions, which for now is restricted to functions analytic on the support of F. Work in this area is currently being pursued.

The proof of Lemma 9.11 is divided into three sections. Sections 9.9 and 9.10 handle the limiting behavior of the centralized M_n, while Section 9.11 analyzes the nonrandom part.

We mention here some simple extensions of results in Billingsley [57] needed in the analysis.

Bi. In the Arzela-Ascoli theorem on p. 221, (8) can be replaced by

$$\sup_{x \in A} |x(t_0)| < \infty$$

for some $t_0 \in [0,1]$. Subsequently (i) of Theorem 12.3, p. 95, can be replaced by

The sequence $\{X_n(t_0)\}$ *is tight for some* $t_0 \in [0,1]$.

Bii. If the sequence $\{X_n\}$ of random elements in $C[0,1]$ is tight, and the sequence $\{a_n\}$ of nonrandom elements in $C[0,1]$ satisfies the first part of Bi. above with $A = \{a_n\}$ and is uniformly equicontinuous ((9) on p. 221), then the sequence $\{X_n + a_n\}$ is tight.

Biii. For any countably dense subset T of $[0,1]$, the finite-dimensional sets (see pp. 19–20) formed from points in T uniquely determine probability measures on $C[0,1]$ (that is, it is a determining class). This implies that a random element of $C[0,1]$ is uniquely determined by its finite-dimensional distributions having points in T.

9.9 Convergence of Finite-Dimensional Distributions

Write for $z \in \mathcal{C}_n$, $M_n(z) = M_n^1(z) + M_n^2(z)$, where

$$M_n^1(z) = p[s_{F^{\mathbf{B}_n}}(z) - \mathrm{E}s_{F^{\mathbf{B}_n}}(z)]$$

and

$$M_n^2(z) = p[s_{\mathrm{E}F^{\mathbf{B}_n}}(z) - s_{F^{y_n,H_n}}(z)],$$

and define $\widehat{M}_n^1(z)$, $\widehat{M}_n^2(z)$ for $z \in \mathcal{C}^+$ in terms of M_n^1, M_n^2 as in (9.8.2). In this section, we will show for any positive integer r the sum

$$\sum_{i=1}^{r} \alpha_i M_n^1(z_i) \qquad (\Im z_i \neq 0)$$

whenever it is real, tight, and, under the assumptions in (2) or (3) of Theorem 9.10, will converge in distribution to a Gaussian random variable. Formula (9.8.4) will also be derived. From this and the result to be obtained in Section 9.11, we will have weak convergence of the finite-dimensional distributions of $\widehat{M}_n(z)$ for all $\in \mathcal{C}^+$, except at the two endpoints. Because of Biii, this will be enough to ensure the uniqueness of any weakly converging subsequence of $\{\widehat{M}_n\}$.

We begin by quoting the following result.

Lemma 9.12. (Theorem 35.12 of Billingsley [56]). *Suppose that for each n, $Y_{n1}, Y_{n2}, \cdots, Y_{nr_n}$ is a real martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_{nj}\}$ having second moments. If, as $n \to \infty$,*

$$\sum_{j=1}^{r_n} \mathrm{E}(Y_{nj}^2 | \mathcal{F}_{n,j-1}) \xrightarrow{i.p.} \sigma^2, \tag{9.9.1}$$

where σ^2 is a positive constant, and, for each $\varepsilon > 0$,

$$\sum_{j=1}^{r_n} \mathrm{E}(Y_{nj}^2 I_{(|Y_{nj}| \geq \varepsilon)}) \to 0, \tag{9.9.2}$$

then

$$\sum_{j=1}^{r_n} Y_{nr_n} \xrightarrow{\mathscr{D}} N(0, \sigma^2).$$

Recalling the truncation and centralization steps, if $\mathbf{C}$ is a matrix with $\|\mathbf{C}\| \leq K$ on B_n^c and $\|\mathbf{C}\| < n^d$ on B_n for some constant d, then by Lemma 9.1, (9.7.8), and (9.7.9), we get (similar to (9.2.17))

$$\mathrm{E}|\mathbf{x}_t^* \mathbf{C} \mathbf{x}_t - \mathrm{tr}\mathbf{C}|^p \leq K_p \|\mathbf{C}\|^p \eta_n^{2p-4} n^{p-1} \leq K_p \eta_n^{2p-4} n^{p-1}, \quad p \geq 2, \tag{9.9.3}$$

where $B_n = \{\|\mathbf{B}_n\| > \mu_1 \text{ or } \lambda_{\min}^{\mathbf{B}_n} < \mu_2\}$.

Let $v = \Im z$. For the following analysis, we will assume $v > 0$. To facilitate notation, we will let $\mathbf{T} = \mathbf{T}_n$. Because of assumption (b) of Theorem 9.10, we may assume $\|\mathbf{T}\| \leq 1$ for all n. Constants appearing in inequalities will be denoted by K and may take on different values from one expression to the next.

In what follows, we use the notation $\mathbf{r}_j$, $\mathbf{D}(z)$, $\mathbf{D}_j(z)$, α_j, δ_j, γ_j, $\hat{\gamma}_j$, $\bar{\gamma}_j$, β_j, $\bar{\beta}_j$ defined in subsections 6.2.2 and 6.2.3 and define

$$b_j = \frac{1}{1 + n^{-1}\mathrm{Etr}\mathbf{T}\mathbf{D}_j^{-1}} \quad \text{and} \quad b = \frac{1}{1 + n^{-1}\mathrm{Etr}\mathbf{T}\mathbf{D}^{-1}}. \tag{9.9.4}$$

Each of β_j, $\bar{\beta}_j$, b_j, and b is bounded in absolute value by $|z|/v$ (see (6.2.5)). We have

$$\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z) = -\mathbf{D}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\beta_j(z),$$

and from Lemma 6.9 for any $p \times p$ $\mathbf{A}$,

$$|\mathrm{tr}(\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z))\mathbf{A}| \leq \frac{\|\mathbf{A}\|}{\Im z}. \tag{9.9.5}$$

For nonrandom $p \times p$ $\mathbf{A}_k$, $k = 1, \cdots, m$ and $\mathbf{B}_l$, $l = 1, \cdots, q$, we shall establish the following inequality:

$$\left|\mathrm{E}\left(\prod_{k=1}^{m}\mathbf{r}_t^*\mathbf{A}_k\mathbf{r}_t\prod_{l=1}^{q}(\mathbf{r}_t^*\mathbf{B}_l\mathbf{r}_t-n^{-1}\mathrm{tr}\mathbf{T}\mathbf{B}_l)\right)\right|$$
$$\le Kn^{-(1\wedge q)}\eta_n^{(2q-4)\vee 0}\prod_{k=1}^{m}\|\mathbf{A}_k\|\prod_{l=1}^{q}\|\mathbf{B}_l\|,\quad m\ge 0,\ q\ge 0. \tag{9.9.6}$$

When $m=0$ and $q=1$, the left side is 0. When $m=0$ and $q>1$, (9.9.6) is a consequence of (9.9.3) and Hölder's inequality. If $m\ge 1$, then by induction on m we have

$$\left|\mathrm{E}\left(\prod_{k=1}^{m}\mathbf{r}_t^*\mathbf{A}_k\mathbf{r}_t\prod_{l=1}^{q}(\mathbf{r}_t^*\mathbf{B}_l\mathbf{r}_t-n^{-1}\mathrm{tr}\mathbf{T}\mathbf{B}_l)\right)\right|$$
$$\le\left|\mathrm{E}\left(\prod_{k=1}^{m-1}\mathbf{r}_t^*\mathbf{A}_k\mathbf{r}_t(\mathbf{r}_t^*\mathbf{A}_p\mathbf{r}_t-n^{-1}\mathrm{tr}\mathbf{T}\mathbf{A}_p)\prod_{l=1}^{q}(\mathbf{r}_t^*\mathbf{B}_l\mathbf{r}_t-n^{-1}\mathrm{tr}\mathbf{T}\mathbf{B}_l)\right)\right|$$
$$+pn^{-1}\|\mathbf{A}_p\|\left|\mathrm{E}\left(\prod_{k=1}^{m-1}\mathbf{r}_t^*\mathbf{A}_k\mathbf{r}_t\prod_{l=1}^{q}(\mathbf{r}_t^*\mathbf{B}_l\mathbf{r}_t-n^{-1}\mathrm{tr}\mathbf{T}\mathbf{B}_l)\right)\right|$$
$$\le Kn^{-1}\eta_n^{(2q-4)\vee 0}\prod_{k=1}^{m}\|\mathbf{A}_k\|\prod_{l=1}^{q}\|\mathbf{B}_l\|.$$

We have proved the case where $q>0$. When $q=0$, (9.9.6) is a trivial consequence of (9.9.3).

Let $\mathrm{E}_0(\cdot)$ denote expectation and $\mathrm{E}_j(\cdot)$ denote conditional expectation with respect to the σ-field generated by $\mathbf{r}_1,\cdots,\mathbf{r}_j$.

Using the martingale decomposition, we have

$$p[s_{F^{\mathbf{B}_n}}(z)-\mathrm{E}s_{F^{\mathbf{B}_n}}(z)]=\mathrm{tr}[\mathbf{D}^{-1}(z)-\mathrm{E}\mathbf{D}^{-1}(z)]$$
$$=-\sum_{j=1}^{n}(\mathrm{E}_j-\mathrm{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j.$$

Write

$$\beta_j(z)=\bar\beta_j(z)-\beta_j(z)\bar\beta_j(z)\hat\gamma_j(z)$$
$$=\bar\beta_j(z)-\bar\beta_j^2(z)\hat\gamma_j(z)+\bar\beta_j^2(z)\beta_j(z)\hat\gamma_j^2(z).$$

Then we have

$$(\mathrm{E}_j-\mathrm{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z)\mathbf{r}_j$$
$$=\mathrm{E}_j\left(\bar\beta_j(z)\alpha_j(z)-\bar\beta_j^2(z)\hat\gamma_j(z)\frac{1}{n}\mathrm{tr}\mathbf{T}\mathbf{D}_j^{-2}(z)\right)$$
$$-(\mathrm{E}_j-\mathrm{E}_{j-1})\bar\beta_j^2(z)(\hat\gamma_j(z)\alpha_j(z)-\beta_j(z)\mathbf{r}_j\mathbf{D}_j^{-2}(z)\mathbf{r}_j\hat\gamma_j^2(z)).$$

Using (9.9.6), we have

$$\begin{aligned}
&\mathrm{E}\left|\sum_{j=1}^{n}(\mathrm{E}_j - \mathrm{E}_{j-1})\bar{\beta}_j^2(z)\hat{\gamma}_j(z)\alpha_j(z)\right|^2 \\
&= \sum_{j=1}^{n} \mathrm{E}|(\mathrm{E}_j - \mathrm{E}_{j-1})\bar{\beta}_j^2(z)\hat{\gamma}_j(z)\alpha_j(z)|^2 \\
&\le 4\sum_{j=1}^{n} \mathrm{E}|\bar{\beta}_j^2(z)\hat{\gamma}_j(z)\alpha_j(z)|^2 = o(1).
\end{aligned}$$

Therefore, $\sum_{j=1}^{n}(\mathrm{E}_j - \mathrm{E}_{j-1})\bar{\beta}_j^2(z)\hat{\gamma}_j(z)\alpha_j(z)$ converges to zero in probability.

By the same argument, we have

$$\sum_{j=1}^{n}(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z)\mathbf{r}_j\mathbf{D}_j^{-2}(z)\mathbf{r}_j\hat{\gamma}_j^2(z) \xrightarrow{i.p.} 0.$$

Therefore we need only consider the sum

$$\sum_{i=1}^{r}\alpha_i\sum_{j=1}^{n}Y_j(z_i) = \sum_{j=1}^{n}\sum_{i=1}^{r}\alpha_i Y_j(z_i),$$

where

$$\begin{aligned}
Y_j(z) &= -\mathrm{E}_j\left(\bar{\beta}_j(z)\alpha_j(z) - \bar{\beta}_j^2(z)\hat{\gamma}_j(z)\frac{1}{n}\mathrm{tr}\mathbf{T}\mathbf{D}_j^{-2}(z)\right) \\
&= -\mathrm{E}_j\frac{d}{dz}\bar{\beta}_j(z)\hat{\gamma}_j(z).
\end{aligned}$$

Again, by using (9.9.6), we obtain

$$\mathrm{E}|Y_j(z)|^4 \le K\left(\frac{|z|^4}{v^4}\mathrm{E}|\alpha_j(z)|^4 + \frac{|z|^8}{v^{16}}\left(\frac{p}{n}\right)^4\mathrm{E}|\hat{\gamma}_j(z)|^4\right) = o(n^{-1}),$$

which implies, for any $\varepsilon > 0$,

$$\begin{aligned}
&\sum_{j=1}^{n}\mathrm{E}\left(\left|\sum_{i=1}^{r}\alpha_i Y_j(z_i)\right|^2 I\left(\left|\sum_{i=1}^{r}\alpha_i Y_j(z_i)\right| \ge \varepsilon\right)\right) \\
&\le \frac{1}{\varepsilon^2}\sum_{j=1}^{n}\mathrm{E}\left|\sum_{i=1}^{r}\alpha_i Y_j(z_i)\right|^4 \to 0
\end{aligned}$$

as $n \to \infty$. Therefore condition (ii) of Lemma 9.12 is satisfied and it is enough to prove, under the assumptions in (2) or (3), for $z_1, z_2 \in \mathbb{C}$ with $\Im(z_j) \neq 0$, that

$$\sum_{j=1}^{n} \mathrm{E}_{j-1}[Y_j(z_1)Y_j(z_2)] \tag{9.9.7}$$

converges in probability to a constant (and to determine the constant).

We show here for future use the tightness of the sequence $\{\sum_{i=1}^r \alpha_i M_n^1(z_i)\}$. From (9.9.6) we easily get $\mathrm{E}|Y_j(z)|^2 = O(n^{-1})$, so that

$$\begin{aligned}\mathrm{E}\left|\sum_{i=1}^{r} \alpha_i \sum_{j=1}^{n} Y_j(z_i)\right|^2 &= \sum_{j=1}^{n} \mathrm{E}\left|\sum_{i=1}^{r} \alpha_i Y_j(z_i)\right|^2 \\ &\le r \sum_{j=1}^{n}\sum_{i=1}^{r} |\alpha_i|^2 \mathrm{E}\left|Y_j(z_i)\right|^2 \le K. \end{aligned} \tag{9.9.8}$$

Consider the sum

$$\sum_{j=1}^{n} \mathrm{E}_{j-1}[\mathrm{E}_j(\bar\beta_j(z_1)\hat\gamma_j(z_1))\mathrm{E}_j(\bar\beta_j(z_2)\hat\gamma_j(z_2))]. \tag{9.9.9}$$

In the j-th term (viewed as an expectation with respect to $\mathbf{r}_{j+1}, \cdots, \mathbf{r}_n$), we apply the d.c.t. to the difference quotient defined by $\bar\beta_j(z)\hat\gamma_j(z)$ to get

$$\frac{\partial^2}{\partial z_2 \partial z_1}(9.9.9) = (9.9.7).$$

Let v_0 be a lower bound on $|\Im z_i|$. For each j, let $\mathbf{A}_j^i = (1/n)\mathbf{T}^{1/2}\mathrm{E}_j\mathbf{D}_j^{-1}(z_i)\mathbf{T}^{1/2}$, $i = 1, 2$. Then $\mathrm{tr}\mathbf{A}_j^i\mathbf{A}_j^{i\,*} \le p(v_0 n)^{-2}$. Using (9.9.3) we therefore see that (9.9.9) is bounded.

We can then appeal to Lemma 2.14. Suppose (9.9.9) converges in probability for each $z_k, z_l \in \{z_i\}$, bounded away from the imaginary axis and having a limit point. Then, by a diagonalization argument, for any subsequence of the natural numbers, there is a further subsequence such that, with probability 1, (9.9.9) converges for each pair z_k, z_l. Applying Lemma 2.14 twice, we see that almost surely (9.9.7) will converge on the subsequence for each pair. That is enough to imply convergence in probability of (9.9.7). Therefore we need only show that (9.9.9) converges in probability.

By the definition of β_1 and b_1, using the martingale decomposition to $\mathrm{tr}\mathbf{T}\mathbf{D}_j(z_i) - \mathrm{E}\mathrm{tr}\mathbf{T}\mathbf{D}_j(z_i)$, we get

$$\begin{aligned}&\mathrm{E}|\bar\beta_1(z_i) - b_1(z_i)|^2 \\ &\le |z|^4 v_0^{-4} n^{-2}\mathrm{E}|\mathrm{tr}\mathbf{T}\mathbf{D}_1(z_i) - \mathrm{E}\mathrm{tr}\mathbf{T}\mathbf{D}_1(z_i)|^2 \\ &\le K\frac{|z_i|^4}{v_0^6} n^{-1}.\end{aligned}$$

Similarly, we have

$$\mathrm{E}|\bar{\beta}_j(z_i) - b_j(z_i)|^2 \le K\frac{|z_i|^4}{v_0^6}n^{-1}.$$

This, together with (9.9.6), implies that

$$\begin{aligned}&\mathrm{E}|\mathrm{E}_{j-1}[\mathrm{E}_j(\bar{\beta}_j(z_1)\hat{\gamma}_j(z_1))\mathrm{E}_j(\bar{\beta}_j(z_2)\hat{\gamma}_j(z_2))]\\&\quad -\mathrm{E}_{j-1}[\mathrm{E}_j(b_j(z_1)\hat{\gamma}_j(z_1))\mathrm{E}_j(b_j(z_2)\hat{\gamma}_j(z_2))]|\\&= o(n^{-1}),\end{aligned}$$

from which

$$\begin{aligned}&\sum_{j=1}^{n}\mathrm{E}_{j-1}[\mathrm{E}_j(\bar{\beta}_j(z_1)\hat{\gamma}_j(z_1))\mathrm{E}_j(\bar{\beta}_j(z_2)\hat{\gamma}_j(z_2))]\\&\quad -\sum_{j=1}^{n} b_j(z_1)b_j(z_2)\mathrm{E}_{j-1}[\mathrm{E}_j(\hat{\gamma}_j(z_1))\mathrm{E}_j(\hat{\gamma}_j(z_2))]\\&\xrightarrow{i.p.} 0.\end{aligned}$$

Thus the goal is to show that

$$\sum_{j=1}^{n} b_j(z_1)b_j(z_2)\mathrm{E}_{j-1}[\mathrm{E}_j(\hat{\gamma}_j(z_1))\mathrm{E}_j(\hat{\gamma}_j(z_2))] \tag{9.9.10}$$

converges in probability and to determine its limit. The latter's second mixed partial derivative will yield the limit of (9.9.7).

We now assume the CSE case, namely $\mathrm{E}X_{11}^2 = o(1/n)$ and $\mathrm{E}|X_{11}|^4 = 2 + o(1)$, so that, using (9.8.6), (9.9.10) becomes

$$\frac{1}{n^2}\sum_{j=1}^{n} b_j(z_1)b_j(z_2)(\mathrm{tr}\mathbf{T}^{1/2}\mathrm{E}_j(\mathbf{D}_j^{-1}(z_1))\mathbf{T}\mathrm{E}_j(\mathbf{D}_j^{-1}(z_2))\mathbf{T}^{1/2} + o(1)A_n),$$

where

$$\begin{aligned}|A_n| &\le K(\mathrm{tr}\mathbf{T}\mathrm{E}_j(\mathbf{D}_j^{-1}(z_1))\mathbf{T}\mathrm{E}_j(\bar{\mathbf{D}}_j^{-1}(z_1))\\&\qquad \times\mathrm{tr}\mathbf{T}\mathrm{E}_j(\mathbf{D}_j^{-1}(z_2))\mathbf{T}\mathrm{E}_j(\bar{\mathbf{D}}_j^{-1}(z_2)))^{1/2} = O(n).\end{aligned}$$

Thus we need only to study the limit of

$$\frac{1}{n^2}\sum_{j=1}^{n} b_j(z_1)b_j(z_2)\mathrm{tr}\mathrm{E}_j(\mathbf{D}_j^{-1}(z_1))\mathbf{T}\mathrm{E}_j(\mathbf{D}_j^{-1}(z_2))\mathbf{T}. \tag{9.9.11}$$

The RSE case ($\mathbf{T}$, X_{11} real, $\mathrm{E}|X_{11}|^4 = 3 + o(1)$) will be double that of the limit of (9.9.11).

Let $\mathbf{D}_{ij}(z) = \mathbf{D}(z) - \mathbf{r}_i\mathbf{r}_i^* - \mathbf{r}_j\mathbf{r}_j^*$,

$$\beta_{ij}(z) = \frac{1}{1 + \mathbf{r}_i^* \mathbf{D}_{ij}^{-1}(z)\mathbf{r}_i}, \quad \text{and} \quad b_{i,j}(z) = \frac{1}{1 + n^{-1}\mathrm{E}\mathrm{tr}\mathbf{T}\mathbf{D}_{i\,j}^{-1}(z)}.$$

We write

$$\mathbf{D}_j(z_1) + z_1\mathbf{I} - \frac{n-1}{n} b_j(z_1)\mathbf{T} = \sum_{i \neq j} \mathbf{r}_i\mathbf{r}_i^* - \frac{n-1}{n} b_j(z_1)\mathbf{T}.$$

Multiplying by $(z_1\mathbf{I} - \frac{n-1}{n} b_j(z_1)\mathbf{T})^{-1}$ on the left, $\mathbf{D}_j^{-1}(z_1)$ on the right, and using

$$\mathbf{r}_i^*\mathbf{D}_j^{-1}(z_1) = \beta_{ij}(z_1)\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1),$$

we get

$$\begin{aligned}
\mathbf{D}_j^{-1}(z_1) = &-\left(z_1\mathbf{I} - \frac{n-1}{n} b_j(z_1)\mathbf{T}\right)^{-1} \\
&+ \sum_{i \neq j} \beta_{ij}(z_1)\left(z_1\mathbf{I} - \frac{n-1}{n} b_j(z_1)\mathbf{T}\right)^{-1} \mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1) \\
&- \frac{n-1}{n} b_j(z_1)\left(z_1\mathbf{I} - \frac{n-1}{n} b_j(z_1)\mathbf{T}\right)^{-1} \mathbf{T}\mathbf{D}_j^{-1}(z_1) \\
= &-\left(z_1\mathbf{I} - \frac{n-1}{n} b_j(z_1)\mathbf{T}\right)^{-1} + b_j(z_1)\mathbf{A}(z_1) \\
&+ \mathbf{B}(z_1) + \mathbf{C}(z_1),
\end{aligned} \tag{9.9.12}$$

where

$$\mathbf{A}(z_1) = \sum_{i \neq j}\left(z_1\mathbf{I} - \frac{n-1}{n} b_j(z_1)\mathbf{T}\right)^{-1} (\mathbf{r}_i\mathbf{r}_i^* - n^{-1}\mathbf{T})\mathbf{D}_{ij}^{-1}(z_1),$$

$$\mathbf{B}(z_1) = \sum_{i \neq j}(\beta_{ij}(z_1) - b_j(z_1))\left(z_1\mathbf{I} - \frac{n-1}{n} b_j(z_1)\mathbf{T}\right)^{-1} \mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1),$$

and

$$\mathbf{C}(z_1) = n^{-1}b_j(z_1)\left(z_1\mathbf{I} - \frac{n-1}{n} b_j(z_1)\mathbf{T}\right)^{-1} \mathbf{T}\sum_{i \neq j}(\mathbf{D}_{ij}^{-1}(z_1) - \mathbf{D}_j^{-1}(z_1)).$$

It is easy to verify, for any real t, that

$$\begin{aligned}
&\left|1 - \frac{t}{z(1 + n^{-1}\mathrm{E}\mathrm{tr}\mathbf{T}\mathbf{D}_j^{-1}(z))}\right|^{-1} \le \frac{|z(1 + n^{-1}\mathrm{E}\mathrm{tr}\mathbf{T}\mathbf{D}_j^{-1}(z))|}{\Im\, z(1 + n^{-1}\mathrm{E}\mathrm{tr}\mathbf{T}\mathbf{D}_j^{-1}(z))} \\
&\le \frac{|z|(1 + p/(nv_0))}{v_0}.
\end{aligned}$$

Thus

$$\left\| \left(z_1 \mathbf{I} - \frac{n-1}{n} b_j(z_1) \mathbf{T} \right)^{-1} \right\| \le \frac{1 + \frac{p}{nv_0}}{v_0}. \tag{9.9.13}$$

Let $\mathbf{M}$ be $p \times p$ and let $\|\mathbf{M}\|$ denote a nonrandom bound on the spectral norm of $\mathbf{M}$ for all parameters governing $\mathbf{M}$ and under all realizations of $\mathbf{M}$. From (9.9.5),

$$|b_{i,j}(z_1) - b_j(z_1)| \le K/n$$

and

$$\mathrm{E}|\beta_{ij} - b_{i,j}|^2 \le K/n.$$

Then, by (9.9.6) and (9.9.13), we get

$$\begin{aligned} \mathrm{E}|\mathrm{tr}\mathbf{B}(z_1)\mathbf{M}| &\le \sum_{i \ne j} \mathrm{E}^{1/2}(|\beta_{i,j}(z_1) - b_j(z_1)|^2) \\ &\quad \times \mathrm{E}^{1/2}\left(\left| \mathbf{r}_i^* \mathbf{D}_{ij}^{-1}(z_1) \mathbf{M} \left(z_1 \mathbf{I} - \frac{n-1}{n} b_j(z_1) \mathbf{T} \right)^{-1} \mathbf{r}_i \right|^2 \right) \\ &\le K \|\mathbf{M}\| n^{1/2}. \end{aligned} \tag{9.9.14}$$

From (9.9.5), we have

$$|\mathrm{tr}C(z_1)\mathbf{M}| \le K\|\mathbf{M}\|. \tag{9.9.15}$$

From (9.9.6) and (9.9.13), we get for $\mathbf{M}$ nonrandom and any j,

$$\begin{aligned} &\mathrm{E}|\mathrm{tr}\mathbf{A}(z_1)\mathbf{M}| \\ &\le \frac{K}{n} \sum_{i \ne j} \mathrm{E}^{1/2} \mathrm{tr} \mathbf{T}^{1/2} \mathbf{D}_{ij}^{-1}(z_1) \mathbf{M} \left(z_1 I - \frac{n-1}{n} b_j(z_1) \mathbf{T} \right)^{-1} \mathbf{T} \\ &\quad \times \left(\bar{z}_1 \mathbf{I} - \frac{n-1}{n} b_j(\bar{z}_1) \mathbf{T} \right)^{-1} \mathbf{M}^* \mathbf{D}_{ij}^{-1}(\bar{z}_1) \mathbf{T}^{1/2} \\ &\le K \|\mathbf{M}\| n^{1/2}. \end{aligned} \tag{9.9.16}$$

We write (using the identity above (9.9.5))

$$\mathrm{tr}[\mathrm{E}_j \mathbf{A}(z_1)] \mathbf{T} \mathbf{D}_j^{-1}(z_2) \mathbf{T} = A_1(z_1, z_2) + A_2(z_1, z_2) + A_3(z_1, z_2), \tag{9.9.17}$$

where

$$\begin{aligned} A_1(z_1, z_2) &= -\mathrm{tr} \sum_{i<j} \left(z_1 \mathbf{I} - \frac{n-1}{n} b_j(z_1) \mathbf{T} \right)^{-1} \mathbf{r}_i \mathbf{r}_i^* [\mathrm{E}_j \mathbf{D}_{ij}^{-1}(z_1)] \\ &\quad \times \mathbf{T} \beta_{ij}(z_2) \mathbf{D}_{ij}^{-1}(z_2) \mathbf{r}_i \mathbf{r}_i^* \mathbf{D}_{ij}^{-1}(z_2) \mathbf{T} \\ &= -\sum_{i<j} \beta_{ij}(z_2) \mathbf{r}_i^* [\mathrm{E}_j \mathbf{D}_{ij}^{-1}(z_1)] \mathbf{T} \mathbf{D}_{ij}^{-1}(z_2) \mathbf{r}_i \mathbf{r}_i^* \mathbf{D}_{ij}^{-1}(z_2) \mathbf{T} \end{aligned}$$

$$\times \left(z_1 \mathbf{I} - \frac{n-1}{n} b_j(z_1) \mathbf{T} \right)^{-1} \mathbf{r}_i,$$

$$A_2(z_1, z_2) = -\mathrm{tr} \sum_{i<j} \left(z_1 \mathbf{I} - \frac{n-1}{n} b_j(z_1) \mathbf{T} \right)^{-1} n^{-1} \mathbf{T} [\mathrm{E}_j \mathbf{D}_{ij}^{-1}(z_1)] \mathbf{T}$$
$$\times \big(\mathbf{D}_j^{-1}(z_2) - \mathbf{D}_{ij}^{-1}(z_2) \big) \mathbf{T},$$

and

$$A_3(z_1, z_2) = \mathrm{tr} \sum_{i<j} \left(z_1 \mathbf{I} - \frac{n-1}{n} b_j(z_1) \mathbf{T} \right)^{-1} (\mathbf{r}_i \mathbf{r}_i^* - n^{-1} \mathbf{T})$$
$$\times [\mathrm{E}_j \mathbf{D}_{ij}^{-1}(z_1)] \mathbf{T} \mathbf{D}_{ij}^{-1}(z_2) \mathbf{T}.$$

We get, from (9.9.5) and (9.9.13),

$$|A_2(z_1, z_2)| \le K, \tag{9.9.18}$$

and, similar to (9.9.14), we have

$$\mathrm{E}|A_3(z_1, z_2)| \le K n^{1/2}.$$

Using (9.9.3) and (9.9.6), we have, for $i < j$,

$$\mathrm{E} \Big| \beta_{ij}(z_2) \mathbf{r}_i^* [\mathrm{E}_j \mathbf{D}_{ij}^{-1}(z_1)] \mathbf{T} \mathbf{D}_{ij}^{-1}(z_2) \mathbf{r}_i \mathbf{r}_i^* \mathbf{D}_{ij}^{-1}(z_2) \mathbf{T}$$
$$\times \left(z_1 \mathbf{I} - \frac{n-1}{n} b_j(z_1) \mathbf{T} \right)^{-1} \mathbf{r}_i - b_j(z_2) n^{-2} \mathrm{tr}([\mathrm{E}_j \mathbf{D}_{ij}^{-1}(z_1)] \mathbf{T}$$
$$\times \mathbf{D}_{ij}^{-1}(z_2) \mathbf{T}) \mathrm{tr} \left(\mathbf{D}_{ij}^{-1}(z_2) \mathbf{T} \left(z_1 \mathbf{I} - \frac{n-1}{n} b_j(z_1) \mathbf{T} \right)^{-1} \mathbf{T} \right) \Big|$$
$$\le K n^{-1/2}$$

(K now depending as well on z_i and y_n). Using (9.9.5), we have

$$\Big| \mathrm{tr} \left(\left[\mathrm{E}_j \mathbf{D}_{ij}^{-1}(z_1) \right] \mathbf{T} \mathbf{D}_{ij}^{-1}(z_2) \mathbf{T} \right) \mathrm{tr} \left(\mathbf{D}_{ij}^{-1}(z_2) \mathbf{T} \left(z_1 \mathbf{I} - \frac{n-1}{n} b_j(z_1) \mathbf{T} \right)^{-1} \mathbf{T} \right)$$
$$- \mathrm{tr} \left(\left[\mathrm{E}_j \mathbf{D}_j^{-1}(z_1) \right] \mathbf{T} \mathbf{D}_j^{-1}(z_2) \mathbf{T} \right) \mathrm{tr} \left(\mathbf{D}_j^{-1}(z_2) \mathbf{T} \left(z_1 \mathbf{I} - \frac{n-1}{n} b_j(z_1) \mathbf{T} \right)^{-1} \mathbf{T} \right) \Big|$$
$$\le K n.$$

It follows that

$$\mathrm{E} \Big| A_1(z_1, z_2) + \frac{j-1}{n^2} b_j(z_2) \mathrm{tr} \left(\left[\mathrm{E}_j \mathbf{D}_j^{-1}(z_1) \right] \mathbf{T} \mathbf{D}_j^{-1}(z_2) \mathbf{T} \right)$$

$$\times \operatorname{tr}\left(\mathbf{D}_j^{-1}(z_2)\mathbf{T}\left(z_1\mathbf{I} - \frac{n-1}{n}b_j(z_1)\mathbf{T}\right)^{-1}\mathbf{T}\right)\Bigg|$$
$$\le Kn^{1/2}. \tag{9.9.19}$$

Therefore, from (9.9.12) to (9.9.19), we can write

$$\begin{aligned}
&\operatorname{tr}\left(\left[\mathrm{E}_j\mathbf{D}_j^{-1}(z_1)\right]\mathbf{T}\mathbf{D}_j^{-1}(z_2)\mathbf{T}\right)\left[1 + \frac{j-1}{n^2}b_j(z_1)b_j(z_2)\right.\\
&\left.\operatorname{tr}\left(\mathbf{D}_j^{-1}(z_2)\mathbf{T}\left(z_1\mathbf{I} - \frac{n-1}{n}b_j(z_1)\mathbf{T}\right)^{-1}\mathbf{T}\right)\right]\\
&= -\operatorname{tr}\left(\left(z_1\mathbf{I} - \frac{n-1}{n}b_j(z_1)\mathbf{T}\right)^{-1}\mathbf{T}\mathbf{D}_j^{-1}(z_2)\mathbf{T}\right) + A_4(z_1, z_2),
\end{aligned}$$

where

$$\mathrm{E}|A_4(z_1, z_2)| \le Kn^{1/2}.$$

Using the expression for $\mathbf{D}_j^{-1}(z_2)$ in (9.9.12) and (9.9.14)–(9.9.16), we find that

$$\begin{aligned}
&\operatorname{tr}(\mathrm{E}_j(\mathbf{D}_j^{-1}(z_1))\mathbf{T}\mathbf{D}_j^{-1}(z_2)\mathbf{T})\\
&\quad\times\left[1 - \frac{j-1}{n^2}b_j(z_1)b_j(z_2)\operatorname{tr}\left(\left(z_2\mathbf{I} - \frac{n-1}{n}b_j(z_2)\mathbf{T}\right)^{-1}\mathbf{T}\right.\right.\\
&\quad\left.\left.\times\left(z_1\mathbf{I} - \frac{n-1}{n}b_j(z_1)\mathbf{T}\right)^{-1}\mathbf{T}\right)\right]\\
&= \operatorname{tr}\left(\left(z_2\mathbf{I} - \frac{n-1}{n}b_j(z_2)\mathbf{T}\right)^{-1}\mathbf{T}\left(z_1\mathbf{I} - \frac{n-1}{n}b_j(z_1)\mathbf{T}\right)^{-1}\mathbf{T}\right)\\
&\quad + A_5(z_1, z_2),
\end{aligned}$$

where

$$\mathrm{E}|A_5(z_1, z_2)| \le Kn^{1/2}.$$

From (9.9.5), we have

$$|b_j(z) - b(z)| \le Kn^{-1}$$

and

$$|b_j(z) - \mathrm{E}\beta_j(z)| \le Kn^{-1/2}.$$

From the formula

$$\underline{s}_n = -\frac{1}{zn}\sum_{j=1}^{n}\beta_j(z)$$

(see (6.2.4)), we get $\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\beta_j(z) = -z\mathrm{E}\underline{s}_n(z)$. It will be shown in Section 9.11 that

$$|\mathrm{E}\underline{s}_n(z) - \underline{s}_n^0(z)| \le Kn^{-1}.$$

Therefore we have

$$\max_j |b_j(z) + z\underline{s}_n^0(z)| \le Kn^{-1/2}, \tag{9.9.20}$$

so that we can write

$$\begin{aligned}
&\mathrm{tr}(\mathrm{E}_j(\mathbf{D}_j^{-1}(z_1))\mathbf{T}\mathbf{D}_j^{-1}(z_2)\mathbf{T}) \\
&\left[1 - \frac{j-1}{n^2}\underline{s}_n^0(z_1)\underline{s}_n^0(z_2)\mathrm{tr}((\mathbf{I}+\underline{s}_n^0(z_2)\mathbf{T})^{-1}\mathbf{T}(\mathbf{I}+\underline{s}_n^0(z_1)\mathbf{T})^{-1}\mathbf{T})\right] \\
= &\frac{1}{z_1z_2}\mathrm{tr}((\mathbf{I}+\underline{s}_n^0(z_2)\mathbf{T})^{-1}\mathbf{T}(\mathbf{I}+\underline{s}_n^0(z_1)\mathbf{T})^{-1}\mathbf{T}) + A_6(z_1,z_2),
\end{aligned} \tag{9.9.21}$$

where

$$\mathrm{E}|A_6(z_1,z_2)| \le Kn^{1/2}.$$

Rewrite (9.9.21) as

$$\begin{aligned}
&\mathrm{tr}(\mathrm{E}_j(\mathbf{D}_j^{-1}(z_1))\mathbf{T}\mathbf{D}_j^{-1}(z_2)\mathbf{T}) \\
&\left[1 - \frac{j-1}{n}y_n\underline{s}_n^0(z_1)\underline{s}_n^0(z_2)\int\frac{t^2\,dH_n(t)}{(1+t\underline{s}_n^0(z_1))(1+t\underline{s}_n^0(z_2))}\right] \\
= &\frac{ny_n}{z_1z_2}\int\frac{t^2\,dH_n(t)}{(1+t\underline{s}_n^0(z_1))(1+t\underline{s}_n^0(z_2))} + A_6(z_1,z_2).
\end{aligned}$$

Using (6.2.14), (6.2.21), and (9.9.20), we find

$$\begin{aligned}
&\left|y_n\underline{s}_n^0(z_1)\underline{s}_n^0(z_2)\int\frac{t^2\,dH_n(t)}{(1+t\underline{s}_n^0(z_1))(1+t\underline{s}_n^0(z_2))}\right| \\
= &\left|y_n\frac{\int\frac{t^2\,dH_n(t)}{(1+t\underline{s}_n^0(z_1))(1+t\underline{s}_n^0(z_2))}}{\left(-z_1+y_n\int\frac{t\,dH_n(t)}{1+t\underline{s}_n^0(z_1)}\right)\left(-z_2+y_n\int\frac{t\,dH_n(t)}{1+t\underline{s}_n^0(z_2)}\right)}\right| \qquad (9.9.22) \\
\le &\left(y_n\frac{\int\frac{t^2\,dH_n(t)}{|1+t\underline{s}_n^0(z_1)|^2}}{\left|-z_1+y_n\int\frac{t\,dH_n(t)}{1+t\underline{s}_n^0(z_1)}\right|^2}\right)^{1/2}\left(y_n\frac{\int\frac{t^2\,dH_n(t)}{|1+t\underline{s}_n^0(z_2)|^2}}{\left|-z_2+y_n\int\frac{t\,dH_n(t)}{1+t\underline{s}_n^0(z_2)}\right|^2}\right)^{1/2} \\
= &\left(\frac{\Im\,\underline{s}_n^0(z_1)y_n\int\frac{t^2\,dH_n(t)}{|1+t\underline{s}_n^0(z_1)|^2}}{\Im\,z_1+\Im\,\underline{s}_n^0(z_1)y_n\int\frac{t^2\,dH_n(t)}{|1+t\underline{s}_n^0(z_1)|^2}}\right)^{1/2} \\
&\times\left(\frac{\Im\,\underline{s}_n^0(z_2)y_n\int\frac{t^2\,dH_n(t)}{|1+t\underline{s}_n^0(z_2)|^2}}{\Im\,z_2+\Im\,\underline{s}_n^0(z_2)y_n\int\frac{t^2\,dH_n(t)}{|1+t\underline{s}_n^0(z_2)|^2}}\right)^{1/2} < 1 \qquad (9.9.23)
\end{aligned}$$

since

$$\frac{\Im z}{\Im \underline{s}_n^0(z) y_n \int \frac{t^2\, dH_n(t)}{|1+t\underline{s}_n^0(z)|^2}}$$

is bounded away from 0. Therefore, using (9.9.20) and letting $a_n(z_1, z_2)$ denote the expression inside the absolute value sign in (9.9.22), we find that (9.9.11) can be written as

$$a_n(z_1, z_2)\frac{1}{n}\sum_{j=1}^{n} \frac{1}{1 - \frac{j-1}{n} a_n(z_1, z_2)} + A_7(z_1, z_2),$$

where

$$\mathrm{E}|A_7(z_1, z_2)| \le K n^{-1/2}.$$

We see then that

$$(9.9.11) \xrightarrow{\text{i.p.}} a(z_1, z_2)\int_0^1 \frac{1}{1 - ta(z_1, z_2)} dt = \int_0^{a(z_1,z_2)} \frac{1}{1-z} dz,$$

where

$$\begin{aligned}
a(z_1, z_2) &= y\underline{s}(z_1)\underline{s}(z_2)\int \frac{t^2\, dH(t)}{(1 + t\underline{s}(z_1))(1 + t\underline{s}(z_2))} \\
&= \frac{\underline{s}(z_1)\underline{s}(z_2)}{\underline{s}(z_2) - \underline{s}(z_1)}\left(y\int \frac{t\, dH(t)}{1 + t\underline{s}(z_1)} - y\int \frac{t\, dH(t)}{1 + t\underline{s}(z_2)}\right) \\
&= 1 + \frac{\underline{s}(z_1)\underline{s}(z_2)(z_1 - z_2)}{\underline{s}(z_2) - \underline{s}(z_1)}.
\end{aligned}$$

Thus the limit of (9.9.7) in probability under the CSE case is

$$\begin{aligned}
&\frac{\partial^2}{\partial z_2 \partial z_1}\int_0^{a(z_1,z_2)} \frac{1}{1-z} dz = \frac{\partial}{\partial z_2}\left(\frac{\frac{\partial}{\partial z_1} a(z_1, z_2)}{1 - a(z_1, z_2)}\right) \\
&= -\frac{\partial}{\partial z_2}\left(\frac{\underline{s}'(z_1)}{\underline{s}(z_1)} + \frac{1}{z_1 - z_2} + \frac{\underline{s}'(z_1)}{\underline{s}(z_2) - \underline{s}(z_1)}\right) \\
&= \frac{\underline{s}'(z_1)\underline{s}'(z_2)}{(\underline{s}(z_2) - \underline{s}(z_1))^2} - \frac{1}{(z_1 - z_2)^2},
\end{aligned}$$

which is half of (9.8.4).

9.10 Tightness of $M_n^1(z)$

We proceed to prove tightness of the sequence of random functions $\widehat{M}_n^1(z)$ for $z \in \mathcal{C}^+$. We will use Theorem 12.3 (p. 95) of Billingsley [57]. From (9.9.8) we see that the condition in Bi is satisfied. We will verify condition (ii) of Theorem 12.3 by proving the moment condition (12.51) of Billingsley [57]. We will show

$$\sup_{n;z_1,z_2\in\mathcal{C}^+} \frac{\mathrm{E}|\widehat{M}_n^1(z_1)-\widehat{M}_n^1(z_2)|^2}{|z_1-z_2|^2}$$

is finite. It is straightforward to verify that this will be true if we can find a $K>0$ for which

$$\sup_{n;z_1,z_2\in\mathcal{C}_n} \frac{\mathrm{E}|M_n^1(z_1)-M_n^1(z_2)|^2}{|z_1-z_2|^2} \le K.$$

From Bii, the results in this and the next section will establish tightness of $\{\widehat{M}_n\}$.

We claim that moments of $\|\mathbf{D}^{-1}(z)\|$, $\|\mathbf{D}_j^{-1}(z)\|$, and $\|\mathbf{D}_{ij}^{-1}(z)\|$ are bounded in n and $z\in\mathcal{C}_n$. This is clearly true for $z\in\mathcal{C}_u$ and for $z\in\mathcal{C}_l$ if $x_l<0$. For $z\in\mathcal{C}_r$ or $z\in\mathcal{C}_l$, if $x_l>0$, we use (9.7.8), (9.7.9), and (9.8.1) on, for example, $\mathbf{B}_{(1)}=\mathbf{B}_n-\mathbf{r}_1\mathbf{r}_1^*$, to get

$$\begin{aligned}\mathrm{E}\|\mathbf{D}_1^{-1}(z)\|^p &\le K_1+v^{-p}\mathrm{P}\Big(\|\mathbf{B}_{(1)}\|\ge\eta_r \text{ or } \lambda_{\min}^{\mathbf{B}_{(1)}}\le\eta_l\Big)\\ &\le K_1+K_2n^p\varepsilon^{-p}n^{-\ell}\le K\end{aligned}$$

for suitably large ℓ. Here, η_r is any fixed number between $\limsup_n\|\mathbf{T}\|(1+\sqrt{y})^2$ and x_r, and, if $x_l>0$, η_l is any fixed number between x_l and $\liminf_n\lambda_{\min}^{\mathbf{T}}(1-\sqrt{y})^2$ (take $\eta_l<0$ if $x_l<0$). Therefore, for any positive p,

$$\max(\mathrm{E}\|\mathbf{D}^{-1}(z)\|^p,\mathrm{E}\|\mathbf{D}_j^{-1}(z)\|^p,\mathrm{E}\|\mathbf{D}_{ij}^{-1}(z)\|^p)\le K_p. \tag{9.10.1}$$

We can use the argument above to extend (9.9.6). Using (9.7.7) and (9.9.6), we get

$$\begin{aligned}&\left|\mathrm{E}\left(a(\mu,\nu)\prod_{l=1}^q(\mathbf{r}_1^*\mathbf{B}_l(\nu)\mathbf{r}_1-n^{-1}\mathrm{tr}\mathbf{T}\mathbf{B}_l(\nu))\right)\right|\\ &\le Kn^{-(1\wedge q)}\eta_n^{(2q-4)\vee 0},\quad q\ge 0,\end{aligned} \tag{9.10.2}$$

where now the matrices $\mathbf{B}_l(v)$ are independent of $\mathbf{r}_1$ and

$$\|\mathbf{B}_l(\nu)\|\le K(1+n^\nu\mathbf{I}(\|\mathbf{B}_n\|\ge\eta_r \text{ or } \lambda_{\min}^{\widetilde{\mathbf{B}}}\le\eta_l))$$

and

$$|a(\mu,\nu)|\le K\Big(1+|\mathbf{r}|^\mu+n^\nu\mathbf{I}(\|\mathbf{B}_n\|\ge\eta_r \text{ or } \lambda_{\min}^{\widetilde{\mathbf{B}}}\le\eta_l)\Big)$$

for some positive μ and ν, with $\widetilde{\mathbf{B}}$ being $\mathbf{B}_n$, or $\mathbf{B}_n$ with one or two of the $\mathbf{r}_j$'s removed.

We would like to inform the reader that, in applications of (9.10.2), $a(\mu,\nu)$ is a product of factors of the form $\beta_1(z)$ or $\mathbf{r}_1^*\mathbf{A}(z)\mathbf{r}_1$ and $\mathbf{A}$ is a product of one or several $\mathbf{D}_j^{-1}(z)$ $\mathbf{D}_j^{-1}(z_i)$, $i=1,2$, or similarly defined $\mathbf{D}^{-1}$ matrices. The matrices $\mathbf{B}_l$ also have this form. For example, we have, for any $z \in \mathcal{C}_n$,

$$|\mathbf{r}_1^*\mathbf{D}_1^{-1}(z_1)\mathbf{D}_1^{-1}(z_2)\mathbf{r}_1| \le |\mathbf{r}_1|^2\|\mathbf{D}_1^{-1}(z_1)\mathbf{D}_1^{-1}(z_2)\|$$
$$\le K|\mathbf{r}_1|^2 + n^5 I(\|\mathbf{B}_n\| \ge \eta_r \text{ or } \lambda_{\min}^{\mathbf{B}_{(1)}} \le \eta_l),$$

where K can be taken to be $\max((x_r-\eta_r)^{-2},(\eta_l-x_l)^{-2},v_0^{-2})$. We have $\|\mathbf{B}_l\|$ obviously satisfying this condition. Since

$$\mathbf{r}_j^*\mathbf{D}^{-1}\mathbf{r}_j = \frac{\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j}{1+\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j} = 1-\beta_j,$$

we assert that $\beta_j(z)$ also satisfies this condition since

$$|\beta_j(z)| = |1-\mathbf{r}_j^*\mathbf{D}^{-1}\mathbf{r}_j|$$
$$\le 1 + K|\mathbf{r}_j|^2 + n^3 I(\|\mathbf{B}_n\| \ge \eta_r \text{ or } \lambda_{\min}^{\mathbf{B}_n} \le \eta_l).$$

In what follows, we shall freely use (9.10.2) without verifying these conditions, even similarly defined β_j functions and $\mathbf{A}$, $\mathbf{B}$ matrices.

For later use, we have

$$\mathbf{D}^{-1}(z) - \mathbf{D}_j^{-1}(z) = -\frac{\mathbf{D}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)}{1+\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j}$$
$$= -\beta_j(z)\mathbf{D}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z). \tag{9.10.3}$$

Let

$$\gamma_j(z) = \mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j - n^{-1}\mathrm{E}\mathrm{tr}(\mathbf{D}_j^{-1}(z)\mathbf{T}).$$

We first derive bounds on the moments of $\gamma_j(z)$ and $\hat\gamma_j(z)$. Using (9.10.2), we have

$$\mathrm{E}|\hat\gamma_j(z)|^p \le K_p n^{-1}\eta_n^{2p-4}, \quad p \text{ even}. \tag{9.10.4}$$

It should be noted that constants obtained do not depend on $z \in \mathcal{C}_n$.

Using Lemma 2.12, (9.10.2), and Hölder's inequality, we have, for all even p,

$$\mathrm{E}|\gamma_1(z) - \hat\gamma_1(z)|^p$$
$$= \mathrm{E}\left|\frac{1}{n}\sum_{j=2}^{n} \mathrm{E}_j\mathrm{tr}\mathbf{T}\mathbf{D}_1(z)^{-1} - \mathrm{E}_{j-1}\mathrm{tr}\mathbf{T}\mathbf{D}_1^{-1}(z)\right|^p$$
$$= \mathrm{E}\left|\frac{1}{n}\sum_{j=2}^{n} \mathrm{E}_j\mathrm{tr}\mathbf{T}(\mathbf{D}_1^{-1}(z) - \mathbf{D}_{1j}^{-1}(z)) - \mathrm{E}_{j-1}\mathrm{tr}\mathbf{T}(\mathbf{D}_1^{-1}(z) - \mathbf{D}_{1j}^{-1}(z))\right|^p$$

$$= \frac{1}{n^p}\mathrm{E}\left|\sum_{j=2}^{n}(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_{1j}(z)\mathbf{r}_j^*\mathbf{D}_{1j}^{-1}(z)\mathbf{T}\mathbf{D}_{1j}^{-1}(z)\mathbf{r}_j\right|^p$$
$$\le \frac{K_p}{n^p}\mathrm{E}\left(\sum_{j=2}^{n}\left|(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_{1j}(z)\mathbf{r}_j^*\mathbf{D}_{1j}^{-1}(z)\mathbf{T}\mathbf{D}_{1j}^{-1}(z)\mathbf{r}_j\right|^2\right)^{p/2}$$
$$\le \frac{K_p}{n^{1+p/2}}\sum_{j=2}^{n}\mathrm{E}|(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_{1j}(z)\mathbf{r}_j^*\mathbf{D}_{1j}^{-1}(z)\mathbf{T}\mathbf{D}_{1j}^{-1}(z)\mathbf{r}_j|^p$$
$$\le K_p n^{-p/2}\mathrm{E}|\beta_{12}(z)\mathbf{r}_2^*\mathbf{D}_{12}^{-1}(z)\mathbf{T}\mathbf{D}_{12}^{-1}(z)\mathbf{r}_2|^p \le K_p n^{-p/2}.$$

It is easy to see that the inequality above is true for all j; i.e.,

$$\mathrm{E}|\gamma_j - \hat{\gamma}_j|^p \le K_p n^{-p/2}.$$

Therefore

$$\mathrm{E}|\gamma_j|^p \le K_p n^{-1}\eta_n^{2p-4}, \quad p \ge 2. \tag{9.10.5}$$

We next prove that $b_j(z)$ are uniformly bounded for all n and $z \in \mathcal{C}_n$. From (9.10.2), we find, for any $p \ge 1$,

$$\mathrm{E}|\beta_j(z)|^p \le K_p. \tag{9.10.6}$$

Since $b_j = \beta_j(z) + \beta_j(z)b_j(z)\gamma_j(z)$, we get from (9.10.5) and (9.10.6),

$$|b_j(z)| = |\mathrm{E}\beta_j(z) + \mathrm{E}\beta_j(z)b_j(z)\gamma_1(z)| \le K_1 + K_2|b_j(z)|n^{-1/2}.$$

Thus, for all large n,

$$|b_j(z)| \le \frac{K_1}{1 - K_2 n^{-1/2}},$$

and subsequently $b_j(z)$ is bounded for all n.

From (9.10.3), we have

$$\begin{aligned}
&\mathbf{D}^{-1}(z_1)\mathbf{D}^{-1}(z_2) - \mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\\
&= (\mathbf{D}^{-1}(z_1) - \mathbf{D}_j^{-1}(z_1))(\mathbf{D}^{-1}(z_2) - \mathbf{D}_j^{-1}(z_2))\\
&\quad +(\mathbf{D}^{-1}(z_1) - \mathbf{D}_j^{-1}(z_1))\mathbf{D}_j^{-1}(z_2) + \mathbf{D}_j^{-1}(z_1)(\mathbf{D}^{-1}(z_2) - \mathbf{D}_j^{-1}(z_2))\\
&= \beta_j(z_1)\beta_j(z_2)\mathbf{D}_j^{-1}(z_1)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_2)\\
&\quad -\beta_j(z_1)\mathbf{D}_j^{-1}(z_1)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2) - \beta_j(z_2)\mathbf{D}_j^{-1}(z_1)\\
&\qquad \times\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_2).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathrm{tr}\left(\mathbf{D}^{-1}(z_1)\mathbf{D}^{-1}(z_2) - \mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\right)\\
&= \beta_j(z_1)\beta_j(z_2)(\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j)^2 - \beta_j(z_1)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j
\end{aligned}$$

$$-\beta_j(z_2)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z_2)\mathbf{D}_j^{-1}(z_1)\mathbf{r}_j. \tag{9.10.7}$$

We write

$$\begin{aligned} s_n(z_1) - s_n(z_2) &= \frac{1}{p}\mathrm{tr}(\mathbf{D}^{-1}(z_1) - \mathbf{D}^{-1}(z_2)) \\ &= \frac{1}{p}(z_1 - z_2)\mathrm{tr}\mathbf{D}^{-1}(z_1)\mathbf{D}^{-1}(z_2). \end{aligned}$$

Therefore, from (9.10.7), we have

$$\begin{aligned} & p\frac{s_n(z_1) - s_n(z_2) - \mathrm{E}(s_n(z_1) - s_n(z_2))}{z_1 - z_2} \\ &= \sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1})\mathrm{tr}\mathbf{D}^{-1}(z_1)\mathbf{D}^{-1}(z_2) \\ &= \sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_1)\beta_j(z_2)(\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j)^2 \\ &\quad - \sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_1)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j \\ &\quad - \sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_2)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z_2)\mathbf{D}_j^{-1}(z_1)\mathbf{r}_j. \end{aligned} \tag{9.10.8}$$

Our goal is to show that the absolute second moment of (9.10.8) is bounded. We begin with the second sum in (9.10.8). We have

$$\begin{aligned} & \sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_1)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j \\ &= \sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1})\big(b_j(z_1)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j \\ &\qquad - \beta_j(z_1)b_j(z_1)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j\gamma_j(z_1)\big) \\ &= \sum_{j=1}^n b_j(z_1)\mathrm{E}_j(\mathbf{r}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j \\ &\qquad - n^{-1}\mathrm{tr}\mathbf{T}^{1/2}\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}^{1/2}) \\ &\quad - \sum_{j=1}^n b_j(z_1)(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_1)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j\gamma_j(z_1) \\ &\equiv W_1 - W_2. \end{aligned}$$

Using (9.10.2), we have

$$\mathrm{E}|W_1|^2 = \sum_{j=1}^{n} |b_j(z_1)|^2 \mathrm{E}|\mathrm{E}_j(\mathbf{r}_j^* \mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j \\ -n^{-1}\mathrm{tr}\mathbf{T}^{1/2}\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}^{1/2})|^2 \le K.$$

Using (9.10.5) and the bounds for $\beta_1(z_1)$ and $\mathbf{r}_1^*\mathbf{D}_1^{-2}(z_1)\mathbf{D}_1^{-1}(z_2)\mathbf{r}_1$ given in the remark to (9.10.2), we have

$$\begin{aligned} &\mathrm{E}|W_2|^2 \\ &= \sum_{j=1}^{n} |b_j(z_1)|^2 \mathrm{E}|(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_1)\mathbf{r}_j^*\mathbf{D}_j^{-2}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j\gamma_j(z_1)|^2 \\ &\le Kn[\mathrm{E}|\gamma_1(z_1)|^2 + v^{-10}p^2 P(\|\mathbf{B}_n\| > \eta_r \text{ or } \lambda_{\min}^{\mathbf{B}_{(1)}} < \eta_l)] \le K. \end{aligned}$$

This argument of course handles the third sum in (9.10.8).

For the first sum in (9.10.8), we have

$$\begin{aligned} &\sum_{j=1}^{n}(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_1)\beta_j(z_2)(\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_j)^2 \\ &= \sum_{j=1}^{n} b_j(z_1)b_j(z_2)(\mathrm{E}_j - \mathrm{E}_{j-1})[(\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j)^2 \\ &\quad -(n^{-1}\mathrm{tr}\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}^{1/2})^2] \\ &\quad -\sum_{j=1}^{n} b_j(z_2)(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_1)\beta_j(z_2)(\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_j)^2\gamma_j(z_2) \\ &\quad -\sum_{j=1}^{n} b_j(z_1)b_j(z_2)(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_1)(\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_{ij}^{-1}(z_2)\mathbf{r}_j)^2\gamma_j(z_1) \\ &= Y_1 - Y_2 - Y_3. \end{aligned}$$

Both Y_2 and Y_3 are handled the same way as W_2 above. Using (9.10.2), we have

$$\begin{aligned} &\mathrm{E}|Y_1|^2 \\ &\le K\sum_{j=1}^{n}\mathrm{E}\left|(\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j)^2 - \left(\frac{1}{n}\mathrm{tr}\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}^{1/2}\right)^2\right|^2 \\ &\le K\sum_{j=1}^{n}2\mathrm{E}|\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j - \frac{1}{n}\mathrm{tr}\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}^{1/2}|^4 \\ &\quad +Ky_n^2\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\left|\left(\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j - \frac{1}{n}\mathrm{tr}\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{T}^{1/2}\right)\right. \end{aligned}$$

$$\times \quad \|\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\|\Big|^2$$
$$\leq K.$$

Therefore, condition (ii) of Theorem 12.3 of Billingsley [57] is satisfied, and we conclude that $\{\widehat{M}_n^1(z)\}$ is tight.

9.11 Convergence of $M_n^2(z)$

The proof of Lemma 9.11 is completed with the verification of $\{\widehat{M}_n^2(z)\}$ for $z \in \mathcal{C}^+$ to be bounded and forms a uniformly equicontinuous family and convergence to (9.8.3) under the assumptions in (2) and to zero under those in (3). As in the previous section, it is enough to verify these statements on $\{M_n^2(z)\}$ for $z \in \mathcal{C}_n$.

Similar to (6.2.25), we have

$$\begin{aligned}
&(\mathrm{E}\underline{s}_n - \underline{s}_n^0)\left(1 - \frac{y_n \int \frac{t^2\,dH_n(t)}{(1+t\mathrm{E}\underline{s}_n)(1+t\underline{s}_n^0)}}{\left(-z + y_n \int \frac{t\,dH_n(t)}{1+t\mathrm{E}\underline{s}_n} - R_n\right)\left(-z + y_n \int \frac{t\,dH_n(t)}{1+t\underline{s}_n^0}\right)}\right)\\
&= (\mathrm{E}\underline{s}_n - \underline{s}_n^0)\left(1 - \frac{y_n \int \frac{\underline{s}_n^0 t^2\,dH_n(t)}{(1+t\mathrm{E}\underline{s}_n)(1+t\underline{s}_n^0)}}{-z + y_n \int \frac{t\,dH_n(t)}{1+t\mathrm{E}\underline{s}_n} - R_n}\right)\\
&= \mathrm{E}\underline{s}_n \underline{s}_n^0 R_n, \qquad (9.11.1)
\end{aligned}$$

where $R_n = y_n n^{-1} \sum_{j=1}^n \mathrm{E}\beta_j d_j (\mathrm{E}\underline{s}_n)^{-1}$,

$$\begin{aligned}
d_j = d_j(z) &= -\mathbf{q}_j^* \mathbf{T}^{1/2} (\mathbf{B}_{(j)} - z\mathbf{I})^{-1} (\mathrm{E}\underline{s}_n \mathbf{T} + \mathbf{I})^{-1} \mathbf{T}^{1/2} \mathbf{q}_j\\
&\quad + (1/p)\mathrm{tr}(\mathrm{E}\underline{s}_n \mathbf{T} + \mathbf{I})^{-1} \mathbf{T} (\mathbf{B}_n - z\mathbf{I})^{-1},\\
\beta_j^{-1} &= 1 + \mathbf{r}_j^* (\mathbf{B}_{(j)} - z\mathbf{I})^{-1} \mathbf{r}_j.
\end{aligned}$$

Thus, by noting $M_n^2(z) = n(\mathrm{E}\underline{s}_n(z) - \underline{s}_n^0)$, to prove (9.8.3) it suffices to show

$$\frac{y_n \int \frac{\underline{s}_n^0 t^2\,dH_n(t)}{(1+t\mathrm{E}\underline{s}_n)(1+t\underline{s}_n^0)}}{-z + y_n \int \frac{t\,dH_n(t)}{1+t\mathrm{E}\underline{s}_n} - R_n} \to y \int \frac{\underline{s}^2 t^2 dH(t)}{(1+t\underline{s})^2} \qquad (9.11.2)$$

and

$$y_n \sum_{j=1}^n \mathrm{E}\beta_j d_j \to \begin{cases} \dfrac{y \int \frac{\underline{s}^2 t^2\,dH(t)}{(1+t\underline{s})^3}}{1 - y \int \frac{\underline{s}^2 t^2\,dH(t)}{(1+t\underline{s})^2}}, & \text{for the RSE case,} \\ 0, & \text{for the CSE case,} \end{cases} \qquad (9.11.3)$$

uniformly for $z \in \mathcal{C}_n$.

To prove the two assertions above, we first prove

$$\sup_{z\in\mathcal{C}_n} |\mathrm{E}\underline{s}_n(z) - \underline{s}(z)| \to 0 \quad \text{as } n \to \infty. \tag{9.11.4}$$

In order to simplify the exposition, we let $\mathcal{C}_1 = \mathcal{C}_u$ or $\mathcal{C}_u \cup \mathcal{C}_l$ if $x_l < 0$ and $\mathcal{C}_2 = \mathcal{C}_2(n) = \mathcal{C}_r$ or $\mathcal{C}_r \cup \mathcal{C}_l$ if $x_l > 0$.

Since $F^{\underline{\mathbf{B}}_n} \xrightarrow{\mathscr{D}} F^{y,H}$ almost surely, we get from the dct (dominated convergence theorem) $\mathrm{E}F^{\underline{\mathbf{B}}_n} \xrightarrow{\mathscr{D}} F^{y,H}$. It is easy to verify that $\mathrm{E}F^{\underline{\mathbf{B}}_n}$ is a proper c.d.f. Since, as z ranges in $\mathcal{C}_1$, the functions $(\lambda - z)^{-1}$ in $\lambda \in [0,\infty)$ form a bounded, equicontinuous family, it follows (see, e.g. , Billingsley [57], Problem 8, p. 17) that

$$\sup_{z\in\mathcal{C}_1} |\mathrm{E}\underline{s}_n(z) - \underline{s}(z)| \to 0.$$

For $z \in \mathcal{C}_2$, we write (η_l, η_r defined as in Section 9.10)

$$\begin{aligned}\mathrm{E}\underline{s}_n(z) - \underline{s}(z) = &\int \frac{1}{\lambda - z} I_{[\eta_l,\eta_r]}(\lambda) d(\mathrm{E}F^{\underline{\mathbf{B}}_n}(\lambda) - F^{y,H}(\lambda)) \\ &+\mathrm{E}\int \frac{1}{\lambda - z} I_{[\eta_l,\eta_r]^c}(\lambda) dF^{\underline{\mathbf{B}}_n}(\lambda).\end{aligned}$$

As above, the first term converges uniformly to zero. For the second term, we use (9.7.8) and (9.7.9) with $\ell \geq 2$ to get

$$\begin{aligned}&\sup_{z\in\mathcal{C}_2} \left|\mathrm{E}\int \frac{1}{\lambda - z} I_{[\eta_l,\eta_r]^c}(\lambda) dF^{\underline{\mathbf{B}}_n}(\lambda)\right| \\ &\leq (\varepsilon_n/n)^{-1}\mathrm{P}(\|\mathbf{B}_n\| \geq \eta_r \text{ or } \lambda_{\min}^{\mathbf{B}_n} \leq \eta_l) \leq Kn\varepsilon_n^{-1}n^{-\ell} \to 0.\end{aligned}$$

Thus (9.11.4) holds.

From the fact that $F^{y_n,H_n} \xrightarrow{\mathscr{D}} F^{y,H}$ along with the fact that $\mathcal{C}$ lies outside the support of $F^{y,H}$, it is straightforward to verify that

$$\sup_{z\in\mathcal{C}} |\underline{s}_n^0(z) - \underline{s}(z)| \to 0 \quad \text{as } n \to \infty. \tag{9.11.5}$$

We then show that, for some constant K,

$$\sup_{n\, z\in\mathcal{C}_n} \|(\mathrm{E}\underline{s}_n(z)\mathbf{T} + \mathbf{I})^{-1}\| < K. \tag{9.11.6}$$

From Lemma 6.10(a), $\|(\mathrm{E}\underline{s}_n(z)\mathbf{T} + \mathbf{I})^{-1}\|$ is bounded by $\max(2, 4v_0^{-1})$ on $\mathcal{C}_u$. Let $x = x_l$ or x_r. Since x is outside the support of $\underline{F}^{y,H}$, it follows from Lemma 6.1 and equation (6.1.6), for any t in the support of H, that $\underline{s}(x)t + 1 \neq 0$. Choose any t_0 in the support of H. Since $\underline{s}(z)$ is continuous on $\mathcal{C}^0 \equiv \{x + iv : v \in [0, v_0]\}$, there exist positive constants δ_1 and μ_0 such that

$$\inf_{z\in\mathcal{C}^0} |\underline{s}(z)t_0 + 1| > \delta_1 \quad \text{and} \quad \sup_{z\in\mathcal{C}^0} |\underline{s}(z)| < \mu_0.$$

Using $H_n \xrightarrow{\mathscr{D}} H$ and (9.11.4), for all large n, there exists an eigenvalue $\lambda^{\mathbf{T}}$ of $\mathbf{T}$ such that $|\lambda^{\mathbf{T}} - t_0| < \delta_1/4\mu_0$ and $\sup_{z\in\mathcal{C}_l\cup\mathcal{C}_r} |\mathrm{E}\underline{s}_n(z) - \underline{s}(z)| < \delta_1/4$. Therefore, we have

$$\inf_{z\in\mathcal{C}_l\cup\mathcal{C}_r} |\mathrm{E}\underline{s}_n(z)\lambda^{\mathbf{T}} + 1| > \delta_1/2,$$

which completes the proof of (9.11.6).

Assuming (9.11.3), with (9.12.1) given later, we see that $\sup_{z\in\mathcal{C}_n} |R_n| \to 0$. This, (9.12.1), (9.11.4)–(9.11.6), and the dct imply the truth of (9.11.2). Therefore, our task remains to prove (9.11.3).

Using the identity $\beta_j = \bar\beta_j - \bar\beta_j^2\hat\gamma_j + \bar\beta_j^2\beta_j\hat\gamma_j^2$ and (9.10.2), we have

$$\begin{aligned}
y_n\sum_{j=1}^n \mathrm{E}\beta_j d_j &= -y_n\sum_{j=1}^n \mathrm{E}\beta_j\Big[\mathbf{q}_j^*\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}^{1/2}\mathbf{q}_j \\
&\quad -\frac{1}{p}\mathrm{tr}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_j^{-1}\Big] \\
&\quad +\frac{1}{n}\mathrm{E}\Big[\beta_j\mathrm{tr}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}(\mathbf{D}^{-1}-\mathbf{D}_j^{-1})\Big] \\
&= y_n\sum_{j=1}^n \mathrm{E}\bar\beta_j^2\Big[\mathbf{q}_j^*\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}^{1/2}\mathbf{q}_j \\
&\quad -\frac{1}{p}\mathrm{tr}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_j^{-1}\Big]\hat\gamma_j \\
&\quad -\frac{1}{n}\mathrm{E}\beta_j^2\mathbf{r}_j^*\mathbf{D}_j^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_j^{-1}\mathbf{r}_j + o(1).
\end{aligned}$$

Using (9.10.2), it can be proven that all of β_j, $\bar\beta_j$, and b_j and similarly defined quantities can be replaced by $-z\underline{s}(z)$. Thus we have

$$\begin{aligned}
&\frac{1}{n}\mathrm{E}\beta_j^2\mathbf{r}^*\mathbf{D}_j^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_j^{-1}\mathbf{r}_j \\
&= \frac{z^2\underline{s}^2}{n^2}\mathrm{Etr}\mathbf{D}_j^{-1}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_j^{-1}\mathbf{T} + o(1).
\end{aligned} \tag{9.11.7}$$

Now, assume the assumptions for CSE hold. By (9.8.6) and (9.10.2),

$$\begin{aligned}
&-y_n\sum_{j=1}^n \mathrm{E}\bar\beta_j^2\Big[\mathbf{q}_j^*\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}^{1/2}\mathbf{q}_j \\
&-\frac{1}{p}\mathrm{tr}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_j^{-1}\Big]\hat\gamma_j
\end{aligned}$$

$$= -\frac{z^2\underline{s}^2}{n^2}\sum_{j=1}^{n} \mathrm{E}\mathrm{tr}\mathbf{D}_j^{-1}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_j^{-1}\mathbf{T} + o(1). \tag{9.11.8}$$

This proves (9.8.3) for the CSE case.

Now, assume the conditions for the RSE case hold. Let us continue to derive the limit for $y_n \sum_{j=1}^n \mathrm{E}\beta_j d_j$. By (9.8.6), we have

$$y_n \sum_{j=1}^{n} \mathrm{E}\beta_j d_j = \frac{z^2\underline{s}^2}{n^2}\sum_{j=1}^{n} \mathrm{E}\mathrm{tr}\mathbf{D}_j^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_j^{-1}\mathbf{T} + o(1)$$

$$= \frac{z^2\underline{s}^2}{n^2}\sum_{j=1}^{n} \mathrm{E}\mathrm{tr}\mathbf{D}_j^{-1}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_j^{-1}\mathbf{T} + o(1). \tag{9.11.9}$$

Using the decomposition (9.9.12) and estimates given there, we have

$$y_n \sum_{j=1}^{n} \mathrm{E}\beta_j d_j = \frac{\underline{s}^2}{n}\mathrm{E}\mathrm{tr}(\underline{s}\mathbf{T}+\mathbf{I})^{-3}\mathbf{T}^2 + \frac{z^4\underline{s}^4}{n^2}\sum_{j=1}^{n} \mathrm{E}\mathrm{tr}\mathbf{A}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{A}\mathbf{T} + o(1), \tag{9.11.10}$$

where

$$\mathbf{A} = \sum_{i\neq j}^{n}\left(z\mathbf{I} - \frac{n-1}{n}b_j(z)\mathbf{T}\right)^{-1}\left[\mathbf{r}_i\mathbf{r}_i^* - \frac{1}{n}\mathbf{T}\right]\mathbf{D}_{i,j}^{-1}$$
$$= \sum_{i\neq j}^{n}\mathbf{D}_{i,j}^{-1}\left[\mathbf{r}_i\mathbf{r}_i^* - \frac{1}{n}\mathbf{T}\right]\left(z\mathbf{I} - \frac{n-1}{n}b_j(z)\mathbf{T}\right)^{-1},$$

where the equivalence of the two expressions of $\mathbf{A}$ above can be seen from the fact that $\mathbf{A}(z) = (\mathbf{A}(\bar{z}))^*$. Substituting the first expression of $\mathbf{A}$ into (9.11.10) for the $\mathbf{A}$ on the left and the second expression for the $\mathbf{A}$ on the right, and noting that $\frac{n-1}{n}b_j(z)$ can be replaced by $-z\underline{s}$, inducing a negligible error uniformly on $\mathcal{C}^+$, we obtain

$$\frac{z^4\underline{s}^4}{n^2}\sum_{j=1}^{n} \mathrm{E}\mathrm{tr}\mathbf{A}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{A}\mathbf{T}$$
$$= \frac{z^2\underline{s}^4}{n^2}\sum_{j=1}^{n}\sum_{i,\ell\neq j} \mathrm{E}\mathrm{tr}(\underline{s}\mathbf{T}+\mathbf{I})^{-2}\mathbf{T}\left(\mathbf{r}_i\mathbf{r}_i^* - \frac{1}{n}\mathbf{T}\right)\mathbf{D}_{i,j}^{-1}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}$$
$$\times\mathbf{D}_{\ell,j}^{-1}\left(\mathbf{r}_\ell\mathbf{r}_\ell^* - \frac{1}{n}\mathbf{T}\right) + o(1). \tag{9.11.11}$$

We claim that the sum of cross terms in (9.11.11) is negligible. Note that the cross terms will be 0 if either $\mathbf{D}_{i,j}$ or $\mathbf{D}_{\ell,j}$ is replaced by $\mathbf{D}_{\ell,i,j}$, where

$$\mathbf{D}_{\ell,i,j} = \mathbf{D}_{i,j} - \mathbf{r}_\ell \mathbf{r}_\ell^* = \mathbf{D}_{\ell,j} - \mathbf{r}_i \mathbf{r}_i^*.$$

Therefore, our assertion follows by the following estimate.

For $i \neq \ell$, by (9.10.2),

$$\begin{aligned}
&\left|\mathrm{Etr}(\underline{s}\mathbf{T}+\mathbf{I})^{-2}\mathbf{T}\Big(\mathbf{r}_i\mathbf{r}_i^* - \frac{1}{n}\mathbf{T}\Big)(\mathbf{D}_{i,j}^{-1} - \mathbf{D}_{i,\ell,j}^{-1})(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\right.\\
&\quad\left.\times(\mathbf{D}_{\ell,j} - \mathbf{D}_{i,\ell,j}^{-1})\Big(\mathbf{r}_\ell\mathbf{r}_\ell^* - \frac{1}{n}\mathbf{T}\Big)\right|\\
&= \left|\mathrm{Etr}(\underline{s}\mathbf{T}+\mathbf{I})^{-2}\mathbf{T}\Big(\mathbf{r}_i\mathbf{r}_i^* - \frac{1}{n}\mathbf{T}\Big)\beta_{i,j,\ell}(\mathbf{D}_{i,\ell,j}^{-1}\mathbf{r}_\ell\mathbf{r}_\ell^*\mathbf{D}_{i,\ell,j}^{-1})(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\right.\\
&\quad\left.\times\beta_{j,\ell,i}\mathbf{D}_{i,\ell,j}^{-1}\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{i,\ell,j}^{-1}\Big(\mathbf{r}_\ell\mathbf{r}_\ell^* - \frac{1}{n}\mathbf{T}\Big)\right|\\
&= \left|\mathrm{Etr}(\underline{s}\mathbf{T}+\mathbf{I})^{-2}\mathbf{T}\Big(\mathbf{r}_i\mathbf{r}_i^* - \frac{1}{n}\mathbf{T}\Big)(\mathbf{D}_{i,\ell,j}^{-1}\mathbf{r}_\ell\mathbf{r}_\ell^*\mathbf{D}_{i,\ell,j}^{-1})(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\right.\\
&\quad\left.\times\mathbf{D}_{i,\ell,j}^{-1}\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{i,\ell,j}^{-1}\Big(\mathbf{r}_\ell\mathbf{r}_\ell^* - \frac{1}{n}\mathbf{T}\Big)\varepsilon_{j,\ell,i}\varepsilon_{j,i,\ell}\beta_{j,\ell,i}\beta_{j,i,\ell}\bar{\beta}_{j,i,\ell}^2\right|\\
&= \mathrm{E}^{1/4}\left|\mathbf{r}_i^*\mathbf{D}_{i,\ell,j}^{-1}\Big(\mathbf{r}_\ell\mathbf{r}_\ell^* - \frac{1}{n}\mathbf{T}\Big)(\underline{s}\mathbf{T}+\mathbf{I})^{-2}\mathbf{T}\Big(\mathbf{r}_i\mathbf{r}_i^* - \frac{1}{n}\mathbf{T}\Big)\mathbf{D}_{i,\ell,j}^{-1}\mathbf{r}_\ell\right|^4\\
&\quad\times\mathrm{E}^{1/4}|\mathbf{r}_\ell^*\mathbf{D}_{i,\ell,j}^{-1}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{D}_{i,\ell,j}^{-1}\mathbf{r}_i|^4 \times o(n^{-1/2}) = o(n^{-1}),
\end{aligned}$$

where

$$\begin{aligned}
\beta_{j,i,\ell}^{-1} &= 1 + \mathbf{r}_\ell^*\mathbf{D}_{i,\ell,j}^{-1}\mathbf{r}_\ell,\\
\bar{\beta}_{j,i,\ell}^{-1} &= 1 + \frac{1}{n}\mathrm{tr}\mathbf{D}_{i,\ell,j}^{-1}\mathbf{T},\\
\varepsilon_{j,i,\ell} &= \mathbf{r}_\ell^*\mathbf{D}_{i,\ell,j}^{-1}\mathbf{r}_\ell - \frac{1}{n}\mathrm{tr}\mathbf{D}_{i,\ell,j}^{-1}.
\end{aligned}$$

Here, we have used the fact that with Hermitian $\mathbf{M}$ independent of $\mathbf{r}_i$ and $\mathbf{r}_\ell$,

$$\begin{aligned}
&\mathrm{E}|\mathbf{r}_i^*\mathbf{M}\mathbf{r}_\ell|^4 \le K\mathrm{E}(\mathbf{r}_i^*\mathbf{M}^2\mathbf{r}_i)^2\\
&\le K[(\mathrm{tr}\mathbf{M}^2)^2 + \mathrm{tr}\mathbf{M}^4],
\end{aligned}$$

and by estimation term by term in the expansion,

$$\mathrm{E}\left|\mathbf{r}_i^*\mathbf{D}_{i,\ell,j}^{-1}\Big(\mathbf{r}_\ell\mathbf{r}_\ell^* - \frac{1}{n}\mathbf{T}\Big)(\underline{s}\mathbf{T}+\mathbf{I})^{-2}\mathbf{T}\Big(\mathbf{r}_i\mathbf{r}_i^* - \frac{1}{n}\mathbf{T}\Big)\mathbf{D}_{i,\ell,j}^{-1}\mathbf{r}_\ell\right|^4 \le K.$$

Hence, we have proved

$$
\begin{aligned}
& y_n \sum_{j=1}^{n} \mathrm{E}\beta_j d_j \\
&= \frac{\underline{s}^2}{n}\mathrm{Etr}(\underline{s}\mathbf{T}+\mathbf{I})^{-3}\mathbf{T}^2 + \frac{z^2\underline{s}^4}{n^2}\sum_{i\neq j}\mathrm{Etr}(\underline{s}\mathbf{T}+\mathbf{I})^{-2}\mathbf{T}\left(\mathbf{r}_i\mathbf{r}_i^* - \frac{1}{n}\mathbf{T}\right) \\
&\quad \times \mathbf{D}_{i,j}^{-1}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{D}_{i,j}^{-1}\left(\mathbf{r}_i\mathbf{r}_i^* - \frac{1}{n}\mathbf{T}\right) + o(1) \\
&= \frac{\underline{s}^2}{n}\mathrm{Etr}(\underline{s}\mathbf{T}+\mathbf{I})^{-3}\mathbf{T}^2 \\
&\quad + \frac{z^2\underline{s}^4}{n^2}\sum_{i\neq j}\mathrm{Etr}(\underline{s}\mathbf{T}+\mathbf{I})^{-2}\mathbf{T}\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{D}_{ij}^{-1}\mathbf{r}_i\mathbf{r}_i^* + o(1) \\
&= \frac{\underline{s}^2}{n}\mathrm{Etr}(\underline{s}\mathbf{T}+\mathbf{I})^{-3}\mathbf{T}^2 + \frac{z^2\underline{s}^4}{n^4}\sum_{i\neq j}\mathrm{tr}(\underline{s}\mathbf{T}+\mathbf{I})^{-2}\mathbf{T}^2 \\
&\quad \times \mathrm{tr}\mathbf{D}_{ij}^{-1}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{D}_{ij}^{-1}\mathbf{T} + o(1) \\
&\quad = \frac{\underline{s}^2}{n}\mathrm{Etr}(\underline{s}\mathbf{T}+\mathbf{I})^{-3}\mathbf{T}^2 + \frac{z^2\underline{s}^4}{n^3}\sum_{j=1}^{n}\mathrm{tr}(\underline{s}\mathbf{T}+\mathbf{I})^{-2}\mathbf{T}^2 \\
&\quad \times \mathrm{tr}\mathbf{D}_j^{-1}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{D}_j^{-1}\mathbf{T} + o(1).
\end{aligned} \tag{9.11.12}
$$

Recalling (9.11.9), we obtain

$$
\begin{aligned}
y_n \sum_{j=1}^{n} \mathrm{E}\beta_j d_j &= \frac{\frac{\underline{s}^2}{n}\mathrm{Etr}(\underline{s}\mathbf{T}+\mathbf{I})^{-3}\mathbf{T}^2}{1 - \frac{\underline{s}^2}{n}\mathrm{tr}((\underline{s}\mathbf{T}+\mathbf{I})^{-2}\mathbf{T}^2)} + o(1) \\
&= \frac{y\int \frac{\underline{s}^2 t^2 dH(t)}{(1+t\underline{s})^3}}{1 - y\int \frac{t^2\underline{s}^2 dH(t)}{(1+t\underline{s})^2}} + o(1).
\end{aligned} \tag{9.11.13}
$$

Therefore, we conclude that in the RSE case

$$
\sup_{z\in\mathcal{C}_n}\left| M_n^2(z) - \frac{y\int \frac{\underline{s}(z)^3 t^2 dH(t)}{(1+t\underline{s}(z))^3}}{\left(1 - y\int \frac{\underline{s}(z)^2 t^2 dH(t)}{(1+t\underline{s}(z))^2}\right)^2}\right| \to 0 \quad \text{as } n\to\infty.
$$

Therefore we get (9.8.3).

Finally, for general standardized x_{ij}, we see that in light of the work above, in order to show that $\{M_n^2(z)\}$ for $z \in \mathcal{C}_n$ is bounded and equicontinuous, it is sufficient to prove that $\{f_n'(z)\}$ is bounded, where

$$
\begin{aligned}
f_n(z) \equiv \sum_{j=1}^{n} \mathrm{E}[&(\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j - n^{-1}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{T})(\mathbf{r}_j^*\mathbf{D}_j^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{r}_j \\
&- n^{-1}\mathrm{tr}\mathbf{D}_j^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{T})].
\end{aligned}
$$

Using (9.9.6), we find

$$|f_n'(z)| \le Kn^{-2}\sum_{j=1}^{n}\Big((\mathrm{E}(\mathrm{tr}\mathbf{D}_j^{-2}\mathbf{T}\overline{\mathbf{D}}_j^{-2}\mathbf{T})\mathrm{E}(\mathrm{tr}\mathbf{D}_j^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}$$
$$\mathbf{T}(\mathrm{E}\overline{\underline{s}}_n\mathbf{T}+\mathbf{I})^{-1}\overline{\mathbf{D}}_j^{-1}\mathbf{T}))^{1/2} + (\mathrm{E}(\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{T}\overline{\mathbf{D}}_j^{-1}\mathbf{T})$$
$$\times\mathrm{E}(\mathrm{tr}\mathbf{D}_j^{-2}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}(\mathrm{E}\overline{\underline{s}}_n\mathbf{T}+\mathbf{I})^{-1}\overline{\mathbf{D}}_j^{-2}\mathbf{T}))^{1/2}$$
$$+|\mathrm{E}\underline{s}_n'|(\mathrm{E}(\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{T}\overline{\mathbf{D}}_j^{-1}\mathbf{T})\mathrm{E}(\mathrm{tr}\mathbf{D}_j^{-1}(\mathrm{E}\underline{s}_n\mathbf{T}+\mathbf{I})^{-2}$$
$$\times\mathbf{T}^3(\mathrm{E}\overline{\underline{s}}_n\mathbf{T}+\mathbf{I})^{-2}\overline{\mathbf{D}}_j^{-1}\mathbf{T}))^{1/2}\Big).$$

Using the same argument that resulted in (9.10.1), it is a simple matter to conclude that $\mathrm{E}\underline{s}_n'(z)$ is bounded for $z \in \mathcal{C}_n$. All the remaining expected values are $O(n)$ due to (9.10.1) and (9.11.6), and we are done.

9.12 Some Derivations and Calculations

This section contains proofs of formulas stated in Section 9.8. We begin by deriving some properties of $\underline{s}(z)$.

9.12.1 Verification of (9.8.8)

We claim that, for any bounded subset S of $\mathbb{C}^+$,

$$\inf_{z\in S} |\underline{s}(z)| > 0. \tag{9.12.1}$$

Suppose not. Then there exists a sequence $\{z_n\} \subset \mathbb{C}^+$ that converges to a number for which $\underline{s}(z_n) \to 0$. From (9.7.1), we must have

$$y\int \frac{t\underline{s}(z_n)}{1+t\underline{s}(z_n)}dH(t) \to 1.$$

But, because H has bounded support, the limit of the left-hand side of the above is obviously 0. The contradiction proves our assertion.

Next, we find a lower bound on the size of the difference quotient $(\underline{s}(z_1) - \underline{s}(z_2))/(z_1 - z_2)$ for distinct $z_1 = x + iv_1$, $z_2 = y + iv_2$, $v_1, v_2 \neq 0$. From (9.7.1), we get

$$z_1 - z_2 = \frac{\underline{s}(z_1) - \underline{s}(z_2)}{\underline{s}(z_1)\underline{s}(z_2)}\left(1 - y\int \frac{\underline{s}(z_1)\underline{s}(z_2)t^2 dH(t)}{(1+t\underline{s}(z_1))(1+t\underline{s}(z_2))}\right).$$

Therefore, from (9.9.22) we can write

$$\frac{\underline{s}(z_1)-\underline{s}(z_2)}{z_1-z_2}=\frac{\underline{s}(z_1)\underline{s}(z_2)}{1-y\int\frac{\underline{s}(z_1)\underline{s}(z_2)t^2dH(t)}{(1+t\underline{s}(z_1))(1+t\underline{s}(z_2))}}$$

and conclude that

$$\left|\frac{\underline{s}(z_1)-\underline{s}(z_2)}{z_1-z_2}\right|\geq\frac{1}{2}|\underline{s}(z_1)\underline{s}(z_2)|. \tag{9.12.2}$$

We proceed to show (9.8.8). Choose $f, g \in \{f_1, \cdots, f_k\}$. Let S_F denote the support of $F^{y,H}$, and let $a \neq 0$, b be such that S_F is a subset of (a, b) on whose closure f and g are analytic.

Assume the z_1 contour encloses the z_2 contour. Using integration by parts twice, first with respect to z_2 and then with respect to z_1, we get

RHS of (9.7.6)

$$=\frac{1}{2\pi^2}\int\int\frac{f(z_1)g'(z_2)}{(\underline{s}(z_1)-\underline{s}(z_2))}\frac{d}{dz_1}\underline{s}(z_1)dz_2dz_1$$

$$=-\frac{1}{2\pi^2}\int\int f'(z_1)g'(z_2)\log(\underline{s}(z_1)-\underline{s}(z_2))dz_1dz_2$$

(where log is any branch of the logarithm)

$$=-\frac{1}{2\pi^2}\int\int f'(z_1)g'(z_2)[\log|\underline{s}(z_1)-\underline{s}(z_2)|+i\arg(\underline{s}(z_1)-\underline{s}(z_2))]dz_1dz_2.$$

We choose the contours to be rectangles with sides parallel to the axes. The inside rectangle intersects the real axis at a and b, and the horizontal sides are a distance $v < 1$ away from the real axis. The outside rectangle intersects the real axis at $a-\varepsilon$, $b+\varepsilon$ (points where f and g remain analytic), with height twice that of the inside rectangle. We let $v \to 0$.

We need only consider the logarithm term and show its convergence since the real part of the arg term disappears (f and g are real-valued on $\mathbb{R}$) in the limit, and the sum (9.7.6) is real. Therefore the arg term also approaches zero.

We split up the log integral into 16 double integrals, each one involving a side from each of the two rectangles. We argue that any portion of the integral involving a vertical side can be neglected. This follows from (9.12.1), (9.12.2), and the fact that z_1 and z_2 remain a positive distance apart, so that $|\underline{s}(z_1)-\underline{s}(z_2)|$ is bounded away from zero. Moreover, at least one of $|\underline{s}(z_1)|$, $|\underline{s}(z_2)|$ is bounded, while the other is bounded by $1/v$, so the integral is bounded by $Kv\log v^{-1} \to 0$.

Therefore we arrive at

$$-\frac{1}{2\pi^2}\int_a^b\int_{a-\varepsilon}^{b+\varepsilon}[(f'(x+i2v)g'(y+iv)+\bar{f}'(x+i2v)\bar{g}'(y+iv))$$
$$\times\log|\underline{s}(x+i2v)-\underline{s}(y+iv)|-(f'(x+i2v)\bar{g}'(y+iv)$$
$$+\bar{f}'(x+i2v)g'(y+iv))\log|\underline{s}(x+i2v)-\overline{\underline{s}}(y+iv)|]dxdy. \tag{9.12.3}$$

Using subscripts to denote real and imaginary parts, we find

$$\begin{aligned}(9.12.3)=&-\frac{1}{\pi^2}\int_a^b\int_{a-\varepsilon}^{b+\varepsilon}[(f_r'(x+i2v)g_r'(y+iv)\\ &-f_i'(x+i2v)g_i'(y+iv))\log|\underline{s}(x+i2v)-\underline{s}(y+iv)|\\ &-(f_r'(x+i2v)g_r'(y+iv)+f_i'(x+i2v)g_i'(y+iv))\\ &\times\log|\underline{s}(x+i2v)-\overline{\underline{s}}(y+iv)|]dxdy\\ =&\frac{1}{\pi^2}\int_a^b\int_{a-\varepsilon}^{b+\varepsilon}f_r'(x+i2v)g_r'(y+iv)\\ &\times\log\left|\frac{\underline{s}(x+i2v)-\overline{\underline{s}}(y+iv)}{\underline{s}(x+i2v)-\underline{s}(y+iv)}\right|dxdy \qquad (9.12.4)\\ &+\frac{1}{\pi^2}\int_a^b\int_{a-\varepsilon}^{b+\varepsilon}f_i'(x+i2v)g_i'(y+iv)\log|(\underline{s}(x+i2v)\\ &-\underline{s}(y+iv))(\underline{s}(x+i2v)-\overline{\underline{s}}(y+iv))|dxdy. \qquad (9.12.5)\end{aligned}$$

We have for any real-valued h analytic on the bounded interval $[\alpha,\beta]$ for all v sufficiently small,

$$\sup_{x\in[\alpha,\beta]}|h_i(x+iv)|\le K|v|, \tag{9.12.6}$$

where K is a bound on $|h'(z)|$ for z in a neighborhood of $[\alpha,\beta]$. Using this and (9.12.1) and (9.12.2), we see that (9.12.5) is bounded in absolute value by $Kv^2\log v^{-1}\to 0$.

For (9.12.4), we write

$$\begin{aligned}&\log\left|\frac{\underline{s}(x+i2v)-\overline{\underline{s}}(y+iv)}{\underline{s}(x+i2v)-\underline{s}(y+iv)}\right|\\ &=\frac{1}{2}\log\left(1+\frac{4\underline{s}_i(x+i2v)\underline{s}_i(y+iv)}{|\underline{s}(x+i2v)-\underline{s}(y+iv)|^2}\right).\end{aligned} \tag{9.12.7}$$

From (9.12.2), we get

$$\text{RHS of (9.12.7)}\le\frac{1}{2}\log\left(1+\frac{16\underline{s}_i(x+i2v)\underline{s}_i(y+iv)}{(x-y)^2|\underline{s}(x+i2v)\underline{s}(y+iv)|^2}\right).$$

From (9.12.1), we have

$$\sup_{\substack{x,y\in[a-\varepsilon,b+\varepsilon]\\ v\in(0,1)}} \frac{\underline{s}_i(x+i2v)\underline{s}_i(y+iv)}{|\underline{s}(x+i2v)\underline{s}(y+iv)|^2} < \infty.$$

Therefore, there exists a $K>0$ for which the right-hand side of (9.12.7) is bounded by

$$\frac{1}{2}\log\left(1+\frac{K}{(x-y)^2}\right) \tag{9.12.8}$$

for $x, y \in [a-\varepsilon, b+\varepsilon]$. It is straightforward to show that (9.12.8) is Lebesgue integrable on bounded subsets of $\mathbb{R}^2$. Therefore, from (9.8.10) and the dominated convergence theorem, we conclude that (9.8.11) is Lebesgue integrable and that (9.8.8) holds.

9.12.2 Verification of (9.8.9)

From (9.7.1), we have

$$\frac{d}{dz}\underline{s}(z) = \frac{\underline{s}^2(z)}{1-y\int\frac{t^2\underline{s}^2(z)}{(1+t\underline{s}(z))^2}dH(t)}.$$

In Silverstein and Choi [267], it is argued that the only places where $\underline{s}'(z)$ can possibly become unbounded are near the origin and the boundary, ∂S_F, of S_F. It is a simple matter to verify

$$\begin{aligned}\mathrm{E}X_f &= \frac{1}{4\pi i}\int f(z)\frac{d}{dz}\mathrm{Log}\left(1-y\int\frac{t^2\underline{s}^2(z)}{(1+t\underline{s}(z))^2}dH(t)\right)dz\\ &= -\frac{1}{4\pi i}\int f'(z)\mathrm{Log}\left(1-y\int\frac{t^2\underline{s}^2(z)}{(1+t\underline{s}(z))^2}dH(t)\right)dz,\end{aligned}$$

where, because of (9.9.22), the arg term for log can be taken from $(-\pi/2,\pi/2)$. We choose a contour as above. From (6.2.22), there exists a $K>0$ such that, for all small v,

$$\inf_{x\in\mathbb{R}}\left|1-y\int\frac{t^2\underline{s}^2(x+iv)}{(1+t\underline{s}(x+iv))^2}dH(t)\right| \ge Kv^2. \tag{9.12.9}$$

Therefore, we see that the integrals on the two vertical sides are bounded by $Kv\log v^{-1}\to 0$. The integral on the two horizontal sides is equal to

$$\frac{1}{2\pi}\int_a^b f_i'(x+iv)\log\left|1-y\int\frac{t^2\underline{s}^2(x+iv)}{(1+t\underline{s}(x+iv))^2}dH(t)\right|dx$$

$$+\frac{1}{2\pi}\int_a^b f_r'(x+iv)\arg\left(1-y\int\frac{t^2\underline{s}^2(x+iv)}{(1+t\underline{s}(x+iv))^2}dH(t)\right)dx. \tag{9.12.10}$$

Using (9.9.22), (9.12.6), and (9.12.9), we see that the first term in (9.12.10) is bounded in absolute value by $Kv\log v^{-1} \to 0$. Since the integrand in the second term converges for all $x \notin \{0\} \cup \partial S_F$ (a countable set), we therefore get (9.8.9) from the dominated convergence theorem.

9.12.3 Derivation of Quantities in Example (1.1)

We now derive $d(y)$ ($y \in (0,1)$) in (1.1.1), (9.8.12), and the variance in (9.8.13). The first two rely on Poisson's integral formula

$$u(z) = \frac{1}{2\pi}\int_0^{2\pi} u(e^{i\theta})\frac{1-r^2}{1+r^2-2r\cos(\theta-\phi)}d\theta, \tag{9.12.11}$$

where u is harmonic on the unit disk in $\mathbb{C}$, and $z = re^{i\phi}$ with $r \in [0,1)$. Making the substitution $x = 1+y-2\sqrt{y}\cos\theta$, we get

$$\begin{aligned} d(y) &= \frac{1}{\pi}\int_0^{2\pi}\frac{\sin^2\theta}{1+y-2\sqrt{y}\cos\theta}\log(1+y-2\sqrt{y}\cos\theta)d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi}\frac{2\sin^2\theta}{1+y-2\sqrt{y}\cos\theta}\log|1-\sqrt{y}e^{i\theta}|^2 d\theta. \end{aligned}$$

It is straightforward to verify that

$$f(z) \equiv -(z-z^{-1})^2(\log(1-\sqrt{y}z)+\sqrt{y}z)-\sqrt{y}(z-z^3)$$

is analytic on the unit disk and that

$$\Re f(e^{i\theta}) = 2\sin^2\theta\log|1-\sqrt{y}e^{i\theta}|^2.$$

Therefore, from (9.12.11), we have

$$d(y) = \frac{f(\sqrt{y})}{1-y} = \frac{y-1}{y}\log(1-y)-1.$$

For (9.8.12), we use (9.8.9). From (9.7.1), with $H(t) = I_{[1,\infty)}(t)$, we have for $z \in \mathbb{C}^+$

$$z = -\frac{1}{\underline{s}(z)} + \frac{y}{1+\underline{s}(z)}. \tag{9.12.12}$$

Solving for $\underline{s}(z)$, we find

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

the square roots defined to yield positive imaginary parts for $z \in \mathbb{C}^+$. As $z \to x \in [a(y), b(y)]$ (limits defined below (1.1.1)), we get

$$\underline{s}(x) = \frac{-(x+1-y)+\sqrt{4y-(x-1-y)^2}\,i}{2x}$$
$$= \frac{-(x+1-y)+\sqrt{(x-a(y))(b(y)-x)}\,i}{2x}.$$

Identity (9.12.12) still holds with z replaced by x, and from it we get

$$\frac{\underline{s}(x)}{1+\underline{s}(x)} = \frac{1+x\underline{s}(x)}{y},$$

so that

$$1 - y\frac{\underline{s}^2(x)}{(1+\underline{s}(x))^2}$$
$$= 1 - \frac{1}{y}\left(\frac{-(x-1-y)+\sqrt{4y-(x-1-y)^2}\,i}{2}\right)^2$$
$$= \frac{\sqrt{4y-(x-1-y)^2}}{2y}\left(\sqrt{4y-(x-1-y)^2}+(x-1-y)i\right).$$

Therefore, from (9.8.9),

$$\mathrm{E}X_f = \frac{1}{2\pi}\int_{a(y)}^{b(y)} f'(x)\tan^{-1}\left(\frac{x-1-y}{\sqrt{4y-(x-1-y)^2}}\right)dx$$
$$= \frac{f(a(y))+f(b(y))}{4} - \frac{1}{2\pi}\int_{a(y)}^{b(y)}\frac{f(x)}{\sqrt{4y-(x-1-y)^2}}dx. \tag{9.12.13}$$

To compute the last integral when $f(x) = \log x$, we make the same substitution as before, arriving at

$$\frac{1}{4\pi}\int_0^{2\pi}\log|1-\sqrt{y}e^{i\theta}|^2 d\theta.$$

We apply (9.12.11), where now $u(z) = \log|1-\sqrt{y}z|^2$, which is harmonic, and $r = 0$. Therefore, the integral must be zero, and we conclude that

$$\mathrm{E}X_{\log} = \frac{\log(a(y)b(y))}{4} = \frac{1}{y}\log(1-y).$$

To derive (9.8.13), we use (9.8.7). Since the z_1, z_2 contours cannot enclose the origin (because of the logarithm), neither can the resulting s_1, s_2 contours. Indeed, either from the graph of $x(\underline{s})$ or from $\underline{s}(x)$, we see that $x > b(y) \Longleftrightarrow \underline{s}(x) \in (-(1+\sqrt{y})^{-1}, 0)$ and $x \in (0, a(y)) \Longleftrightarrow \underline{s}(x) < (\sqrt{y}-1)^{-1}$. For our analysis, it is sufficient to know that the s_1, s_2 contours, nonintersecting and both taken in the positive direction, enclose $(y-1)^{-1}$ and -1, but not 0. Assume the s_2 contour encloses the s_1 contour. For fixed s_2, using (9.12.12) we have

$$\int \frac{\log(z(s_1))}{(s_1-s_2)^2}\,ds_1 = \int \frac{\frac{1}{s_1^2} - \frac{y}{(1+s_1)^2}}{-\frac{1}{s_1} + \frac{y}{1+s_1}} \frac{1}{(s_1-s_2)}\,ds_1$$
$$= \int \frac{(1+s_1)^2 - ys_1^2}{ys_1(s_1-s_2)} \left(\frac{-1}{s_1+1} + \frac{1}{s_1 - \frac{1}{y-1}} \right) ds_1$$
$$= 2\pi i \left(\frac{1}{s_2+1} - \frac{1}{s_2 - \frac{1}{y-1}} \right).$$

Therefore,

$$\mathrm{Var}X_{\log} = \frac{1}{\pi i} \int \left(\frac{1}{s+1} - \frac{1}{s-\frac{1}{y-1}} \right) \log(z(s))ds$$
$$= \frac{1}{\pi i} \int \left[\frac{1}{s+1} - \frac{1}{s-\frac{1}{y-1}} \right] \log\left(\frac{s-\frac{1}{y-1}}{s+1} \right) ds$$
$$-\frac{1}{\pi i} \int \left[\frac{1}{s+1} - \frac{1}{s-\frac{1}{y-1}} \right] \log(s) ds.$$

The first integral is zero since the integrand has antiderivative

$$-\frac{1}{2}\left[\log\left(\frac{s-\frac{1}{y-1}}{s+1} \right) \right]^2,$$

which is single-valued along the contour. Therefore, we conclude that

$$\mathrm{Var}X_{\log} = -2[\log(-1) - \log((y-1)^{-1})] = -2\log(1-y).$$

9.12.4 Verification of Quantities in Jonsson's Results

Finally, we compute expressions for (9.8.14) and (9.8.15). Using (9.12.13), we have

$$\mathrm{E}X_{x^r} = \frac{(a(y))^r + (b(y))^r}{4} - \frac{1}{2\pi}\int_{a(y)}^{b(y)} \frac{x^r}{\sqrt{4y-(x-1-y)^2}}dx$$
$$= \frac{(a(y))^r + (b(y))^r}{4} - \frac{1}{4\pi}\int_0^{2\pi} |1-\sqrt{y}e^{i\theta}|^{2r}d\theta$$
$$= \frac{(a(y))^r + (b(y))^r}{4} - \frac{1}{4\pi}\int_0^{2\pi} \sum_{j,k=0}^{r}\binom{r}{j}\binom{r}{k}(-\sqrt{y})^{j+k}e^{i(j-k)\theta}d\theta$$
$$= \frac{1}{4}((1-\sqrt{y})^{2r}+(1+\sqrt{y})^{2r}) - \frac{1}{2}\sum_{j=0}^{r}\binom{r}{j}^2 y^j,$$

which is (9.8.14).

For (9.8.15), we use (9.8.7) and rely on observations made in deriving (9.8.13). For $y \in (0,1)$, the contours can again be made enclosing -1 and not the origin. However, because of the fact that (9.7.6) derives from (9.8.5) and the support of $F^{y,I_{[1,\infty)}}$ on $\mathbb{R}^+$ is $[a(y), b(y)]$, we may also take the contours in the same way when $y > 1$. The case $y = 1$ simply follows from the continuous dependence of (9.8.7) on y.

Keeping s_2 fixed, we have on a contour within 1 of -1,

$$\int \frac{(-\frac{1}{s_1}+\frac{y}{1+s_1})^{r_1}}{(s_1-m_2)^2}ds_1$$
$$= y^{r_1}\int\left(\frac{1}{s_1+1}+\frac{1-y}{y}\right)^{r_1}(1-(s_1+1))^{-r_1}(s_2+1)^{-2}$$
$$\times\left(1-\frac{s_1+1}{s_2+1}\right)^{-2}ds_1$$
$$= y^{r_1}\int\sum_{k_1=0}^{r_1}\binom{r_1}{k_1}\left(\frac{1-y}{y}\right)^{k_1}(1+s_1)^{k-r_1}$$
$$\times\sum_{j=0}^{\infty}\binom{r_1+j-1}{j}(s_1+1)^j(s_2+1)^{-2}\sum_{\ell=1}^{\infty}\ell\left(\frac{s_1+1}{s_2+1}\right)^{\ell-1}ds_1$$
$$= 2\pi i y^{r_1}\sum_{k_1=0}^{r_1-1}\sum_{\ell=1}^{r_1-k_1}\binom{r_1}{k_1}\left(\frac{1-y}{y}\right)^{k_1}\binom{2r_1-1-(k_1+\ell)}{r_1-1}\ell(s_2+1)^{-\ell-1}.$$

Therefore,

$$\mathrm{Cov}(X_{x^{r_1}}, X_{x^{r_2}}) = -\frac{i}{\pi}y^{r_1+r_2}\sum_{k_1=0}^{r_1-1}\sum_{\ell=1}^{r_1-k_1}\binom{r_1}{k_1}\left(\frac{1-y}{y}\right)^{k_1}$$
$$\times\binom{2r_1-1-(k_1+\ell)}{r_1-1}\ell\int(s_2+1)^{-\ell-1}\sum_{k_2=0}^{r_2}\binom{r_2}{k_2}\left(\frac{1-y}{y}\right)^{k_2}$$

$$\times (s_2+1)^{k_2-r_2} \sum_{j=0}^{\infty} \binom{r_2+j-1}{j} (s_2+1)^j ds_2$$

$$= 2y^{r_1+r_2} \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2} \binom{r_1}{k_1}\binom{r_2}{k_2}$$

$$\times \left(\frac{1-y}{y}\right)^{k_1+k_2} \sum_{\ell=1}^{r_1-k_1} \ell \binom{2r_1-1-(k_1+\ell)}{r_1-1}\binom{2r_2-1-k_2+\ell}{r_2-1},$$

which is (9.8.15), and we are done.

9.12.5 Verification of (9.7.8) and (9.7.9)

We verify these two bounds by modifying the proof in Theorem 5.10. We present the following theorem.

Theorem 9.13. *For each fixed n, suppose that $x_{ij} = x_{ij}^{(n)}$, $i = 1, \cdots, p$, $j = 1, \cdots, n$ are independent complex random variables satisfying* $\mathrm{E}x_{ij} = 0$, $\max_i \sum_{j=1}^n |1 - \mathrm{E}|x_{ij}|^2| = o(n)$, $\max_{i,j,n} \mathrm{E}|x_{ij}|^4 < \infty$, *and* $|x_{ij}| \le \eta_n \sqrt{n}$, *where η_n are positive constants with $\eta_n \to 0$ and $\eta_n n^{1/4} \to \infty$. Let $\mathbf{S}_n = (1/n)\mathbf{X}\mathbf{X}^*$, where $\mathbf{X} = (X_{ij})$ is $p \times n$ with $p/n \to y > 0$ as $n \to \infty$. Then, for any $\mu > (1+\sqrt{y})^2$ and any $\ell > 0$,*

$$\mathrm{P}(\lambda_{\max}(\mathbf{S}_n) > \mu) = o(n^{-\ell}).$$

Moreover, if $y \in (0,1)$, then for any $\mu < (1-\sqrt{y})^2$ and any $\ell > 0$,

$$\mathrm{P}(\lambda_{\min}(\mathbf{S}_n) < \mu) = o(n^{-\ell}).$$

Proof. We assume first that $y \in (0,1)$. We follow along the proof of Theorem 5.10. The conclusions of Lemmas 5.12 and 5.13–5.15 need to be improved from "almost sure" statements to ones reflecting tail probabilities. We shall denote the augmented lemmas with primes ($'$) after the numbers.

For Lemma 5.12, it has been shown that for the Hermitian matrices $\mathbf{T}(l)$ defined there, and even integers m_n satisfying $m_n/\log n \to \infty$, $m_n \eta_n^{1/3}/\log n \to 0$, and $m_n/(\eta_n\sqrt{n}) \to 0$,

$$\mathrm{E}\mathrm{tr}\mathbf{T}^{2m_n}(l) \le n^2((2l+1)(l+1))^{2m_n}(p/n)^{m_n(l-1)}(1+o(1))^{4m_n l}$$

(see (5.2.18)). Therefore, writing $m_n = k_n \log n$, for any $\varepsilon > 0$ there exists an $a \in (0,1)$ such that, for all large n,

$$\begin{aligned} &\mathrm{P}(\mathrm{tr}\mathbf{T}(l) > (2l+1)(l+1)y^{(l-1)/2} + \varepsilon) \\ &\le n^2 a^{m_n} = n^{2+k_n \log a} = o(n^{-\ell}) \end{aligned} \qquad (9.12.14)$$

for any positive ℓ. We call (9.12.14) Lemma 5.12′.

We next replace Lemma 5.13 with the following one.

Lemma 5.13′. *Under the conditions of Theorem 9.13, for any* $\varepsilon > 0$, $f \geq 2$, *and* $\ell > 0$,

$$\mathrm{P}\left(n^{-f/2} \max_{i \leq p} \left| \sum_{j=1}^{n} (|x_{ij}|^f - \mathrm{E}|x_{ij}|^f) \right| > \varepsilon \right) = o(n^{-\ell}).$$

Proof. Similar to the estimation for moments of S_1 given in the proof of Lemma 9.1, choosing an even integer $k_n \sim \eta_n^{-1/2} \nu^{-1} \log n$, we have

$$\begin{aligned} &\mathrm{E}\left| n^{-f/2} \sum_{j=1}^{n} (|x_{ij}|^f - \mathrm{E}|x_{ij}|^f) \right|^k \\ &\leq \sum_{1 \leq s \leq k/2} n^{-s} \eta_n^{-4s} \nu^s s^k \\ &\leq \nu(n\eta_n^4)^{-1} (40\eta_n)^{kf} = o(n^\ell), \end{aligned}$$

for $f \geq 2$ and any fixed ℓ, where ν is the super bound of the fourth moments of the underlying variables. This completes the proof of the lemma.

Redefining the matrix $\mathbf{Y}_n^{(f)}$ in Lemma 5.13 to be $[|X_{uv}|^f]$, Lemma 5.13′ states that, for any ε and ℓ,

$$\begin{aligned} &\mathrm{P}(\lambda_{\max}\{n^{-1}\mathbf{Y}_n^{(1)}\mathbf{Y}_n^{(1)^*}\} > 7 + \varepsilon) = o(n^{-\ell}), \\ &\mathrm{P}(\lambda_{\max}\{n^{-2}\mathbf{Y}_n^{(2)}\mathbf{Y}_n^{(2)^*}\} > y + \varepsilon) = o(n^{-\ell}), \\ &\mathrm{P}(\lambda_{\max}\{n^{-f}\mathbf{Y}_n^{(f)}\mathbf{Y}_n^{(f)^*}\} > \varepsilon) = o(n^{-\ell}) \quad \text{for any } f > 2. \end{aligned}$$

For the first estimation, we have

$$\lambda_{\max}\{n^{-1}\mathbf{Y}_n^{(1)}\mathbf{Y}_n^{(1)^*}\} \leq \mathbf{T}_n(1) + \frac{1}{n} \max_i \sum_{j=1}^{n} |x_{ij}|^2.$$

Thus,

$$\begin{aligned} &\mathrm{P}(\lambda_{\max}\{n^{-1}\mathbf{Y}_n^{(1)}\mathbf{Y}_n^{(1)^*}\} > 7 + \varepsilon) \\ &\leq \mathrm{P}(\|\mathbf{T}_n(1)\| \geq 6 + \varepsilon/2) + \mathrm{P}\left(\frac{1}{n} \max_i \sum_{j=1}^{n} |x_{ij}|^2 \geq 1 + \varepsilon/2\right) \\ &\leq o(n^{-\ell}), \end{aligned}$$

where we have used Lemma 5.12′ for $\mathrm{P}(\|\mathbf{T}_n(1)\| \geq 6 + \varepsilon/2) = o(n^{-\ell})$. The second probability can be estimated by Lemma 5.13′.

For the second and the third estimations, we use the Gerŝgorin bound[2]

$$\begin{aligned}
&\lambda_{\max}\{n^{-f}\mathbf{Y}_n^{(f)}\mathbf{Y}_n^{(f)^*}\}\\
&\le \max_i n^{-f}\sum_{j=1}^{n}|x_{ij}|^{2f}\\
&\quad+\max_i n^{-f}\sum_{k\neq i}\sum_{j=1}^{n}|x_{ij}|^f|x_{kj}|^f\\
&\le \max_i n^{-f}\sum_{j=1}^{n}|x_{ij}|^{2f}+\left(\max_i n^{-f/2}\sum_{j=1}^{n}|x_{ij}|^f\right)\\
&\quad\times\left(\max_j n^{-f/2}\sum_{k=1}^{p}|x_{kj}|^f\right). \qquad (9.12.15)
\end{aligned}$$

When $f > 1$, then

$$n^{-f}\sum_{j=1}^{n}\mathrm{E}|x_{ij}^{2f}| \le \eta_n^{2f-2}\to 0.$$

Thus, in application of Lemma 5.13′, we may use

$$\mathrm{P}\left(\max_i n^{-f/2}\sum_{j=1}^{n}|x_{ij}|^f \ge \varepsilon = o(n^{-\ell})\right), \text{ for } f>1,$$

$$\begin{aligned}
&\mathrm{P}\left(\lambda_{\max}\{n^{-2}\mathbf{Y}_n^{(2)}\mathbf{Y}_n^{(2)^*}\}>y+\varepsilon\right)\\
&\le \mathrm{P}\left(\max_i n^{-2}\sum_{j=1}^{n}|x_{ij}|^4\ge\frac{\varepsilon}{2}\right)+\mathrm{P}\left(\max_i \frac{1}{n}\sum_{j=1}^{n}|x_{ij}^2|\ge 1+\frac{\varepsilon}{2+y}\right)\\
&\quad+\mathrm{P}\left(\max_j\frac{1}{n}\sum_{i=1}^{p}|x_{ij}|^2\ge y+\frac{\varepsilon}{2+y}\right)\\
&= o(n^{-\ell}).
\end{aligned}$$

For $f > 2$, we have

[2] Gerŝgorin's theorem states that any eigenvalue of the matrix $\mathbf{A}=(a_{ij})$ must be enclosed in one of the circles with center a_{kk} and radius $\sum_{j\neq k}|a_{jk}|$. Its proof is simple. Let λ be an eigenvalue of $\mathbf{A}$ with eigenvector $\mathbf{x}=(x_1,\cdots,x_k)'$. Suppose $|x_k|=\max_j(|x_j|)$. Then the conclusion follows from the equality

$$(a_{kk}-\lambda)x_k=-\sum_{j\neq k}a_{kj}x_j.$$

$$
\begin{aligned}
&\mathrm{P}(\lambda_{\max}\{n^{-f}\mathbf{Y}_n^{(f)}\mathbf{Y}_n^{(f)^*}\} > \varepsilon) \\
&\le \mathrm{P}\left(n^{-f}\max_i \sum_{j=1}^n |x_{1j}|^{2f} > \varepsilon/2\right) + \mathrm{P}\left(n^{-f/2}\max_i \sum_{j=1}^n |x_{ij}|^{f} > \sqrt{\varepsilon/2}\right) \\
&\quad + \mathrm{P}\left(n^{-f/2}\max_j \sum_{k=1}^p |X_{kj}|^{f} > \sqrt{\varepsilon/2}\right) \\
&= o(n^{-\ell}).
\end{aligned}
$$

The proofs of Lemmas 5.14′ and 5.15′ are handled using the arguments in the proof of Theorem 5.10 and those used above: each quantity L_n in the proof of Theorem 5.10 that is $o(1)$ a.s. can be shown to satisfy $\mathrm{P}(|L_n| > \varepsilon) = o(n^{-\ell})$.

From Lemmas 5.12′ and 5.15′, there exists a positive C such that, for every integer $k > 0$ and positive ε and ℓ,

$$
\mathrm{P}(\|\mathbf{T} - y\mathbf{I}\|^k > Ck^4 2^k y^{k/2} + \varepsilon) = o(n^{-\ell}). \tag{9.12.16}
$$

For given $\varepsilon > 0$, let integer $k > 0$ be such that

$$
|2\sqrt{y}(1 - (Ck^4)^{1/k})| < \varepsilon/2.
$$

Then

$$
2\sqrt{y} + \varepsilon > 2\sqrt{y}(Ck^4)^{1/k} + \varepsilon/2 \ge (Ck^4 2^k y^{k/2} + (\varepsilon/2)^k)^{1/k}.
$$

Therefore, from (9.12.16), we get, for any $\ell > 0$,

$$
\mathrm{P}(\|\mathbf{T} - y\mathbf{I}\| > 2\sqrt{y} + \varepsilon) = o(n^{-\ell}). \tag{9.12.17}
$$

From Lemma 9.1 with $\mathbf{A} = \mathbf{I}$ and $p = [\log n]$, and (9.12.17), we get for any fixed positive ε and ℓ,

$$
\begin{aligned}
&\mathrm{P}(\|\mathbf{S}_n - (1+y)\mathbf{I}\| > 2\sqrt{y} + \varepsilon) \\
&\le \mathrm{P}(\|\mathbf{S}_n - \mathbf{I} - \mathbf{T}\| > \varepsilon/2) + o(n^{-\ell}) \\
&= \mathrm{P}\left(\max_{i \le p}\left| n^{-1}\sum_{j=1}^n |X_{ij}|^2 - 1\right| > \varepsilon/2\right) + o(n^{-\ell}) = o(n^{-\ell}).
\end{aligned}
$$

Therefore, for any positive $\mu > (1+\sqrt{y})^2$ and $\ell > 0$,

$$
\begin{aligned}
&\mathrm{P}(\lambda_{\max}(\mathbf{S}_n) > \mu) \\
&= \mathrm{P}(\lambda_{\max}(\mathbf{S}_n - (1+y)\mathbf{I}) > \mu - (1+\sqrt{y})^2 + 2\sqrt{y}) \\
&\le \mathrm{P}(\|\mathbf{S}_n - (1+y)\mathbf{I}\| > 2\sqrt{y} + \mu - (1+\sqrt{y})^2) = o(n^{-\ell}).
\end{aligned}
$$

Similarly, if $\mu < (1-\sqrt{y})^2$ and $\ell > 0$,

$$
\mathrm{P}(\lambda_{\min}(\mathbf{S}_n) < \mu)
$$

$$\leq \mathrm{P}(\|\mathbf{S}_n - (1+y)\mathbf{I}\| > 2\sqrt{y} + (1-\sqrt{y})^2 - \mu) = o(n^{-\ell}).$$

For $y > 1$ and $\mu > (1+\sqrt{y})^2$, choose $\underline{\mu}$ such that $(1+1/\sqrt{y})^2 < \underline{\mu} < (n/p)\mu$ for all n sufficiently large. Then, for these n and any $\ell > 0$,

$$\begin{aligned}&\mathrm{P}(\lambda_{\max}(\mathbf{S}_n) > \mu) = \mathrm{P}(\lambda_{\max}(1/p)\mathbf{X}^*\mathbf{X} > (n/p)\mu)\\ &\leq \mathrm{P}(\lambda_{\max}(1/p)\mathbf{X}^*\mathbf{X} > \underline{\mu}) = o(n^{-\ell}).\end{aligned}$$

Finally, for $y = 1$ and any $\mu > 4$, let $\underline{y} < 1$ be such that $\underline{y}\mu > (1+\sqrt{\underline{y}})^2$. Let m_n be a sequence of positive integers for which $p/(n+m_n) \to \underline{y}$. Notice that $n/(n+m_n)$ also converges to $\underline{y}$. Let $\underline{\mathbf{X}}$ be $p \times m_n$ with entries independent of $\mathbf{X}$ and distributed the same as those of $\mathbf{X}$. Choose $\underline{\mu}$ satisfying $(1+\sqrt{\underline{y}})^2 < \underline{\mu} < (n/(n+m_n))\mu$ for all large n. For these n and any $\ell > 0$, we have

$$\begin{aligned}&\mathrm{P}(\lambda_{\max}(\mathbf{S}_n) > \mu) \leq \mathrm{P}(\lambda_{\max}(\mathbf{S}_n + (1/n)\underline{\mathbf{X}}\underline{\mathbf{X}}^*) > \mu)\\ &\leq \mathrm{P}(\lambda_{\max}(1/(n+m_n))(\mathbf{X}\mathbf{X}^* + \underline{\mathbf{X}}\underline{\mathbf{X}}^*) > \underline{\mu}) = o(n^{-\ell})\end{aligned}$$

and we are done.

9.13 CLT for the *F*-Matrix

The multivariate F-matrix, its LSD, and the Stieltjes transform of its LSD are defined and derived in Section 4.4. To facilitate the reading, we repeat them here. If $p/n_1 \to y_1 \in (0,\infty)$ and $p/n_2 \to y_2 \in (0,1)$, then the LSD has density

$$\begin{aligned}p(x) &= \begin{cases} \frac{\sqrt{4x-((1-y_1)+x(1-y_2))^2}}{2\pi x(y_1+y_2x)}, & \text{when } 4x-((1-y_1)+x(1-y_2))^2 > 0,\\ 0, & \text{otherwise.}\end{cases}\\ &= \begin{cases} \frac{(1-y_2)\sqrt{(b-x)(x-a)}}{2\pi x(y_1+y_2x)}, & \text{when } a < x < b,\\ 0, & \text{otherwise,}\end{cases}\end{aligned}$$

where $a, b = \left(\frac{1\mp\sqrt{y_1+y_2-y_1y_2}}{1-y_2}\right)^2$. The LSD will have a point mass $1-1/y_1$ at the origin when $y_1 > 1$.

Its Stieltjes transform (see Section 4.4) is

$$s(z) = \frac{(1-y_1) - z(y_1+y_2) + \sqrt{\Big((1-y_1)+z(1-y_2)\Big)^2 - 4z}}{2z(y_1+zy_2)},$$

from which we have

$$\underline{s}(z) = \frac{y_1(1-y_1) - z(2y_2 - y_1y_2 + y_1^2) + y_1\sqrt{\Big((1-y_1)+z(1-y_2)\Big)^2 - 4z}}{2z(y_1+zy_2)},$$

where $\underline{s}(z) = -\frac{1-y_1}{z} + y_1 s(z)$.

The CLT of the F-matrix has many important applications in multivariate statistical analysis. For example, in multivariate linear regression models $X = \beta Z + \epsilon$, where $\beta = (\beta_1, \beta_2)$ is a parameter matrix, the log likelihood ratio statistic T, up to a constant multiplier, for the testing problem

$$H_0 : \beta_1 = \beta_1^* \ vs. \ H_1 : \beta_1 \neq \beta_1^*$$

can be expressed as a functional of the empirical spectral distribution of the F-matrix,

$$T = \int f(x) dF^{\{n_1,n_2\}}(x) = -\frac{1}{p}\sum_i \log(1+\lambda_i),$$

where the λ_i's are the eigenvalues of an F-matrix, $F^{\{n_1,n_2\}}(x)$ its ESD, and $f(x) = -\log(1+x)$. Similarly, it is known that the log likelihood ratio test of equality of covariance matrices of two populations $H_0 : \Sigma_1 = \Sigma_2$ is equivalent to a functional of the empirical spectral distribution of the F-matrix with $f(x) = \frac{y_1+y_2}{y_2}\log(y_2x+y_1) - \log(x)$. It is known that the Wilks approximation for log likelihood ratio statistics does not work well as the dimension p proportionally increases with the sample size and thus we have to find an alternative limiting theorem to form the hypothesis test. We see then the importance in investigating the CLT of the LSS associated with multivariate F-matrices.

Throughout this section, we assume that

$$y_{n_1} = p/n_1 \to y_1 \in (0,\infty) \quad \text{and} \quad y_{n_2} = p/n_2 \to y_2 \in (0,1)$$

as $\min(n_1, n_2) \to \infty$.

Let $s^{\{n_1,n_2\}}(z)$ denote the Stieltjes transform of the ESD $F^{\{n_1,n_2\}}(x)$ of the F-matrix $\mathbf{S}_1\mathbf{S}_2^{-1}$ and $s^{\{y_1,y_2\}}(z)$ denote the Stieltjes transform of the LSD $F^{\{y_1,y_2\}}(x)$. Let $\underline{s}^{\{n_1,n_2\}}(z) = -\frac{1-y_{n_1}}{z} + y_{n_1}s^{\{n_1,n_2\}}(z)$ and $\underline{s}^{\{y_1,y_2\}}(z) = -\frac{1-y_1}{z} + y_1 s^{\{y_1,y_2\}}(z)$. For brevity, $s^{\{y_1,y_2\}}(z)$ and $\underline{s}^{\{y_1,y_2\}}(z)$ will simply be written as $s(z)$ and $\underline{s}(z)$.

Let $s_{n_2}(z)$ denote the Stieltjes transform of the ESD $F_{n_2}(x)$ of $\mathbf{S}_2$ and $s_{y_2}(z)$ denote the Stieltjes transform of the LSD $F_{y_2}(x)$ of $\mathbf{S}_2$. Let $\underline{s}_{n_2}(z) = -\frac{1-y_{n_2}}{z} + y_{n_2}s_{n_2}(z)$ and $\underline{s}_{y_2}(z) = -\frac{1-y_2}{z} + y_2 s_{y_2}(z)$.

Let $H_{n_2}(x)$ and $H_{y_2}(x)$ denote the ESD and LSD of $\mathbf{S}_2^{-1}$. Note that $\frac{1}{\lambda}$ is a positive eigenvalue of $\mathbf{S}_2$ if λ is a positive eigenvalue of $\mathbf{S}_2^{-1}$. Hence, we have $H_{n_2}(x-0) = 1 - F_{n_2}(1/x)$ and $H_{y_2}(x) = 1 - F_{y_2}(1/x)$ for all $x > 0$.

9.13.1 CLT for LSS of the F-Matrix

Let $F^{\{n_1,n_2\}}(x)$ and $F^{\{y_{n_1},y_{n_2}\}}(x)$ denote the ESD and LSD of the F-matrix $\mathbf{S}_1\mathbf{S}_2^{-1}$. The LSS of the F-matrix for functions $f_1,\ldots,f_k$ is

$$\left(\int f_1(x)d\widetilde{G}_{n_1,n_2}(x),\cdots,\int f_k(x)d\widetilde{G}_{n_1,n_2}(x)\right),$$

where

$$\widetilde{G}_{\{n_1,n_2\}}(x)=p\left(F^{\{n_1,n_2\}}(x)-F^{\{y_{n_1},y_{n_2}\}}(x)\right).$$

In fact, we have

$$\int f_i(x)d\widetilde{G}_{n_1,n_2}(x)=\int f_i(x)\ d\left[p\cdot\left(F^{\{n_1,n_2\}}(x)-F^{\{y_{n_1},y_{n_2}\}}(x)\right)\right]$$

$$=\sum_{j=1}^{p}f_i(\lambda_j)-p\cdot\int_{a_n}^{b_n}\frac{f_i(x)\cdot(1-y_{n_2})\sqrt{(b_n-x)(x-a_n)}}{2\pi x\cdot(y_{n_1}+y_{n_2}x)}dx$$

for $i=1,\cdots,k$, where $a_n=\frac{(1-h_n)^2}{(1-y_{n_2})^2}$ and $b_n=\frac{(1+h_n)^2}{(1-y_{n_2})^2}$, $h_n^2=y_{n_1}+y_{n_2}-y_{n_1}y_{n_2}$, and the λ_j's are the eigenvalues of the F-matrix $\mathbf{S}_1\mathbf{S}_2^{-1}$.

We shall establish the following theorem due to Zheng [310].

Theorem 9.14. *Assume that the X-variables satisfy the condition*

$$\frac{1}{n_1p}\sum_{ij}E|X_{ij}^4|\cdot I(|X_{ij}|\geq\sqrt{n}\eta)\to 0,$$

for any fixed $\eta>0$, and the Y-variables satisfy a similar condition. In addition, we assume:

(a) $\{X_{i_1j_1},Y_{i_2j_2},i_1,j_1,i_2,j_2\}$ are independent. The moments satisfy $EX_{i_1j_1}=EY_{i_2j_2}=0$, $E|X_{i_1j_1}|^2=E|Y_{i_2j_2}|^2=1$, $E|X_{i_1j_1}|^4=\beta_x+\kappa+1$, and $E|X_{i_2j_2}|^4=\beta_y+\kappa+1$, where $\kappa=2$ if both the X-variables and Y-variables are real and $\kappa=1$ if they are complex. Furthermore, we assume $EX_{i_1j_1}^2=0$ and $EY_{i_2j_2}^2=0$ if the variables are all complex.

(b) $y_{n_1}=\frac{p}{n_1}\to y_1\in(0,+\infty)$ and $y_{n_2}=\frac{p}{n_2}\to y_2\in(0,1)$.

(c) $f_1,\cdots,f_k$ are functions analytic in an open region containing the interval $[a,b]$, where

$$a=\frac{(1-\sqrt{y_1})^2}{(1-y_2)^2}\quad\textit{and}\quad b=\frac{(1+\sqrt{y_1})^2}{(1-y_2)^2}.$$

Let $h^2=y_1+y_2-y_1y_2$. Then the random vector

$$\begin{pmatrix} \int f_1(x)d\widetilde{G}_{n_1,n_2}(x) \\ \vdots \\ \int f_k(x)d\widetilde{G}_{n_1,n_2}(x) \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^{p} f_1(\lambda_j) - p\int_{a_n}^{b_n} \frac{f_1(x)(1-y_{n_2})\sqrt{(b_n-x)(x-a_n)}}{2\pi x\cdot(y_{n_1}+y_{n_2}x)}dx \\ \vdots \\ \sum_{j=1}^{p} f_k(\lambda_j) - p\int_{a_n}^{b_n} \frac{f_k(x)(1-y_{n_2})\sqrt{(b_n-x)(x-a_n)}}{2\pi x\cdot(y_{n_1}+y_{n_2}x)}dx \end{pmatrix}$$

converges weakly to a Gaussian vector $(X_{f_1},\cdots,X_{f_k})'$ *with means*

$$\mathrm{E}X_f = \lim_{r\downarrow 1}\left[(9.13.1)+(9.13.2)+(9.13.3)+(9.13.4)\right],$$

where

$$\frac{\kappa-1}{4\pi i}\oint_{|\xi|=1} f_i\left(\frac{|1+h\xi|^2}{(1-y_2)^2}\right)\left[\frac{1}{\xi-r^{-1}}+\frac{1}{\xi+r^{-1}}-\frac{1}{\xi-\frac{\sqrt{y_2}}{h}}-\frac{1}{\xi+\frac{\sqrt{y_2}}{h}}\right]\ d\xi \tag{9.13.1}$$

$$\frac{\beta_x\cdot y_1(1-y_2)^2}{2\pi i\cdot h^2}\oint_{|\xi|=1} f_i\left(\frac{|1+h\xi|^2}{(1-y_2)^2}\right)\frac{1}{(\xi+\frac{y_2}{h})^3}\ d\xi \tag{9.13.2}$$

$$\frac{\kappa-1}{4\pi i}\oint_{|\xi|=1} f_i\left(\frac{|1+h\xi|^2}{(1-y_2)^2}\right)\left[\frac{1}{\xi-\frac{\sqrt{y_2}}{h}}+\frac{1}{\xi+\frac{\sqrt{y_2}}{hr}}-\frac{2}{\xi+\frac{y_2}{h}}\right]\ d\xi \tag{9.13.3}$$

$$\frac{\beta_y\cdot(1-y_2)}{4\pi i}\oint_{|\xi|=1} f_i\left(\frac{|1+h\xi|^2}{(1-y_2)^2}\right)\frac{\xi^2-\frac{y_2}{h^2}}{(\xi+\frac{y_2}{h})^2}\left[\frac{1}{\xi-\frac{\sqrt{y_2}}{h}}\frac{1}{\xi+\frac{\sqrt{y_2}}{h}}-\frac{2}{\xi+\frac{y_2}{h}}\right]\ d\xi \tag{9.13.4}$$

and covariance functions

$$\mathrm{Cov}(X_{f_i},X_{f_j}) = \lim_{r\downarrow 1}\left[(9.13.5)+(9.13.6)+(9.13.7)\right],$$

where

$$-\frac{\kappa}{4\pi^2}\oint_{|\xi_1|=1}\oint_{|\xi_2|=1}\frac{f_i\left(\frac{|1+h\xi_1|^2}{(1-y_2)^2}\right)f_j\left(\frac{|1+h\xi_2|^2}{(1-y_2)^2}\right)}{(\xi_1-r\xi_2)^2}\ d\xi_1 d\xi_2 \tag{9.13.5}$$

$$-\frac{\beta_x\cdot y_1(1-y_2)^2}{4\pi^2h^2}\oint_{|\xi_1|=1}\frac{f_i\left(\frac{|1+h\xi_1|^2}{(1-y_2)^2}\right)}{(\xi_1+\frac{y_2}{h})^2}d\xi_1\oint_{|\xi_2|=1}\frac{f_j\left(\frac{|1+h\xi_2|^2}{(1-y_2)^2}\right)}{(\xi_2+\frac{y_2}{h})^2}d\xi_2 \tag{9.13.6}$$

$$-\frac{\beta_y \cdot y_2(1-y_2)^2}{4\pi^2 h^2} \oint_{|\xi_1|=1} \frac{f_i\left(\frac{|1+h\xi_1|^2}{(1-y_2)^2}\right)}{(\xi_1+\frac{y_2}{h})^2} d\xi_1 \oint_{|\xi_2|=1} \frac{f_j\left(\frac{|1+h\xi_2|^2}{(1-y_2)^2}\right)}{(\xi_2+\frac{y_2}{h})^2} d\xi_2. \quad (9.13.7)$$

9.14 Proof of Theorem 9.14

Before proceeding with the proof of Theorem 9.14, we first present some lemmas.

9.14.1 Lemmas

Throughout this section, we assume that both the X- and Y-variables are truncated and renormalized as described in the next subsection. Also, we will use the notation defined in Subsections 6.2.2 and 6.2.3 with $\mathbf{T} = \mathbf{S}_2^{-1}$.

Lemma 9.15. *Suppose the conditions of Theorem 9.14 hold. Then, for any z with $\Im z > 0$, we have*

$$\max_{i,j} \left| \mathbf{e}_i' \mathrm{E}_j \mathbf{T}^{1/2} \mathbf{D}_j^{-1}(z) \mathbf{T}^{1/2} \mathbf{e}_i + \frac{1+z\underline{s}(z)}{zy_1\underline{s}(z)} \right| \xrightarrow{i.p.} 0 \quad (9.14.1)$$

and

$$\max_{i,j} \left| \mathbf{e}_i' \mathrm{E}_j \mathbf{T}^{1/2} \mathbf{D}_j^{-1}(z) \mathbf{T}^{1/2} \left(\underline{s}(z)\mathbf{T}+\mathbf{I}\right)^{-1} \mathbf{e}_i + \frac{1}{z}\int \frac{x \cdot dF_{y_2}(x)}{(x+\underline{s}(z))^2} \right| \xrightarrow{i.p.} 0, \quad (9.14.2)$$

where $\mathbf{e}_i = (\underbrace{0,\cdots,0}_{i-1},1,0,\cdots,0)'$ *and* E_j *denotes the conditional expectation with respect to the* σ*-field generated by* $\mathbf{X}_{\cdot 1}, \cdots, \mathbf{X}_{\cdot j}$ *and* $\mathbf{S}_2$ *with the convention that* E_0 *is the conditional expectation given* $\mathbf{S}_2$*.*

Similarly, we have

$$\max_{i,j} \left| \mathbf{e}_i' \mathrm{E}_{-j} \left(\mathbf{S}_2 - \frac{1}{n_2}\mathbf{Y}_{\cdot j}\mathbf{Y}_{\cdot j}^* - z\mathbf{I}\right)^{-1} \mathbf{e}_i + \frac{1}{z}\cdot\frac{1}{\underline{s}_{y_2}(z)+1} \right| \to 0 \quad \text{in } p, \quad (9.14.3)$$

where E_{-j} *for* $j \in [1, n_2]$ *denotes the conditional expectation given* $\mathbf{Y}_j, \mathbf{Y}_{j+1}, ..., \mathbf{Y}_{n_2}$*, while* E_{-n_2-1} *denotes unconditional expectation.*

Proof. First, we claim that for any random matrix $\mathbf{M}$ with a nonrandom bound $\|\mathbf{M}\| \le K$, for any fixed $t > 0$, $i \le p$, and z with $|\Im z| = v > 0$, we have

$$\mathrm{P}(\sup_{j\le n_2}|\mathrm{E}_j\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(z)\mathbf{T}^{1/2}\mathbf{M}\mathbf{e}_i-\mathrm{E}_j\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}\mathbf{M}\mathbf{e}_i|\ge\varepsilon)=o(p^{-t}). \tag{9.14.4}$$

In fact,

$$\begin{aligned}&|\mathrm{E}_j\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(z)\mathbf{T}^{1/2}\mathbf{M}\mathbf{e}_i-\mathrm{E}_j\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}\mathbf{M}\mathbf{e}_i|\\ &=\left|\mathrm{E}_j\frac{\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(z)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{T}^{1/2}\mathbf{M}\mathbf{e}_i}{1+\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j}\right|\le K\mathrm{E}_j|\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(z)\mathbf{r}_j|^2.\end{aligned}$$

By noting $\frac{1}{n}\left|\mathbf{e}_i'(\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(z)\mathbf{T}\mathbf{D}_j^{-1}(z)\mathbf{T}^{1/2})\mathbf{e}_i\right|\le K/n$ and applying Lemma 9.1 by choosing $l=[\log n]$, one can easily prove (9.14.4).

To show the convergence of (9.14.1), we consider

$$\mathbf{e}_i'\mathbf{T}^{1/2}\mathrm{E}_j\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}\mathbf{e}_i=\mathrm{E}_j\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}\mathbf{e}_i.$$

Note that

$$\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}=(\mathbf{S}_1-z\cdot\mathbf{S}_2)^{-1}\equiv\widetilde{\mathbf{D}}^{-1}(z).$$

That is, the limits of the diagonal elements of

$$\mathrm{E}_j\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}=\mathbf{E}_j\widetilde{\mathbf{D}}^{-1}(z)$$

are identical. To this end, employing Kolmogorov's inequality for martingales, we have

$$\begin{aligned}I_{i,l}&\equiv\mathrm{P}\left(\sup_{-n_2\le j\le n_1}\left|\mathrm{E}_j\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}\mathbf{e}_i-\mathrm{E}\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}\mathbf{e}_i\right|\ge\varepsilon\right)\\ &\le\varepsilon^{-4}\mathrm{E}\left|\mathrm{E}_{n_1}\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}\mathbf{e}_i-\mathrm{E}_{-n_2-1}\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}\mathbf{e}_i\right|^4\\ &=\varepsilon^{-4}\mathrm{E}\left|\sum_{k=-n_2}^{n_1}(\mathrm{E}_k-\mathrm{E}_{k-1})\mathbf{e}_i'\widetilde{\mathbf{D}}^{-1}(z)\mathbf{e}_i\right|^4\\ &=\mathrm{E}\left|\sum_{k=-n_2}^{n_1}(\mathrm{E}_k-\mathrm{E}_{k-1})\mathbf{e}_i'\left(\widetilde{\mathbf{D}}^{-1}(z)-\widetilde{\mathbf{D}}_{j\ k}^{-1}(z)\right)\mathbf{e}_i\right|^4,\end{aligned}$$

where

$$\widetilde{\mathbf{D}}_k=\begin{cases}\widetilde{\mathbf{D}}-\frac{1}{n_1}\mathbf{X}_k\mathbf{X}_k^*, & \text{if } k>0,\\ \widetilde{\mathbf{D}}+\frac{z}{n_2}\mathbf{Y}_{-k+1}\mathbf{Y}_{-k+1}^*, & \text{if } k\le 0.\end{cases}$$

Thus, by Burkholder's inequality,

$$I_{i,l}\le\frac{K}{\varepsilon^4n_1^4}\mathrm{E}\left|\sum_{k=1}^{n_1}(\mathrm{E}_k-\mathrm{E}_{k-1})\frac{\mathbf{e}_i'\widetilde{\mathbf{D}}_k^{-1}\mathbf{X}_{\cdot k}\mathbf{X}_{\cdot k}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{e}_i}{1+\mathbf{X}_{\cdot k}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{X}_{\cdot k}/n_1}\right|^4$$

$$
\begin{aligned}
&+\frac{K}{\varepsilon^4 n_2^4}\mathrm{E}\left|\sum_{k=-n_2}^{0}(\mathrm{E}_k-\mathrm{E}_{k-1})\frac{ze_i'\widetilde{\mathbf{D}}_{-k+1}^{-1}\mathbf{Y}_{\cdot -k+1}\mathbf{Y}_{\cdot -k+1}^*\widetilde{\mathbf{D}}_{-k+1}^{-1}\mathbf{e}_i}{1-z\mathbf{Y}_{\cdot -k+1}^*\widetilde{\mathbf{D}}_{-k+1}^{-1}\mathbf{Y}_{\cdot -k+1}/n_2}\right|^4\\
&\le \frac{K}{\varepsilon^4 n_1^4}\cdot\left[\mathrm{E}\left(\sum_{k=1}^{n_1}\mathrm{E}_{k-1}\left|\frac{\mathbf{e}_i'\widetilde{\mathbf{D}}_k^{-1}\mathbf{X}_{\cdot k}\mathbf{X}_{\cdot k}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{e}_i}{1+\mathbf{X}_{\cdot k}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{X}_{\cdot k}/n_1}\right|^2\right)^2\right.\\
&\left.+\sum_{k=1}^{n_1}\mathrm{E}\left|\frac{\mathbf{e}_i'\widetilde{\mathbf{D}}_k^{-1}\mathbf{X}_{\cdot k}\mathbf{X}_{\cdot k}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{e}_i}{1+\mathbf{X}_{\cdot k}^*\widetilde{\mathbf{D}}_{jk}^{-1}\mathbf{X}_{\cdot k}/n_1}\right|^4\right]\\
&+\frac{K}{\varepsilon^4 n_2^4}\left[\mathrm{E}\left(\sum_{k=-n_2}^{0}\mathrm{E}_{k-1}\left|\frac{z\mathbf{e}_i'\widetilde{\mathbf{D}}_{-k+1}^{-1}\mathbf{Y}_{\cdot -k+1}\mathbf{Y}_{\cdot -k+1}^*\widetilde{\mathbf{D}}_{-k+1}^{-1}\mathbf{e}_i}{1-z\mathbf{Y}_{\cdot -k+1}^*\widetilde{\mathbf{D}}_{-k+1}^{-1}\mathbf{Y}_{\cdot -k+1}/n_2}\right|^2\right)^2\right.\\
&\left.+\sum_{k=-n_2}^{0}\left|\mathrm{E}\frac{z\mathbf{e}_i'\widetilde{\mathbf{D}}_{-k+1}^{-1}\mathbf{Y}_{\cdot -k+1}\mathbf{Y}_{\cdot -k+1}^*\widetilde{\mathbf{D}}_{-k+1}^{-1}\mathbf{e}_i}{1-z\mathbf{Y}_{\cdot -k+1}^*\widetilde{\mathbf{D}}_{-k+1}^{-1}\mathbf{Y}_{\cdot -k+1}/n_2}\right|^4\right].
\end{aligned}
$$

When $k>0$ and $|\Im z|=v>0$, (i.e., z is on the horizontal part of the contour $\mathcal{C}$), it has been proved that

$$
\left|\frac{1}{1+\mathbf{X}_{\cdot k}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{X}_{\cdot k}/n_1}\right|\le\frac{|z|}{v}.
$$

Therefore, by Lemma 9.1, we have

$$
\begin{aligned}
&\frac{K}{\varepsilon^4 n_1^4}\cdot\mathrm{E}\left(\sum_{k=1}^{n_1}\mathrm{E}_{k-1}\left|\frac{\mathbf{e}_i'\widetilde{\mathbf{D}}_k^{-1}\mathbf{X}_{\cdot k}\mathbf{X}_{\cdot k}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{e}_i}{1+\mathbf{X}_{\cdot k}^*\widetilde{\mathbf{D}}_{jk}^{-1}\mathbf{X}_{\cdot k}/n_1}\right|^2\right)^2\\
&\le\frac{K}{\varepsilon^4 n_1^4}\cdot\mathrm{E}\left(\sum_{k=1}^{n_1}\mathrm{E}_{k-1}\left|\mathbf{e}_i'\widetilde{\mathbf{D}}_k^{-1}\mathbf{X}_{\cdot k}\mathbf{X}_{\cdot k}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{e}_i\right|^2\right)^2=O(n_1^{-2})
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{K}{n_1^4}\cdot\sum_{k=1}^{n_1}\mathrm{E}\left|\frac{\mathbf{e}_i'\widetilde{\mathbf{D}}_k^{-1}\mathbf{X}_{\cdot k}\mathbf{X}_{\cdot 1}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{e}_i}{1+\mathbf{X}_{\cdot k}\widetilde{\mathbf{D}}_k^{-1}\mathbf{X}_{\cdot k}/n_1}\right|^4\\
&\le\frac{K|z|^4}{v^4 n_1^3}\cdot\sum_{k=1}^{n_1}\mathrm{E}\left|\mathbf{e}_i'\widetilde{\mathbf{D}}_k^{-1}\mathbf{X}_{\cdot k}\mathbf{X}_{\cdot 1}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{e}_i\right|^4=o(n_1^{-1}).
\end{aligned}
$$

Furthermore, by noting that

$$
\left|\frac{1}{1-z\mathbf{Y}_{\cdot -k+1}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{Y}_{\cdot -k+1}/n_2}\right|=\left|\frac{\bar z}{\bar z-|z|^2\mathbf{Y}_{\cdot -k+1}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{Y}_{\cdot -k+1}/n_2}\right|\le\frac{|z|}{v},
$$

we can similarly prove that

$$\frac{K}{\varepsilon^4 n_2^4}\mathrm{E}\left(\sum_{k=-n_2}^{0}\mathrm{E}_{k-1}\left|\frac{z\mathbf{e}_i'\widetilde{\mathbf{D}}_{-k+1}^{-1}\mathbf{Y}_{\cdot -k+1}\mathbf{Y}_{\cdot -k+1}^*\widetilde{\mathbf{D}}_{-k+1}^{-1}\mathbf{e}_i}{1-z\mathbf{Y}_{\cdot -k+1}^*\widetilde{\mathbf{D}}_{-k+1}^{-1}\mathbf{Y}_{\cdot -k+1}/n_2}\right|^2\right)^2 = O(n_2^{-2})$$

and

$$\frac{K|z|^4}{n_2^4}\cdot\sum_{k=-n_2}^{0}\mathrm{E}\left|\frac{\mathbf{e}_i'\widetilde{\mathbf{D}}_k^{-1}\mathbf{Y}_{\cdot -k+1}\mathbf{Y}_{\cdot -k+1}^*\widetilde{\mathbf{D}}_{jk}^{-1}\mathbf{e}_i}{1-z\mathbf{Y}_{\cdot -k+1}^*\widetilde{\mathbf{D}}_k^{-1}\mathbf{Y}_{\cdot -k+1}/n_2}\right|^4 = o(n_2^{-1}).$$

We therefore obtain

$$\max_{i,j}\left|\mathrm{E}_j\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}\mathbf{e}_i - \mathrm{E}\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}e_i\right| \to 0 \text{ in p.}$$

If the X and Y variables are identically distributed, then

$$E\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}\mathbf{e}_i = \frac{1}{p}\mathrm{tr}\mathrm{E}\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}.$$

Similar to (9.9.20), we have

$$\begin{aligned}\mathrm{E}\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)\mathbf{T}^{1/2}\mathbf{e}_i &= \frac{1}{p}\mathrm{E}\left[\mathrm{tr}\left(\mathbf{T}\mathbf{D}^{-1}(z)\right)\right] = \frac{n_1}{p}\cdot\left(\frac{1}{b_p(z)}-1\right)\\ &\to -\frac{1+z\underline{s}(z)}{zy_1\underline{s}(z)} = -\int\frac{dF_{y_2}(x)}{z(x+\underline{s})}.\end{aligned}$$

Thus, (9.14.1) follows.

We should in fact show that the limit above holds true under the conditions of Theorem 9.14. Let $\widetilde{\mathbf{D}}_{j,w} = \widetilde{\mathbf{D}} - \frac{1}{n_1}\mathbf{X}_j\mathbf{X}_j^* + \frac{1}{n_1}\mathbf{W}_j\mathbf{W}_j^*$, where $\mathbf{W}_j$ consists of iid entries distributed as X_{11}; that is, we change the j-th term $\frac{1}{n_1}\mathbf{X}_j\mathbf{X}_j^*$ with an analogue $\frac{1}{n_1}\mathbf{W}_j\mathbf{W}_j^*$ with iid entries. We have

$$\begin{aligned}&\mathrm{E}\mathbf{e}_i'\widetilde{\mathbf{D}}^{-1}\mathbf{e}_i - \mathrm{E}\mathbf{e}_i'\widetilde{\mathbf{D}}_{j,w}^{-1}\mathbf{e}_i = \mathrm{E}\mathbf{e}_i'(\widetilde{\mathbf{D}}^{-1}-\widetilde{\mathbf{D}}_j^{-1})\mathbf{e}_i - \mathrm{E}\mathbf{e}_i'(\widetilde{\mathbf{D}}_{j,w}^{-1}-\widetilde{\mathbf{D}}_j^{-1})\mathbf{e}_i\\ &= n_1^{-1}\mathrm{E}\mathbf{e}_i'\widetilde{\mathbf{D}}_j^{-1}\left[\mathbf{X}_j\mathbf{X}_j^*\beta_j - \mathbf{W}_j\mathbf{W}_j^*\beta_{j,w}\right]\widetilde{\mathbf{D}}_j^{-1}\mathbf{e}_i,\end{aligned}$$

where $\beta_{j,w} = (1+n_1^{-1}\mathbf{W}_j^*\widetilde{\mathbf{D}}_j^{-1}\mathbf{W}_j)^{-1}$. Let $\hat{\beta}_j = (1+n_1^{-1}\mathrm{tr}\widetilde{\mathbf{D}}_j^{-1})^{-1}$, $\hat{\gamma}_j = n_1^{-1}[\mathbf{X}_j^*\widetilde{\mathbf{D}}_j^{-1}\mathbf{X}_j - \mathrm{tr}\widetilde{\mathbf{D}}_j]$, and $\hat{\gamma}_{j,w} = n_1^{-1}[\mathbf{W}_j^*\widetilde{\mathbf{D}}_j^{-1}\mathbf{W}_j - \mathrm{tr}\widetilde{\mathbf{D}}_j]$. Noting that $\beta_{j,w} = \hat{\beta}_j - \hat{\beta}_j\beta_{j,w}\hat{\gamma}_{j,w}$ and a similar decomposition for β_j, we have

$$\begin{aligned}&\left|\mathrm{E}\mathbf{e}_i'\widetilde{\mathbf{D}}^{-1}\mathbf{e}_i - \mathrm{E}\mathbf{e}_i'\widetilde{\mathbf{D}}_{j,w}^{-1}\mathbf{e}_i\right|\\ &= n_1^{-1}\left|\mathrm{E}\mathbf{e}_i'\widetilde{\mathbf{D}}_j^{-1}\left[\mathbf{X}_j\mathbf{X}_j^*\hat{\beta}_j\beta_j\hat{\gamma}_j - \mathbf{W}_j\mathbf{W}_j^*\beta_{j,w}\hat{\beta}_j\hat{\gamma}_{j,w}\right]\widetilde{\mathbf{D}}_j^{-1}\mathbf{e}_i\right|\end{aligned}$$

$$\leq \frac{K}{n_1}\left[\left(\mathrm{E}|\mathbf{e}_i'\widetilde{\mathbf{D}}_j^{-1}\mathbf{X}_j|^4\mathrm{E}|\hat{\gamma}_j|^2\right)^{1/2} + \left(\mathrm{E}|\mathbf{e}_i'\widetilde{\mathbf{D}}_j^{-1}\mathbf{W}_j|^4\mathrm{E}|\hat{\gamma}_{j,w}|^2\right)^{1/2}\right]$$
$$= O(n_1^{-3/2}).$$

Using the same approach, we can replace all terms $\frac{1}{n_1}\mathbf{X}_j\mathbf{X}_j^*$ in $\mathbf{S}_1$ by $\frac{1}{n_1}\mathbf{W}_j\mathbf{W}_j^*$. The total error will be bounded by $O(n_1^{-1/2})$. Also, we replace all terms in $\mathbf{S}_2$ by iid entries with a total error bounded by $O(n_2^{-1/2})$. Then, using the argument in the last paragraph, we can show that (9.14.1) holds under the conditions of Theorem 9.14.

In (9.14.1), letting $y_2 \to 0$ or $\mathbf{T} = \mathbf{I}$, we obtain

$$\max_{i,j}\left|\mathbf{e}_i'\mathrm{E}_j\left(\mathbf{S}_1 - z\mathbf{I}\right)^{-1}\mathbf{e}_i + \frac{1}{z(\underline{s}_{y_1}(z)+1)}\right| \to 0 \quad \text{in p.}$$

By symmetry of $\mathbf{S}_1$ and $\mathbf{S}_2$, we have

$$\max_{i,j}\left|\mathbf{e}_i'\mathrm{E}_{j,y}\left(\mathbf{S}_2 - z\mathbf{I}\right)^{-1}\mathbf{e}_i + \frac{1}{z(\underline{s}_{y_2}(z)+1)}\right| \to 0 \quad \text{in p.}$$

This proves (9.14.3). Note that $-\frac{1}{z(\underline{s}_{y_2}(z)+1)} = \underline{s}_{y_2}(z)$.

Finally, we consider the limits of $\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}^{1/2}\mathbf{e}_i$. Using the decomposition (9.9.12) and the similar arguments (9.9.13)–(9.9.16), one can prove that

$$\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}^{-1}(z)(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}^{1/2}\mathbf{e}_i = -z^{-1}\mathbf{e}_i'\mathbf{T}^{1/2}(\mathbf{I}+\underline{s}(z)\mathbf{T})^{-2}\mathbf{T}^{1/2}\mathbf{e}_i + o(1),$$

where $o(1)$ is uniform in $i \leq p$. To find the limit of the RHS of the above, we note that

$$\begin{aligned}
&\mathbf{e}_i'\mathbf{T}^{1/2}(\mathbf{I}+\underline{s}(z)\mathbf{T})^{-2}\mathbf{T}^{1/2}\mathbf{e}_i = \mathbf{e}_i'(\mathbf{S}_2+\underline{s}(z)\mathbf{I})^{-2}\mathbf{S}_2\mathbf{e}_i\\
&= \mathbf{e}_i'(\mathbf{S}_2+\underline{s}(z)\mathbf{I})^{-1}\mathbf{e}_i - \underline{s}(z)\mathbf{e}_i'(\mathbf{S}_2+\underline{s}(z)\mathbf{I})^{-2}\mathbf{e}_i\\
&= \mathbf{e}_i'(\mathbf{S}_2+\underline{s}(z)\mathbf{I})^{-1}\mathbf{e}_i + \underline{s}(z)\frac{d}{ds}\mathbf{e}_i'(\mathbf{S}_2+s\mathbf{I})^{-1}\mathbf{e}_i|s=\underline{s}(z).
\end{aligned}$$

By (9.14.3), we have

$$\begin{aligned}
&\mathbf{e}_i'\mathbf{T}^{1/2}(\mathbf{I}+\underline{s}(z)\mathbf{T})^{-2}\mathbf{T}^{1/2}\mathbf{e}_i\\
&\to \int\frac{dF_{y_2}(x)}{(x+\underline{s}(z))^2} - \int\frac{\underline{s}(z)dF_{y_2}(x)}{(x+\underline{s}(z))^2} = \int\frac{xdF_{y_2}(x)}{(x+\underline{s}(z))^2}.
\end{aligned}$$

This is (9.14.2). The proof of Lemma 9.15 is complete.

Lemma 9.16. *Let $s_0(z) = \underline{s}_{y_2}(-\underline{s}(z))$. Then the following identities hold*

$$z = -\frac{s_0(z)(s_0(z)+1-y_1)}{(1-y_2)s_0(z)+1} \quad \text{and} \quad \underline{s}(z) = \frac{(1-y_2)s_0(z)+1}{s_0(z)(s_0(z)+1)},$$

$$1 - y_1 \int \frac{\underline{s}^2(z)dF_{y_2}(x)}{(x+\underline{s}(z))^2} = \frac{(1-y_2)s_0^2(z)+2s_0(z)+1-y_1}{(1-y_2)s_0^2(z)+2s_0(z)+1},$$

$$\int \frac{dF_{y_2}(x)}{x+\underline{s}(z)} = \frac{s_0(z)}{(1-y_2)s_0(z)+1},$$

$$\int \frac{x \cdot dF_{y_2}(x)}{(x+\underline{s}(z))^2} = \frac{s_0^2(z)}{(1-y_2)s_0^2(z)+2s_0(z)+1},$$

$$s_0'(z) = -\frac{((1-y_2)s_0(z)+1)^2}{(1-y_2)s_0^2(z)+2s_0(z)+1-y_1},$$

and

$$\begin{aligned}\underline{s}'(z) &= \frac{-(1-y_2)\left(s_0(z)+\frac{1}{1-y_2}\right)^2+\frac{y_2}{(1-y_2)}}{s_0^2(z)\cdot(s_0(z)+1)^2}\cdot s_0'(z)\\ &= -\frac{(1-y_2)s_0^2+2s_0+1}{s_0^2(s_0+1)^2}\cdot s_0'(z).\end{aligned}$$

Proof. Because

$$\underline{s}_{y_2}(z) = -\frac{1-y_2}{z} + y_2 \cdot s_{y_2}(z), \tag{9.14.5}$$

we get $s_{y_2}'(z) = -\frac{1-y_2}{y_2}\cdot\frac{1}{z^2}+\frac{1}{y_2}\cdot\underline{s}_{y_2}'(z)$; that is,

$$s_{y_2}'(-\underline{s}(z)) = -\frac{1-y_2}{y_2}\frac{1}{(\underline{s}(z))^2}+\frac{1}{y_2}\cdot\underline{s}_{y_2}'(-\underline{s}(z)).$$

So we have

$$\int \frac{dF_{y_2}(x)}{(x+\underline{s}(z))^2} = s_{y_2}'(-\underline{s}(z)) = -\frac{1-y_2}{y_2}\cdot\frac{1}{(\underline{s}(z))^2}+\frac{1}{y_2}\cdot\underline{s}_{y_2}'(-\underline{s}(z)). \tag{9.14.6}$$

Therefore,

$$\begin{aligned} z &= -\frac{1}{\underline{s}(z)} + \int \frac{y_1 dF_{y_2}(x)}{x+\underline{s}(z)}\\ &= -\frac{1}{\underline{s}(z)} - \frac{y_1(1-y_2)}{y_2\underline{s}(z)} + \frac{y_1}{y_2}\underline{s}_{y_2}(-\underline{s}(z)) \qquad (9.14.7)\\ &= \frac{y_1+y_2-y_1y_2}{y_2}\cdot\frac{1}{-\underline{s}(z)}+\frac{y_1}{y_2}\underline{s}_{y_2}(-\underline{s}(z)).\end{aligned}$$

Using the notation $h^2 = y_1 + y_2 - y_1y_2$ and differentiating both sides of the identity above, we obtain

$$1 = \frac{h^2}{y_2(\underline{s}(z))^2}\underline{s}'(z) - \frac{y_1}{y_2}\underline{s}'_{y_2}(-\underline{s}(z))\underline{s}'(z).$$

This implies $\underline{s}'(z) = \frac{y_2(\underline{s}(z))^2}{h^2 - y_1(\underline{s}(z))^2\underline{s}'_{y_2}(-\underline{s}(z))}$ or

$$y_1(\underline{s}(z))^2\underline{s}'_{y_2}(-\underline{s}(z)) = h^2 - \frac{y_2(\underline{s}(z))^2}{\underline{s}'(z)}. \tag{9.14.8}$$

We herewith remind the reader that $\underline{s}'_{y_2}(-\underline{s}(z)) = \frac{d}{d\xi}\underline{s}_{y_2}(\xi)|_{\xi=-\underline{s}(z)}$ instead of $\frac{d}{dz}\underline{s}_{y_2}(-\underline{s}(z))$. So, by (9.14.6) and (9.14.8), we have

$$1 - y_1\int\frac{(\underline{s}(z))^2 dF_{y_2}(x)}{(x+\underline{s}(z))^2} = \frac{h^2}{y_2} - \frac{y_1(\underline{s}(z))^2\underline{s}'_2(-\underline{s}(z))}{y_2} = \frac{(\underline{s}(z))^2}{\underline{s}'(z)}. \tag{9.14.9}$$

The Stieltjes transform $\underline{s}_2(z)$ satisfies $z = -\frac{1}{\underline{s}_{y_2}(z)} + \frac{y_2}{1+\underline{s}_{y_2}(z)}$. Differentiating both sides, we obtain $1 = \left(\frac{1}{(\underline{s}_{y_2}(z))^2} - \frac{y_2}{(1+\underline{s}_{y_2}(z))^2}\right)\underline{s}'_{y_2}(z)$. Therefore, $\underline{s}'_{y_2}(z) = \frac{(\underline{s}_{y_2}(z))^2}{1-y_2(\underline{s}_{y_2}(z))^2(1+\underline{s}_{y_2}(z))^{-2}}$ and thus

$$\underline{s}'_{y_2}(-\underline{s}(z)) = \frac{[\underline{s}_{y_2}(-\underline{s}(z))]^2}{1 - y_2\cdot[\underline{s}_{y_2}(-\underline{s}(z))]^2\cdot[1+\underline{s}_{y_2}(-\underline{s}(z))]^{-2}}. \tag{9.14.10}$$

Because $z = -\frac{1}{\underline{s}_{y_2}(z)} + \frac{y_2}{1+\underline{s}_{y_2}(z)}$, then we have

$$-\underline{s}(z) = -\frac{1}{\underline{s}_{y_2}(-\underline{s}(z))} + \frac{y_2}{1+\underline{s}_{y_2}(-\underline{s}(z))} = -\frac{(1-y_2)\underline{s}_{y_2}(-\underline{s}(z))+1}{\underline{s}_{y_2}(-\underline{s}(z))\cdot(\underline{s}_{y_2}(-\underline{s}(z))+1)}.$$

Recall that $s_0 = \underline{s}_{y_2}(-\underline{s}(z))$. By (9.14.7), we then obtain the first two conclusions of the lemma,

$$\underline{s}(z) = \frac{(1-y_2)s_0+1}{s_0\cdot(s_0+1)} \quad \text{and} \quad z = -\frac{s_0(s_0+1-y_1)}{(1-y_2)s_0+1}. \tag{9.14.11}$$

Differentiating the second identity in (9.14.11), we obtain

$$1 = -\frac{[(2s_0+1-y_1)\left((1-y_2)s_0+1\right) - s_0(s_0+1-y_1)(1-y_2)]\,\underline{s}'_0}{\left((1-y_2)s_0+1\right)^2}.$$

Solving $\underline{s}'_0$, we obtain the sixth assertion of the lemma

$$\underline{s}'_0 = -\frac{\left((1-y_2)s_0+1\right)^2}{(1-y_2)s_0^2+2s_0+1-y_1}.$$

By the identity $\underline{s}_{y_2}(-\underline{s}(z)) = \frac{1-y_2}{\underline{s}(z)} + y_2\cdot s_{y_2}(-\underline{s}(z))$, we obtain

$$\int \frac{dF_{y_2}(x)}{x+\underline{s}(z)} = s_{y_2}(-\underline{s}(z)) = \frac{\underline{s}_{y_2}(-\underline{s}(z))}{y_2} - \frac{1-y_2}{y_2}\cdot\frac{1}{\underline{s}(z)}$$
$$= \frac{\underline{s}_{y_2}(-\underline{s}(z))}{y_2} - \frac{1-y_2}{y_2}\cdot\frac{\underline{s}_{y_2}(-\underline{s}(z))(\underline{s}_{y_2}(-\underline{s}(z))+1)}{(1-y_2)\underline{s}_{y_2}(-\underline{s}(z))+1}$$
$$= \frac{\underline{s}_{y_2}(-\underline{s}(z))}{(1-y_2)\underline{s}_{y_2}(-\underline{s}(z))+1} = \frac{s_0}{(1-y_2)s_0+1}.$$

This is the fourth conclusion of the lemma. By (9.14.9), (9.14.10), and (9.14.11), we obtain the third conclusion of the lemma,

$$1-y_1\underline{s}^2\int\frac{dF_{y_2}(x)}{(x+\underline{s})^2} = \frac{h^2}{y_2} - \frac{y_1}{y_2}\frac{((1-y_2)s_0+1)^2\cdot s_0^2(1+s_0)^2}{s_0^2(s_0+1)^2\cdot[(1+s_0)^2-y_2\cdot s_0^2]}$$
$$= \frac{(1-y_2)s_0^2+2s_0+1-y_1}{(1-y_2)s_0^2+2s_0+1}, \tag{9.14.12}$$

where $s_0^2+\frac{2}{1-y_2}s_0+\frac{1-y_1}{1-y_2} = \left(s_0+\frac{1+h}{1-y_2}\right)\cdot\left(s_0+\frac{1-h}{1-y_2}\right)$. Thus,

$$\int\frac{x\cdot dF_{y_2}(x)}{(x+\underline{s}(z))^2} = \int\frac{dF_{y_2}(x)}{x+\underline{s}(z)} - \underline{s}(z)\int\frac{dF_{y_2}(x)}{(x+\underline{s}(z))^2}$$
$$= \frac{s_0}{(1-y_2)s_0+1} - \frac{s_0(s_0+1)}{[(1-y_2)s_0^2+2s_0+1]\cdot(1-y_2)(s_0+\frac{1}{1-y_2})}$$
$$= \frac{s_0^2}{(1-y_2)s_0^2+2s_0+1}.$$

This is the fifth line of the lemma.

By (9.14.11), we obtain the last line of the lemma,

$$\underline{s}'(z) = \frac{(1-y_2)\left(s_0(s_0+1)-(s_0+\frac{1}{1-y_2})(2s_0+1)\right)s_0'}{s_0^2(s_0+1)^2}$$

$$= \frac{-(1-y_2)\left(s_0+\frac{1}{1-y_2}\right)^2+\frac{y_2}{(1-y_2)}}{m_0^2\cdot(s_0+1)^2}\underline{s}_0' = -\frac{(1-y_2)s_0^2+2s_0+1}{s_0^2(s_0+1)^2}\cdot\underline{s}_0'.$$

The proof of Lemma 9.16 is completed.

Lemma 9.17. *Let $\mathbf{B}_n = \mathbf{S}_2^{-1/2}\mathbf{S}_1\mathbf{S}_2^{-1/2}$ with $\mathbf{T}=\mathbf{S}_2^{-1}$, under condition (a) of Theorem 9.14. When applying Lemma 9.11, the additional terms for the mean and covariance functions are*

$$\frac{\beta_x\cdot y_1\cdot\underline{s}^3(z)\cdot\int\frac{dF_{y_2}(x)}{x+\underline{s}(z)}\int\frac{x\cdot dF_{y_2}(x)}{(x+\underline{s}(z))^2}}{1-y_1\int\underline{s}^2(z)(x+\underline{s}(z))^{-2}dF_{y_2}(x)} \tag{9.14.13}$$

and

$$\beta_x \cdot y_1 \cdot \int \frac{\underline{s}'(z_1) \cdot x \cdot dF_{y_2}(x)}{(x+\underline{s}(z_1))^2} \int \frac{\underline{s}'(z_2) \cdot x \cdot dF_{y_2}(x)}{(x+\underline{s}(z_2))^2}. \tag{9.14.14}$$

For $\mathbf{B}_n = \mathbf{S}_2$ *with* $\mathbf{T} = \mathbf{I}$, *under condition (a) of Theorem 9.14, the additional terms for mean and covariance functions, with* z *replaced by* $-\underline{s}$, *reduce to*

$$\frac{\beta_y \cdot y_2 \cdot \underline{s}_{y_2}^3(-\underline{s}(z)) \cdot (1+\underline{s}_{y_2}(-\underline{s}(z)))^{-3}}{1 - y_2 \cdot \underline{s}_{y_2}^2(-\underline{s}(z)) \cdot (1+\underline{s}_{y_2}(-\underline{s}(z)))^{-2}} \tag{9.14.15}$$

and

$$\beta_y \cdot y_2 \cdot \frac{\underline{s}'_{y_2}(-\underline{s}(z_1))}{(1+\underline{s}_{y_2}(-\underline{s}(z_1)))^2} \cdot \frac{\underline{s}'_{y_2}(-\underline{s}(z_2))}{(1+\underline{s}_{y_2}(-\underline{s}(z_2)))^2}. \tag{9.14.16}$$

Proof. We consider the case $\mathbf{B}_n = \mathbf{T}^{1/2}\mathbf{S}_1\mathbf{T}^{1/2}$, where $\mathbf{T} = \mathbf{S}_2^{-1}$ and $\mathbf{D}_j(z) = \mathbf{B}_n - zI - \gamma_j\gamma_j^*$, $\gamma_j = \frac{1}{\sqrt{n_1}}\mathbf{T}^{1/2}\mathbf{X}_{\cdot j}$ in more general conditions. Going through the proof of Lemma 9.11 under the conditions of Theorem 9.14, we find that the process

$$M_n(z) = n_1\left[\underline{s}^{\{n_1,n_2\}}(z) - \underline{s}_{F^{\{y_{n_1},H_{n_2}\}}}(z)\right]$$

is still tight, where $\underline{s}_{F^{\{y_{n_1},H_{n_2}\}}}(z)$ is the unique root, which has the same sign for the imaginary part as that of z, to the equation

$$z = -\frac{1}{\underline{s}_{F^{\{y_{n_2},H_{n_2}\}}}} + y_{n_2}\int \frac{t}{1+t\underline{s}_{F^{\{y_{n_2},H_{n_2}\}}}} dH_{n_2}(t).$$

Also, its finite-dimensional distribution still satisfies the Lindeberg condition and thus $M_n(z)$ tends to a Gaussian process. Thus, we need only recalculate the asymptotic mean and covariance functions. Checking the proof of Lemma 9.11, one finds that the equations (9.9.7) and (9.11.1) give the covariance and mean functions of the limiting process $M(z)$ as

$$\mathrm{Cov}(M(z_1), M(z_2)) = \frac{\partial^2 D(z_1,z_2)}{\partial z_1 \partial z_2}$$

and

$$\mathrm{E}M(z) = \frac{-\underline{s}(z)A_p(z)}{1 - y_1\int \underline{s}^2(z)(x+\underline{s}(z))^{-2}dF_{y_2}(t)},$$

where $D(z_1,z_2)$ is the limit of

$$D_p(z_1,z_2) = b_p(z_1)b_p(z_2)\sum_{j=1}^{n_1}\mathrm{E}_{j-1}\left[\mathrm{E}_j\left(\frac{1}{n_1}\mathbf{X}_{\cdot j}^*\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(z_1)\mathbf{T}^{1/2}\mathbf{X}_{\cdot j} - \frac{1}{n_1}\mathrm{tr}\mathbf{T}\mathbf{D}_j^{-1}(z_1)\right) \times E_j\left(\frac{1}{n_1}\mathbf{X}_{\cdot j}^* T^{1/2}D_j^{-1}(z_2)T^{1/2}\mathbf{X}_{\cdot j} - \frac{1}{n_1}\mathrm{tr}TD_j^{-1}(z_2)\right)\right] \tag{9.14.17}$$

and $A(z)$ is the limit of

$$\begin{aligned} &A_p(z) \\ &= \frac{b_p^2}{n_1^2}\sum_{j=1}^{n}\Bigg\{\mathrm{E}\mathrm{tr}\mathbf{D}_j^{-1}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_j^{-1}\mathbf{T} - n_1^2\mathrm{E}\left[\left(\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j - \frac{1}{n_1}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{T}\right)\right. \\ &\quad \times\left.\left.\left(\mathbf{r}_j^*\mathbf{D}_j^{-1}(\underline{s}(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{r}_j - \frac{1}{n_1}\mathrm{tr}\mathbf{D}_j^{-1}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\right)\right]\right\}. \qquad (9.14.18) \end{aligned}$$

Applying (9.8.6) to the limits of D_p and A_p, under the conditions of Theorem 9.14, the limit of the term induced by the first term on the RHS of (9.8.6) should be added to the expression of the asymptotic covariance; that is, (9.14.14). Also, the limit of $\mathrm{tr}(\mathbf{T}^{1/2}\mathbf{D}_1^{-1}(z)(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}^{1/2}\circ\mathbf{T}^{1/2}\mathbf{D}_1^{-1}(z)\mathbf{T}^{1/2})$ should be added to the asymptotic mean; that is, (9.14.13). Please note that these terms may not have limits for general $\mathbf{T}$ as assumed in Theorem 9.10, but for $\mathbf{T}=\mathbf{S}_2^{-1}$ their limits do exist because of Lemma 9.15. Except for the terms of the mean and covariance functions given in Lemma 9.11, the additional terms to these functions are derived as follows.

We first consider $\mathrm{E}M(z)$. By Lemma 9.15, we have

$$\begin{aligned} A_p &= \frac{b_p^2}{n_1^2}\sum_{j=1}^{n}\Bigg\{\beta_x\sum_{i=1}^{p}\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}_j^{-1}\mathbf{T}^{1/2}\mathbf{e}_i\cdot\mathbf{e}_i'\mathbf{T}^{1/2}\mathbf{D}_j^{-1}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}^{1/2}\mathbf{e}_i \\ &\qquad -(\kappa-1)\mathrm{E}\mathrm{tr}\mathbf{D}_j^{-1}\mathbf{T}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{D}_j^{-1}\mathbf{T}\Bigg\} + o(1) \\ &= y_1\beta_x\underline{s}^2\int\frac{dF_{y_2}(x)}{x+\underline{s}(z)}\int\frac{x\cdot dF_{y_2}(x)}{(x+\underline{s}(z))^2} \\ &\qquad -\frac{(\kappa-1)z^2\underline{s}^2}{n_1}\Big\{\mathrm{E}\mathrm{tr}\mathbf{D}^{-1}\mathbf{T}(\underline{s}\mathbf{T}+\mathbf{I})^{-1}\mathbf{D}^{-1}\mathbf{T}\Big\} + o(1). \end{aligned}$$

The limit of the second term on the RHS of the above can be derived being the same as given in Section 9.11. Thus the additional term to the mean is as given in Lemma 9.17; that is,

$$\frac{\beta_x y_1\cdot\underline{s}^3(z)\int\frac{dF_{y_2}(x)}{x+\underline{s}(z)}\int\frac{x\cdot dF_{y_2}(x)}{(x+\underline{s}(z))^2}}{1-y_1\int\underline{s}^2(z)(x+\underline{s}(z))^{-2}dF_{y_2}(x)}.$$

The additional term to D_p is

$$\frac{\beta_x b_p(z_1)b_p(z_2)}{n_1^2}\sum_{j=1}^{n_1}\sum_{i=1}^{p}\mathbf{e}_i'\mathbf{T}^{1/2}\mathrm{E}_j\mathbf{D}_j^{-1}(z_1)\mathbf{T}^{1/2}\mathbf{e}_i\cdot\mathbf{e}_i'\mathbf{T}^{1/2}\mathrm{E}_j\mathbf{D}_j^{-1}(z_2)\mathbf{T}^{1/2}\mathbf{e}_i,$$

which, by applying Lemma 9.15, tends to

$$\beta_x y_1 \int \frac{\underline{s}(z_1) dF_{y_2}(x)}{x + \underline{s}(z_1)} \int \frac{\underline{s}(z_2) dF_{y_2}(x)}{x + \underline{s}(z_2)}.$$

Thus the additional term for $\mathrm{Cov}(M(z_1), M(z_2))$ is

$$\beta_x y_1 \cdot \int \frac{\underline{s}'(z_1) \cdot x \cdot dF_{y_2}(x)}{(x + \underline{s}(z_1))^2} \int \frac{\underline{s}'(z_2) \cdot x \cdot dF_{y_2}(x)}{(x + \underline{s}(z_2))^2}.$$

The proof of the second part is just a simple application of the first part. The proof of Lemma 9.17 is completed.

9.14.2 Proof of Theorem 9.14

Following the techniques of truncation, centralization, and normalization procedures as done in Subsection 9.7.1, we can assume the following additional conditions hold:

- There is a sequence $\eta = \eta_n \downarrow 0$ such that $|X_{jk}| \le \eta\sqrt{p}$ and $|Y_{jk}| \le \eta\sqrt{p}$.
- $EX_{jk} = EY_{jk} = 0$ and $E|X_{jk}^2| = E|Y_{jk}^2| = 1$.
- $|EX_{jk}^4| = \beta_x + \kappa + 1 + o(1)$ and $E|Y_{jk}^4| = \beta_y + \kappa + 1 + o(1)$.
- For the complex case, $EX_{jk}^2 = o(p^{-1})$ and $EY_{jk}^2 = o(p^{-1})$.

Write

$$n_1 \left[\underline{s}^{\{n_1,n_2\}}(z) - \underline{s}^{\{y_{n_1},y_{n_2}\}}(z)\right] = n_1 \left[\underline{s}^{\{n_1,n_2\}}(z) - \underline{s}^{\{y_{n_1},H_{n_2}\}}(z)\right]$$
$$+ n_1 \left[\underline{s}^{\{y_{n_1},H_{n_2}\}}(z) - \underline{s}^{\{y_{n_1},y_{n_2}\}}(z)\right],$$

where $\underline{s}^{\{y_{n_1},H_{n_2}\}}(z)$ and $\underline{s}^{\{y_{n_1},y_{n_2}\}}(z)$ are unique roots, whose imaginary parts have the same sign as that of z, to the equations

$$z = -\frac{1}{\underline{s}^{\{y_{n_1},H_{n_2}\}}} + y_{n_1} \cdot \int \frac{t \cdot dH_{n_2}(t)}{1 + t\underline{s}^{\{y_{n_1},H_{n_2}\}}}$$
$$= -\frac{1}{\underline{s}^{\{y_{n_1},H_{n_2}\}}} + y_{n_1} \cdot \int \frac{dF_{n_2}(t)}{t + \underline{s}^{\{y_{n_1},H_{n_2}\}}}$$

and

$$z = -\frac{1}{\underline{s}^{\{y_{n_1},y_{n_2}\}}} + y_{n_1} \cdot \int \frac{dF_{y_{n_2}}(t)}{t + \underline{s}^{\{y_{n_1},y_{n_2}\}}}.$$

We proceed with the proof in two steps.
Step 1. Consider the conditional distribution of

$$n_1 \left[\underline{s}^{\{n_1,n_2\}}(z) - \underline{s}^{\{y_{n_1},H_{n_2}\}}(z)\right], \tag{9.14.19}$$

given the σ-field generated by all the possible $\mathbf{S}_{n_2}$'s, which we will call $\mathscr{S}_2$. By Lemma 9.17, we have proved that the conditional distribution of

$$n_1 \left[\underline{s}^{\{n_1,n_2\}}(z) - \underline{s}^{\{y_{n_1},H_{n_2}\}}(z)\right] = p\left[s^{\{n_1,n_2\}}(z) - s^{\{y_{n_1},H_{n_2}\}}(z)\right]$$

given $\mathscr{S}_2$ converges to a Gaussian process $M_1(z)$ on the contour $\mathcal{C}$ with mean function

$$\mathrm{E}\left(M_1(z)|\mathscr{S}_2\right) = \frac{(\kappa-1)y_1 \int \underline{s}(z)^3 x[x+\underline{s}(z)]^{-3} dF_{y_{n_2}}(x)}{\left[1 - y_1 \int \underline{s}^2(z)(x+\underline{s}(z))^{-2} dF_{y_{n_2}}(x)\right]^2} + (9.14.13) \tag{9.14.20}$$

for $z \in \mathcal{C}$ and the covariance function

$$\mathrm{Cov}(M_1(z_1), M_1(z_2)|\mathscr{S}_2) = \kappa \cdot \left(\frac{\underline{s}'(z_1)\cdot\underline{s}'(z_2)}{(\underline{s}(z_1)-\underline{s}(z_2))^2} - \frac{1}{(z_1-z_2)^2}\right) + (9.14.14) \tag{9.14.21}$$

for $z_1, z_2 \in \mathcal{C}$. Note that the mean and covariance of the limiting distribution are independent of the conditioning $\mathscr{S}_2$, which shows that the limiting distribution of this part is independent of the limit of the next part because the asymptotic mean and covariances are nonrandom.

Step 2. We consider the CLT of

$$n_1[\underline{s}^{\{y_{n_1},H_{n_2}\}}(z) - \underline{s}^{\{y_{n_1},y_{n_2}\}}(z)] = p[s^{\{y_{n_1},H_{n_2}\}}(z) - s^{\{y_{n_1},y_{n_2}\}}(z)]. \tag{9.14.22}$$

By (9.7.1), under the conditions of Theorem 9.14, we have the equation

$$\begin{aligned} z &= -\frac{1}{\underline{s}^{\{y_{n_1},y_{n_2}\}}} + y_{n_1}\cdot \int \frac{t}{1+t\underline{s}^{\{y_{n_1},y_{n_2}\}}} dH_{y_{n_2}}(t) \\ &= -\frac{1}{\underline{s}^{\{y_{n_1},y_{n_2}\}}} + y_{n_1}\cdot \int \frac{dF_{y_{n_2}}(t)}{t+\underline{s}^{\{y_{n_1},y_{n_2}\}}}. \end{aligned}$$

On the other hand, $\underline{s}^{y_{n_1},H_{n_2}}$ is the solution to the equation

$$\begin{aligned} z &= -\frac{1}{\underline{s}^{\{y_{n_1},H_{n_2}\}}} + y_{n_1}\cdot \int \frac{t\cdot dH_{n_2}(t)}{1+t\cdot\underline{s}^{\{y_{n_1},H_{n_2}\}}} \\ &= -\frac{1}{\underline{s}^{\{y_{n_1},H_{n_2}\}}} + y_{n_1}\cdot \int \frac{dF_{n_2}(t)}{t+\underline{s}^{\{y_{n_1},H_{n_2}\}}}. \end{aligned}$$

By the definition of the Stieltjes transform, the two equations above become

$$\begin{aligned} z &= -\frac{1}{\underline{s}^{\{y_{n_1},y_{n_2}\}}} + y_{n_1}\cdot s_{y_{n_2}}(-\underline{s}^{\{y_{n_1},y_{n_2}\}}), \\ z &= -\frac{1}{\underline{s}^{\{y_{n_1},H_{n_2}\}}} + y_{n_1}\cdot s_{n_2}(-\underline{s}^{\{y_{n_1},H_{n_2}\}}). \end{aligned}$$

Upon taking the difference of the two identities above, we obtain

$$0 = \frac{\underline{s}^{\{y_{n_1},H_{n_2}\}} - \underline{s}^{\{y_{n_1},y_{n_2}\}}}{\underline{s}^{\{y_{n_1},y_{n_2}\}} \cdot \underline{s}^{y_{n_1},H_{n_2}}} + y_{n_1}\Big[s_{n_2}(-\underline{s}^{\{y_1,H_{n_2}\}}) - s_{n_2}(-\underline{s}^{\{y_{n_1},y_{n_2}\}})$$

$$+s_{n_2}(-\underline{s}^{\{y_{n_1},y_{n_2}\}}) - s_{y_{n_2}}(-\underline{s}^{\{y_{n_1},y_{n_2}\}})\Big]$$

$$= \frac{\underline{s}^{y_{n_1},H_{n_2}} - \underline{s}^{\{y_{n_1},y_{n_2}\}}}{\underline{s}^{\{y_{n_1},y_{n_2}\}} \cdot \underline{s}^{y_{n_1},H_{n_2}}} - y_{n_1}\int \frac{(\underline{s}^{y_{n_1},H_{n_2}} - \underline{s}^{\{y_{n_1},y_{n_2}\}})dF_{n_2}(t)}{(t+\underline{s}^{y_{n_1},H_{n_2}})(t+\underline{s}^{\{y_{n_1},y_{n_2}\}})}$$

$$+y_{n_1}\left[s_{n_2}(-\underline{s}^{\{y_{n_1},y_{n_2}\}}) - s_{y_{n_2}}(-\underline{s}^{\{y_{n_1},y_{n_2}\}})\right].$$

Therefore, we have

$$n_1 \cdot \left[\underline{s}^{\{y_{n_1},H_{n_2}\}}(z) - \underline{s}^{\{y_{n_1},y_{n_2}\}}(z)\right]$$

$$= -y_{n_1} \cdot \underline{s}^{\{y_{n_1},y_{n_2}\}} \cdot \ \underline{s}^{\{y_{n_1},H_{n_2}\}} \cdot \frac{n_1\left[s_{n_2}(-\underline{s}^{\{y_{n_1},y_{n_2}\}}) - s_{y_{n_2}}(-\underline{s}^{\{y_{n_1},y_{n_2}\}})\right]}{1 - y_{n_1} \cdot \int \frac{\underline{s}^{\{y_{n_1},y_{n_2}\}} \cdot \underline{s}^{\{y_{n_1},H_{n_2}\}} dF_{n_2}(t)}{\left(t+\underline{s}^{\{y_{n_1},y_{n_2}\}}\right)\cdot\left(t+\underline{s}^{\{y_{n_1},H_{n_2}\}}\right)}}$$

$$= -\underline{s}^{\{y_{n_1},y_{n_2}\}} \cdot \underline{s}^{y_{n_1},H_{n_2}} \cdot \frac{n_2\left[\underline{s}_{n_2}(-\underline{s}^{\{y_{n_1},y_{n_2}\}}) - \underline{s}_{y_{n_2}}(-\underline{s}^{\{y_{n_1},y_{n_2}\}})\right]}{1 - y_{n_1} \cdot \int \frac{\underline{s}^{\{y_{n_1},y_{n_2}\}} \cdot \ \underline{s}^{\{y_{n_1},H_{n_2}\}} dF_{n_2}(t)}{\left(t+\underline{s}^{\{y_{n_1},y_{n_2}\}}\right)\ \left(t+\underline{s}^{\{y_{n_1},H_{n_2}\}}\right)}}.$$

Consider the CLT for

$$n_2 \cdot \left[\underline{s}_{n_2}\left(-\underline{s}^{\{y_{n_1},y_{n_2}\}}(z)\right) - \underline{s}_{y_{n_2}}\left(-\underline{s}^{\{y_{n_1},y_{n_2}\}}(z)\right)\right].$$

Because, for any $z \in \mathbb{C}^+$, $\underline{s}^{\{y_{n_1},y_{n_2}\}}(z) \to \underline{s}(z)$, to consider the limiting distribution of

$$n_2 \cdot \left[\underline{s}_{n_2}\left(-\underline{s}^{\{y_{n_1},y_{n_2}\}}(z)\right) - \underline{s}_{y_{n_2}}\left(-\underline{s}^{\{y_{n_1},y_{n_2}\}}(z)\right)\right],$$

one only needs to consider the CLT for

$$n_2 \cdot \left[\underline{s}_{n_2}\left(-\underline{s}(z)\right) - \underline{s}_{y_{n_2}}\left(-\underline{s}(z)\right)\right],$$

it can be shown that when z runs along $\mathcal{C}$ clockwise, $-\underline{s}(z)$ will enclose the support of F_{y_2} clockwise without intersecting the support. Then, by Lemma 9.17 and Lemma 9.11 (with minor modification), we have $n_2[\underline{s}_{n_2}(-\underline{s}(z)) - \underline{s}_{y_{n_2}}(-\underline{s}(z))]$ converging weakly to a Gaussian process $M_2(\cdot)$ on $z \in \mathcal{C}$ with mean function

$$\mathrm{E}(M_2(z)) = (\kappa - 1) \cdot \frac{y_2 \cdot [\underline{s}_{y_2}(-\underline{s}(z))]^3 \cdot [1 + \underline{s}_{y_2}(-\underline{s}(z))]^{-3}}{\left[1 - y_2 \cdot \left(\frac{\underline{s}_{y_2}(-\underline{s}(z))}{1+\underline{s}_{y_2}(-\underline{s}(z))}\right)^2\right]^2} + (9.14.15) \tag{9.14.23}$$

and $\mathrm{Cov}(M_2(z_1), M_2(z_2)) =$

$$\kappa \cdot \left(\frac{\underline{s}'_{y_2}(-\underline{s}(z_1)) \cdot \underline{s}'_{y_2}(-\underline{s}(z_2))}{[\underline{s}_{y_2}(-\underline{s}(z_1)) - \underline{s}_{y_2}(-\underline{s}(z_2))]^2} - \frac{1}{(\underline{s}(z_1) - \underline{s}(z_2))^2} \right) + (9.14.16). \quad (9.14.24)$$

for $z_1, z_2 \in \mathcal{C}$. Because $\dfrac{-\underline{s}^{\{y_{n_1}, y_{n_2}\}}(z) \cdot \underline{s}^{\{y_{n_1}, H_{n_2}\}}(z)}{1 - y_{n_1} \cdot \int \frac{\underline{s}^{\{y_{n_1}, y_{n_2}\}}(z) \cdot \underline{s}^{\{y_{n_1}, H_{n_2}\}} dF_{n_2}(t)}{\left(t + \underline{s}^{\{y_{n_1}, y_{n_2}\}}(z)\right) \left(t + \underline{s}^{\{y_{n_1}, H_{n_2}\}}\right)}}$ converges to $-\underline{s}'(z) = \dfrac{-\underline{s}^2(z)}{1 - y_1 \cdot \underline{s}^2(z) \cdot \int \frac{dF_{y_2}(t)}{[t + \underline{s}(z)]^2}}$, we then conclude that

$$n_1 \cdot \left[\underline{s}^{\{y_{n_1}, H_{n_2}\}}(z) - \underline{s}^{\{y_{n_1}, y_{n_2}\}}(z) \right]$$

converges weakly to a Gaussian process $M_3(\cdot)$ satisfying

$$M_3(z) = -\underline{s}'(z) \cdot M_2(z),$$

with the means $\mathrm{E}(M_3(z)) = -\underline{s}'(z) \cdot \mathrm{E}M_2(z)$ and covariance functions $\mathrm{Cov}(M_3(z_1), M_3(z_2)) = \underline{s}'(z_1)\underline{s}'(z_2) \cdot \mathrm{Cov}(M_2(z_1), M_2(z_2))$. Because the limit of

$$n_1 \cdot \left[\underline{s}^{\{n_1, n_2\}}(z) - \underline{s}^{\{y_{n_1}, H_{n_2}\}}(z) \right]$$

conditioned on $\mathscr{S}_2$ is independent of the ESD of S_{n_2}, we know that the limits of

$$n_1 \cdot \left[\underline{s}^{\{n_1, n_2\}}(z) - \underline{s}^{\{y_{n_1}, H_{n_2}\}}(z) \right] \quad \text{and} \quad n_1 \cdot \left[\underline{s}^{\{y_{n_1}, H_{n_2}\}}(z) - \underline{s}^{\{y_{n_1}, y_{n_2}\}}(z) \right]$$

are asymptotically independent. Thus we have $n_1 \cdot \left[\underline{s}^{\{n_1, n_2\}}(z) - \underline{s}^{\{y_{n_1}, y_{n_2}\}}(z) \right]$ converging weakly to a Gaussian process $M_1(z) + M_3(z)$, where $M_1(z)$ and $M_3(z)$ are independent. Thus, the mean function will be

$$\mathrm{E}(M_1(z) + M_3(z)) = (9.14.25) + (9.14.26) + (9.14.27) + (9.14.28),$$

where

$$(\kappa - 1) \cdot \frac{y_1 \int \underline{s}^3(z) x [x + \underline{s}(z)]^{-3} dF_{y_2}(x)}{\left[1 - y_1 \int \underline{s}^2(z)(x + \underline{s}(z))^{-2} dF_{y_2}(x)\right]^2} \quad (9.14.25)$$

$$+\beta_x \cdot \frac{y_1 \cdot \underline{s}^3(z) \cdot \int \frac{dF_{y_2}(x)}{x + \underline{s}(z)} \int \frac{x \cdot dF_{y_2}(x)}{(x + \underline{s}(z))^2}}{1 - y_1 \int \underline{s}^2(z)(x + \underline{s}(z))^{-2} dF_{y_2}(x)} \quad (9.14.26)$$

$$-(\kappa - 1) \cdot \underline{s}'(z) \frac{y_2 \cdot [\underline{s}_{y_2}(-\underline{s}(z))]^3 \cdot [1 + \underline{s}_{y_2}(-\underline{s}(z))]^{-3}}{\left[1 - y_2 \cdot \left(\frac{\underline{s}_{y_2}(-\underline{s}(z))}{1 + \underline{s}_{y_2}(-\underline{s}(z))} \right)^2 \right]^2} \quad (9.14.27)$$

$$-\beta_y \cdot \underline{s}'(z)\frac{y_2 \cdot \underline{s}_{y_2}^3(-\underline{s}(z)) \cdot (1+\underline{s}_{y_2}(-\underline{s}(z)))^{-3}}{1-y_2 \cdot \underline{s}_{y_2}^2(-\underline{s}(z)) \cdot (1+\underline{s}_{y_2}(-\underline{s}(z)))^{-2}} \tag{9.14.28}$$

and

$$\begin{aligned}&\mathrm{Cov}(M_1(z_1)+M_3(z_1), M_1(z_2)+M_3(z_2))\\&=(9.14.29)+(9.14.30)+(9.14.31)-\frac{\kappa}{(z_1-z_2)^2},\end{aligned}$$

where

$$\beta_x \cdot y_1 \cdot \int \frac{\underline{s}'(z_1) \cdot x \cdot dF_{y_2}(x)}{(x+\underline{s}(z_1))^2} \int \frac{\underline{s}'(z_2) \cdot x \cdot dF_{y_2}(x)}{(x+\underline{s}(z_2))^2} \tag{9.14.29}$$

$$\kappa \cdot \frac{\underline{s}'(z_1)\underline{s}'(z_2)\underline{s}'_{y_2}(-\underline{s}(z_1)) \cdot \underline{s}'_{y_2}(-\underline{s}(z_2))}{[\underline{s}_{y_2}(-\underline{s}(z_1))-\underline{s}_{y_2}(-\underline{s}(z_2))]^2} \tag{9.14.30}$$

$$\beta_y \cdot y_2 \cdot \frac{\underline{s}'(z_1)\underline{s}'_{y_2}(-\underline{s}(z_1))}{(1+\underline{s}_{y_2}(-\underline{s}(z_1)))^2} \cdot \frac{\underline{s}'(z_2)\underline{s}'_{y_2}(-\underline{s}(z_2))}{(1+\underline{s}_{y_2}(-\underline{s}(z_2)))^2}. \tag{9.14.31}$$

If we choose the contour $\mathcal{C}$ enclosing the interval $[a, b]$, where $a, b = \frac{(1\mp h)^2}{(1-y_2)^2}$, we show that $\int f(x)dG(x) = -\frac{1}{2\pi i}\int_{\mathcal{C}} f(z)s_G(z)dz$ with probability 1 for all large n. In fact, if $y_{n_1} < 1$, then by the exact spectrum separation theorem, with probability 1, for all large p, all eigenvalues of the F-matrix fall within the contour $\mathcal{C}$ and hence the equality above is true. When $y_1 > 1$, by the exact spectrum separation, the F-matrix has exactly $p-n$ zero eigenvalues and all the n positive eigenvalues fall in the contour $\mathcal{C}$. The equality of $\int f(x)dG(x) = -\frac{1}{2\pi i}\int_{\mathcal{C}} f(z)s_G(z)dz$ remains true for all large p.

Then, we obtain that the CLT for the LSS of the F-matrix

$$\left(\int f_1(x)\widetilde{G}_{n_1,n_2}(x), \cdots, \int f_k(x)d\widetilde{G}_{n_1,n_2}(x)\right)$$

converges weakly to a Gaussian vector $(X_{f_1}, \cdots, X_{f_k})$, where

$$\mathrm{E}X_{f_i} = -\frac{1}{2\pi i}\oint f_i(z)E(M_1(z)+M_3(z))dz$$

and

$$\begin{aligned}&\mathrm{Cov}(X_{f_i}, X_{f_j})\\&=-\frac{1}{4\pi^2}\oint\oint f_i(z)f_j(z)\mathrm{Cov}(M_1(z_1)+M_3(z_1), M_1(z_2)+M_3(z_2))dz_1dz_2.\end{aligned}$$

Recall that $s_0(z) = \underline{s}_{y_2}(-\underline{s}(z))$. Then, by Lemma 9.16, we have

$$-\frac{\oint f_i(z) \cdot (9.14.25)dz}{2\pi i} = \frac{\kappa-1}{4\pi i}\oint f_i(z)d\log\left(\frac{(1-y_2)s_0^2(z)+2s_0(z)+1-y_1}{(1-y_2)s_0^2(z)+2s_0(z)+1}\right)$$

$$-\frac{\oint f_i(z)\cdot(9.14.26)dz}{2\pi i} = \frac{\beta_x\cdot y_1}{2\pi i}\oint\frac{f_i(z)}{(s_0(z)+1)^3}ds_0(z)$$

$$-\frac{\oint f_i(z)\cdot(9.14.27)dz}{2\pi i} = \frac{\kappa-1}{4\pi i}\oint f_i(z)\ \ d\log\left(1-\frac{y_2\cdot s_0^2(z)}{(1+s_0(z))^2}\right)$$

$$-\frac{\oint f_i(z)\cdot(9.14.28)dz}{2\pi i}$$
$$= \frac{\beta_y}{4\pi i}\oint f_i(z)\left(1-\frac{y_2\cdot s_0^2(z)}{(1+s_0(z))^2}\right)\ \ d\log\left(1-\frac{y_2\cdot s_0^2(z)}{(1+s_0(z))^2}\right)$$

and

$$-\frac{\oint\oint f_i(z_1)f_j(z_2)\cdot(9.14.29)dz}{4\pi^2} = -\frac{\beta_x\cdot y_1}{4\pi^2}\oint\oint\frac{f_i(z_1)f_j(z_2)ds_0(z_1)ds_0(z_2)}{(m_0(z_1)+1)^2(s_0(z_2)+1)^2}$$

$$-\frac{\oint\oint f_i(z_1)f_j(z_2)\cdot(9.14.30)dz}{4\pi^2} = -\frac{\kappa}{4\pi^2}\oint\oint\frac{f_i(z_1)f_j(z_2)ds_0(z_1)ds_0(z_2)}{(s_0(z_1)-s_0(z_2))^2}$$

$$-\frac{\oint\oint f_i(z_1)f_j(z_2)\cdot(9.14.31)dz}{4\pi^2} = -\frac{\beta_y\cdot y_2}{4\pi^2}\oint\oint\frac{f_i(z_1)f_j(z_2)ds_0(z_1)ds_0(z_2)}{(s_0(z_1)+1)^2(s_0(z_2)+1)^2}.$$

The support set of the limiting spectral distribution $F_{y_1,y_2}(x)$ of the F-matrix is

$$\left[a=\frac{(1-h)^2}{(1-y_2)^2},\ b=\frac{(1+h)^2}{(1-y_2)^2}\right], \tag{9.14.32}$$

when $y_1\le 1$ or the interval above with a singleton $\{0\}$ when $y_1>1$. Because $-\underline{s}(a)$ and $-\underline{s}(b)$ are real numbers outside the support set $[(1-\sqrt{y_2})^2,\ (1+\sqrt{y_2})^2]$ of $F_{y_2}(x)$, then by (9.14.11) we know that $\underline{s}_{y_2}(-\underline{s}(a))$ and $\underline{s}_{y_2}(-\underline{s}(b))$ are real numbers that are the real roots of equations

$$a=\frac{\underline{s}_{y_2}(-\underline{s}(a))\cdot[\underline{s}_{y_2}(-\underline{s}(a))+1-y_1]}{\left[\underline{s}_{y_2}(-\underline{s}(a))-\frac{1}{y_2-1}\right]\cdot(y_2-1)}$$

and

$$b=\frac{\underline{s}_{y_2}(-\underline{s}(b))\cdot[\underline{s}_{y_2}(-\underline{s}(b))+1-y_1]}{\left[\underline{s}_{y_2}(-\underline{s}(b))-\frac{1}{y_2-1}\right]\cdot(y_2-1)}.$$

So we obtain $\underline{s}_{y_2}(-\underline{s}(b))=-\frac{1+h}{1-y_2}$ and $\underline{s}_{y_2}(-\underline{s}(a))=-\frac{1-h}{1-y_2}$. Clearly, when z runs in the positive direction around the support interval $[a,b]$ of $F^{\{y_1,y_2\}}(x)$, $\underline{s}_{y_2}(-\underline{s}(z))$ runs in the positive direction around the interval

$$I=\left(-\frac{1+h}{1-y_2},\quad -\frac{1-h}{1-y_2}\right).$$

Let $s_0(z) = -\frac{1+hr\xi}{1-y_2}$, where $r > 1$ but very close to 1, $|\xi| = 1$. Then, by Lemma 9.16, we have

$$z = -\frac{s_0(z)(s_0(z)+1-y_1)}{(1-y_2)(s_0(z)+\frac{1}{1-y_2})}.$$

This shows that when ξ runs a cycle along the unit circle anticlockwise, z runs a cycle anticlockwise and the cycle encloses the interval $[a,b]$, where $a,b = \frac{(1\mp h)^2}{(1-y_2)^2}$. Therefore, one can make $r \downarrow 1$ as finding their values, so we have

$$-\frac{1}{2\pi i}\oint f_i(z)\cdot(9.14.25)dz = \frac{\kappa-1}{4\pi i}\lim_{r\downarrow 1}\oint_{|\xi|=1} f_i\left(\frac{|1+h\xi|^2}{(1-y_2)^2}\right) \times\left[\frac{1}{\xi-r^{-1}}+\frac{1}{\xi+r^{-1}}-\frac{1}{\xi-\frac{\sqrt{y_2}}{h}}-\frac{1}{\xi+\frac{\sqrt{y_2}}{h}}\right]d\xi \tag{9.14.33}$$

$$-\frac{1}{2\pi i}\oint f_i(z)\cdot(9.14.26)dz = \frac{\beta_x\cdot y_1\cdot(1-y_2)^2}{2\pi i\cdot h^2} \times\oint_{|\xi|=1} f_i\left(\frac{|1+h\xi|^2}{(1-y_2)^2}\right)\frac{1}{(\xi+\frac{y_2}{h})^3}\,d\xi \tag{9.14.34}$$

$$-\frac{1}{2\pi i}\oint f_i(z)\cdot(9.14.27)dz = \frac{\kappa-1}{4\pi i}\oint_{|\xi|=1} f_i \times\left(\frac{|1+h\xi|^2}{(1-y_2)^2}\right)\left[\frac{1}{\xi-\frac{\sqrt{y_2}}{hr}}+\frac{1}{\xi+\frac{\sqrt{y_2}}{h}}-\frac{2}{\xi+\frac{y_2}{h}}\right]d\xi \tag{9.14.35}$$

$$-\frac{1}{2\pi i}\oint f_i(z)\cdot(9.14.28)dz = \frac{\beta_y\cdot(1-y_2)}{4\pi i}\oint_{|\xi|=1} f_i\left(\frac{|1+h\xi|^2}{(1-y_2)^2}\right) \times\frac{\xi^2-\frac{y_2}{h^2r^2}}{(\xi+\frac{y_2}{hr})^2}\left[\frac{1}{\xi-\frac{\sqrt{y_2}}{h}}+\frac{1}{\xi+\frac{\sqrt{y_2}}{h}}-\frac{2}{\xi+\frac{y_2}{h}}\right]d\xi. \tag{9.14.36}$$

By Lemma 9.16, we have

$$\underline{s}'(z) = -\frac{(1-y_2)s_0^2(z)+2s_0+1}{s_0^2(z)\cdot(s_0(z)+1)^2}\cdot s_0'(z) = \frac{(1-y_2)^2}{hr}\cdot\frac{\left(\xi+\frac{\sqrt{y_2}}{hr}\right)\left(\xi-\frac{\sqrt{y_2}}{hr}\right)\xi'}{\left(\xi+\frac{y_2}{hr}\right)^2\left(\xi+\frac{1}{hr}\right)^2}$$

and

$$\underline{s}(z) = -\frac{(1-y_2)^2}{hr}\frac{\xi}{(\xi+\frac{1}{hr})(\xi+\frac{y_2}{hr})}.$$

Making the variable change gives $s_0(z_j) = -\frac{1+hr_j\xi_j}{1-y_2}$, where $r_2 > r_1 > 1$. Because the quantities are independent of $r_2 > r_1 > 1$ provided they are small enough, one can make $r_2 \downarrow 1$ when finding their values. That is,

$$-\frac{1}{4\pi^2}\oint\oint f_i(z_1)f_j(z_2)\cdot(9.14.29)dz_1dz_2 =$$

$$-\frac{\beta_x\cdot y_1(1-y_2)^2}{4\pi^2\cdot h^2}\oint_{|\xi_1|=1}\frac{f_i\left(\frac{|1+h\xi_1|^2}{(1-y_2)^2}\right)}{(\xi_1+\frac{y_2}{h})^2}d\xi_1\oint_{|\xi_2|=1}\frac{f_j\left(\frac{|1+h\xi_2|^2}{(1-y_2)^2}\right)}{(\xi_2+\frac{y_2}{h})^2}d\xi_2 \quad (9.14.37)$$

$$-\frac{1}{4\pi^2}\oint\oint f_i(z_1)f_j(z_2)\cdot(9.14.30)dz_1dz_2 =$$

$$-\frac{\kappa}{4\pi^2}\lim_{r\downarrow 1}\oint_{|\xi_1|=1}\oint_{|\xi_2|=1}\frac{f_i\left(\frac{|1+h\xi_1|^2}{(1-y_2)^2}\right)f_j\left(\frac{|1+h\xi_2|^2}{(1-y_2)^2}\right)}{(\xi_1-r\xi_2)^2}\,d\xi_1d\xi_2 \qquad (9.14.38)$$

$$-\frac{1}{4\pi^2}\oint\oint f_i(z_1)f_j(z_2)\cdot(9.14.31)dz_1dz_2 = -\frac{\beta_y\cdot y_2(1-y_2)^2}{4\pi^2\cdot h^2}$$

$$\oint_{|\xi_1|=1}\frac{f_i\left(\frac{|1+h\xi_1|^2}{(1-y_2)^2}\right)}{(\xi_1+\frac{y_2}{h})^2}d\xi_1\oint_{|\xi_2|=1}\frac{f_j\left(\frac{|1+h\xi_2|^2}{(1-y_2)^2}\right)}{(\xi_2+\frac{y_2}{h})^2}d\xi_2. \qquad (9.14.39)$$

This finishes the proof of Theorem 9.14.

9.15 CLT for the LSS of a Large Dimensional Beta-Matrix

As a consequence of the result on the F-matrix, we establish a CLT for the LSS of the beta-matrix $\boldsymbol{\beta}_{\{n_1,n_2\}} = \mathbf{S}_2(\mathbf{S}_2 + d\cdot\mathbf{S}_1)^{-1}$, a matrix function of the F-matrix, where d is a positive number. If λ is an eigenvalue of the beta-matrix $\boldsymbol{\beta}_{\{n_1,n_2\}}$, then $\frac{1}{d}\left(\frac{1}{\lambda}-1\right)$ is an eigenvalue of the F-matrix $\mathbf{S}_1\mathbf{S}_2^{-1}$, therefore, the ESD of the beta-matrix is

$$F_\beta^{\{n_1,n_2\}}(x) = 1 - F^{\{n_1,n_2\}}\left(\frac{1}{d}\left(\frac{1}{x}-1\right)_-\right),\ x>0,$$

where $F^{\{n_1,n_2\}}(x_-)$ is the left-limit at x; that is $1-F^{\{n_1,n_2\}}\left(\frac{1}{d}\left(\frac{1}{x}-1\right)\right)-\frac{1}{p}$ if x is an eigenvalue of $\boldsymbol{\beta}_{\{n_1,n_2\}}$. Similarly, we obtain

$$F_\beta^{\{y_{n_1},y_{n_2}\}}(x) = 1 - F^{\{y_{n_1},y_{n_2}\}}\left(\frac{1}{d}\left(\frac{1}{x}-1\right)_-\right).$$

Then, we have the following lemma.

Lemma 9.18. *For the beta-matrix* $\mathbf{S}_2\left(\mathbf{S}_2+d\cdot\mathbf{S}_1\right)^{-1}$*, where* d *is a positive number, we have*

$$\left(\int f_1(x)d\widehat{G}_{n_1,n_2}(x),\cdots,\int f_k(x)d\widehat{G}_{n_1,n_2}(x)\right)$$

$$= -\left(\int f_1\left(\tfrac{1}{dx+1}\right)d\widetilde{G}_{n_1,n_2}(x),\cdots,\int f_k\left(\tfrac{1}{dx+1}\right)d\widetilde{G}_{n_1,n_2}(x)\right),$$

where $\widehat{G}_{n_1,n_2}(x) = p\left(F_\beta^{\{n_1,n_2\}}(x) - F_\beta^{\{y_{n_1},y_{n_2}\}}(x)\right)$ *and* $F_\beta^{\{y_1,y_2\}}(x)$ *is the LSD of the beta-matrix.*

As an application of the F-matrix, the following theorem establishes a CLT for the LSS of a beta-matrix that is a matrix function of the F-matrix and is useful in large dimensional data analysis.

Theorem 9.19. *The LSS of the beta-matrix* $\mathbf{S}_2(\mathbf{S}_2 + d\mathbf{S}_1)^{-1}$, *where* d *is a positive number, is*

$$\left(\int f_1(x)d\widehat{G}_{n_1,n_2}(x),\cdots,\int f_k(x)d\widehat{G}_{n_1,n_2}(x)\right). \tag{9.15.1}$$

Under the conditions in (a), (b) and (i), (ii), (9.15.1) converges weakly to a Gaussian vector $(X_{f_1},\cdots,X_{f_k})$ *whose means and covariances are the same as in Theorem 9.14 except* $f_i(x)$ *and* $f_j(x)$ *are replaced by* $-f_i\left(\frac{1}{d\cdot x+1}\right)$ *and* $-f_j\left(\frac{1}{d\cdot x+1}\right)$, *respectively.*

9.16 Some Examples

Here, we give asymptotic means and variance-covariances of some often-used LSSs of the F-matrix when $\mathbf{S}_1$ and $\mathbf{S}_2$ are real variables. These results can be used directly for many problems in multivariate statistical analysis.

Example 9.20. If $f = \log(a+bx)$, $f' = \log(a'+b'x)$, and $a, a', b, b' > 0$, then

$$EX_f = \frac{1}{2}\log\left(\frac{(c^2-d^2)h^2}{(ch-y_2d)^2}\right)$$

and

$$\mathrm{Cov}(X_f, X_{f'}) = 2\log\left(\frac{cc'}{(cc'-dd')}\right),$$

where $c > d > 0$, $c' > d' > 0$ satisfying $c^2 + d^2 = a(1-y_2)^2 + b(1+h^2)$, $c'^2 + d'^2 = a'(1-y_2)^2 + b'(1+h^2)$, $cd = bh$, and $c'd' = b'h$.

Proof. In fact, we have

$$\mathrm{E}(X_f) = \lim_{r\downarrow 1}\frac{1}{4\pi i}\oint_{|\xi|=1}\log(|c+d\xi|^2)\left(\frac{1}{r\xi+1}+\frac{1}{r\xi-1}-\frac{2}{\xi+h^{-1}y_2}\right)d\xi$$

$$
\begin{aligned}
&= \lim_{r\downarrow 1} \frac{1}{4\pi i} \oint_{|\xi|=1} \log(|c + d\xi^{-1}|^2) \bigg(\frac{1}{r\xi^{-1}+1} + \frac{1}{r\xi^{-1}-1} \\
&\quad - \frac{2}{\xi^{-1} + h^{-1}y_2} \bigg) \xi^{-2} d\xi \\
&= \lim_{r\downarrow 1} \bigg\{ \frac{1}{8\pi i} \oint_{|\xi|=1} \log(|c + d\xi|^2) \bigg(\frac{1}{r\xi+1} + \frac{1}{r\xi-1} - \frac{2}{\xi + h^{-1}y_2} \\
&\quad + \frac{1}{\xi(r+\xi)} + \frac{1}{\xi(r-\xi)} - \frac{2}{\xi(1 + h^{-1}y_2\xi)} \bigg) d\xi \bigg\} \\
&= \lim_{r\downarrow 1} \Re \bigg\{ \frac{1}{8\pi i} \int_{|\xi|=1} \log((c + d\xi)^2) \bigg(\frac{1}{r\xi+1} + \frac{1}{r\xi-1} - \frac{2}{\xi + h^{-1}y_2} \\
&\quad + \frac{1}{\xi(r+\xi)} + \frac{1}{\xi(r-\xi)} - \frac{2}{\xi(1 + h^{-1}y_2\xi)} \bigg) d\xi \bigg\} \\
&= \frac{1}{4} \bigg(\log[(c^2 - d^2)^2] - 2\log[(c - y_2 d h^{-1})^2] \bigg) = \frac{1}{2} \log \left(\frac{(c^2 - d^2)h^2}{(ch - y_2 d)^2} \right).
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
&\mathrm{Cov}(X_f, X_{f'}) \\
&= -\lim_{r\downarrow 1} \frac{1}{2\pi^2} \oint \oint_{|\xi_1|=|\xi_2|=1} \log(|c + d\xi_1|^2) \log(|c' + d'\xi_2|^2) \frac{d\xi_1 d\xi_2}{(\xi_1 - r\xi_2)^2} \\
&= -\lim_{r\downarrow 1} \frac{1}{4\pi^2} \oint_{|\xi_1|=1} \log(|c + d\xi_1|^2) d\xi_1 \bigg[\oint_{|\xi_2|=1} \\
&\quad \Big(\log((c' + d'\xi_2)^2) + \log((c' + d'\bar{\xi}_2)^2) \Big) \frac{d\xi_2}{(\xi_1 - r\xi_2)^2} \bigg] \\
&\quad \left(\text{since } \log(|c' + d'\xi_2|^2) = \frac{1}{2} \Big[\log((c' + d'\xi_2)^2) + \log((c' + d'\bar{\xi}_2)^2) \Big] \right) \\
&= -\lim_{r\downarrow 1} \frac{1}{4\pi^2} \oint_{|\xi_1|=1} \log(|c + d\xi_1|^2) d\xi_1 \bigg[\oint_{|\xi_2|=1} \log((c' + d'\xi_2)^2 \Big) \\
&\quad \left(\frac{1}{(\xi_1 - r\xi_2)^2} + \frac{1}{(\xi_1\xi_2 - r)^2} \right) d\xi_2 \bigg] \text{ (transforming } \xi_2^{-1} = \bar{\xi}_2 \to \xi_2) \\
&= \lim_{r\downarrow 1} \frac{d'}{\pi i} \oint_{|\xi_1|=1} \frac{1}{c'r^2 + d'r\xi_1} \log(|c + d\xi_1|^2) d\xi_1 \text{(second term is analytic)} \\
&= \frac{d'}{2\pi i} \oint_{|\xi_1|=1} \frac{1}{c' + d'\xi_1} \Big(\log((c + d\xi_1)^2) + \log((c + d\bar{\xi_1})^2) \Big) d\xi_1 \\
&= \frac{d'}{2\pi i} \oint_{|\xi_1|=1} \log((c + d\xi_1)^2) \left(\frac{1}{c' + d'\xi_1} + \frac{1}{\xi_1(c'\xi_1 + d')} \right) d\xi_1 \\
&= \log(c^2) - \log(c - dd'/c')^2 = 2\log\left(\frac{cc'}{cc' - dd'} \right).
\end{aligned}
$$

The proof is complete.

Example 9.21. For any positive integers $k \geq r \geq 1$ and $f(x) = x^r$ and $g(x) = x^k$, we have

$$\begin{aligned}
&E(X_f) \\
&= \frac{1}{4\pi i} \lim_{l \downarrow 1} \oint_{|\xi|=1} \frac{(1+h\xi)^r(1+h\xi^{-1})^r}{(1-y_2)^{2r}} \left(\frac{1}{\xi + l^{-1}} + \frac{1}{\xi - l^{-1}} - \frac{2}{\xi + \frac{y_2}{hl}} \right) d\xi \\
&= \frac{1}{2(1-y_2)^{2r}} \Bigg[(1-h)^{2r} + (1+h)^{2r} - 2(1-y_2)^r \left(1 - \frac{h^2}{y_2}\right)^r \\
&- \sum_{\substack{i \leq r,\ j \geq 0,\ k \geq 0 \\ i-j=2k+1}} \binom{r}{j}\binom{r}{i} h^{j+i} + 2 \sum_{\substack{i \leq r,\ j \geq 0,\ k \geq 0 \\ i-j=k+1}} \binom{r}{j}\binom{r}{i} h^{j+i} \left(-\frac{h}{y_2}\right)^{k-1} \Bigg]
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Cov}(X_f, X_g) \\
&= -\frac{1}{2\pi^2} \lim_{l \downarrow 1} \oint\oint_{|\xi_1|=|\xi_2|=1} \frac{|1+h\xi_1|^{2r} \cdot |1+h\xi_2|^{2k}}{(1-y_2)^{2r+2k} \cdot (\xi_1 - l\xi_2)^2} d\xi_1 d\xi_2 \\
&= \frac{2 \cdot r! \cdot k!}{(y_2-1)^{2r+2k}} \sum_{j=0}^{r-1} (j+1) \cdot \\
&\quad \left\{ \left(\sum_{l_3=j+1}^{\left[\frac{1+j+r}{2}\right]} \frac{(y_1+y_2-y_1y_2+1)^{r-2l_3} \cdot (y_1+y_2-y_1y_2)^{l_3}}{(-1-j+l_3)! \cdot (1+j+r-2l_3)! \cdot l_3!} \right) \right. \\
&\quad \left. \times \left(\sum_{l_3'=0}^{\left[\frac{k-j-1}{2}\right]} \frac{(y_1+y_2-y_1y_2+1)^{k-2l_3'} \cdot (y_1+y_2-y_1y_2)^{l_3'}}{(j+1+l_3')! \cdot (k-j-1-2l_3')! \cdot l_3'!} \right) \right\},
\end{aligned}$$

where $[a]$ is the integer part of a; that is, the maximum integer less than or equal to a.

Example 9.22. If $f = e^x$, then

$$\begin{aligned}
E(X_f) = \frac{1}{2} \Bigg[& e^{\frac{(1-h)^2}{(1-y_2)^2}} + e^{\frac{(1+h)^2}{(1-y_2)^2}} - 2e^{\frac{\left(1-\frac{h^2}{y_2}\right)}{(1-y_2)^2}} \\
& - e^{\frac{2}{(1-y_2)^2}} \sum_{\substack{j,\ k,\ l \geq 0 \\ j-k=2l+1}} \frac{h^{j+k}}{j!k!(1-y_2)^{2j+2k}} \\
& - 2e^{\frac{2}{(1-y_2)^2}} \sum_{\substack{j,\ k,\ l \geq 0 \\ j-k=2l+1}} \frac{h^{j+k}}{j!k!(1-y_2)^{2j+2k}} \left(-\frac{h}{y_2}\right)^{l-1} \Bigg].
\end{aligned}$$

$$\operatorname{Var}(X_f) = -\frac{1}{2\pi^2}\lim_{l\downarrow 1}\oint_{|\xi_1|=1}\oint_{|\xi_2|=1}\frac{e^{\frac{|1+h\xi_1|^2}{(1-y_2)^2}}\cdot e^{\frac{|1+h\xi_2|^2}{(1-y_2)^2}}}{(\xi_1-l\xi_2)^2}d\xi_1 d\xi_2$$

$$= \sum_{j,\ k=1}^{+\infty}\frac{1}{j!k!}\left[\lim_{1\downarrow 1}\left(-\frac{1}{2\pi^2}\oint_{|\xi_1|=1}\oint_{|\xi_2|=1}\frac{|1+h\xi_1|^{2j}\cdot|1+h\xi_2|^{2k}}{(1-y_2)^{2j+2k}\cdot(\xi_1-r\xi_2)^2}d\xi_1 d\xi_2\right)\right].$$

Chapter 10
Eigenvectors of Sample Covariance Matrices

Thus far, all results in this book have been concerned with the limiting behavior of eigenvalues of large dimensional random matrices. As mentioned in the introduction, the development of RMT has been attributed to the investigation of the energy level distribution of a large number of particles in QM; in other words, the original interests of RMT were confined to eigenvalue distributions of large dimensional random matrices. In the beginning, most of the important results in RMT were related to a certain deterministic behavior, to the extent that their empirical distributions tend toward nonrandom ones as the dimension tends to infinity. Moreover, this behavior is invariant under the distribution of the variables making up the matrix.

Along with the rapid development and wide application of modern computer techniques in various disciplines, large dimensional data analysis has sprung up, resulting in wide application of the theory of spectral analysis of large dimensional random matrices to various areas, such as statistics, signal processing, finance, and economics. Stemming from practical applications, RMT has deepened its interest toward the investigation of second-order accuracy of the ESD, as introduced in the previous chapter. Meanwhile, practical applications of RMT have also raised the need to understand the limiting behavior of eigenvectors of large dimensional random matrices. For example, in PCA (principal component analysis), the eigenvectors corresponding to a few of the largest eigenvalues of random matrices (that is, the directions of the principal components) are of special interest. Therefore, the limiting behavior of eigenvectors of large dimensional random matrices becomes an important issue in RMT.

However, the investigation on eigenvectors has been relatively weaker than that on eigenvalues in the literature due to the difficulty of mathematical formulation since the dimension increases with the sample size. In the literature there were found only five papers, by Silverstein, up to 1990, concerning real sample covariance matrices, until Bai, Miao, and Pan [22]. In this chapter, we shall introduce some known results and some conjectures.

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_10,

10.1 Formulation and Conjectures

As an introduction, we first consider the behavior of eigenvectors of the $p \times p$ sample covariance matrix $\mathbf{S}_p$ studied in Chapter 3 when the x_{ij} are real, standardized, and iid. We will examine properties on the orthogonal $p \times p$ matrix $\mathbf{O}_p$, with columns containing the eigenvectors of $\mathbf{S}$, which we will call the eigenmatrix of $\mathbf{S}_p$, when viewed as a random element in $\mathscr{O}_p$, the space of $p \times p$ orthogonal matrices. This space is measurable when considered as a metric space with the metric taken as the operator norm of the difference of two matrices. Ambiguities do arise in defining the eigenmatrix of $\mathbf{S}_p$, due to the fact that there are 2^p different choices by changing the directions of the column eigenvectors. Whenever an eigenvalue has multiplicity greater than 1, the eigenmatrix will have infinitely many different choices for a given $\mathbf{S}_p$. However, it is later shown that there is a natural way to define a measure ν_p on $\mathscr{O}_p$ for which we can write $\mathbf{S}_p$ in its spectral decomposition $\mathbf{O}_p \Lambda_p \mathbf{O}_p'$ with Λ_p diagonal, its diagonal entries being the eigenvalues of $\mathbf{S}_p$, arranged, say, in ascending order, $\mathbf{O}_p$ orthogonal, columns consisting of the eigenvectors of $\mathbf{S}_p$, and $\mathbf{O}_p$ being ν_p-distributed.

We will investigate the behavior of $\mathbf{O}_p$ both with respect to its random versus deterministic tendencies and possible nondependence on the distribution of x_{11}. The former issue is readily settled when one considers x_{11} to be $N(0,1)$. Indeed, in this case, $n\mathbf{S}_p$ is a Wishart matrix, and the behavior of its eigenvectors is known. Before a description of this behavior can be made, some further definitions and properties need to be introduced.

10.1.1 Haar Measure and Haar Matrices

Besides being measurable, $\mathscr{O}_p$ forms a group under matrix multiplication. It is also a *compact topological group*: it is compact and the mappings $f_1 : \mathscr{O}_p \times \mathscr{O}_p \to \mathscr{O}_p$ and $f_2 : \mathscr{O}_p \to \mathscr{O}_p$ defined by $f_1(\mathbf{O}_1, \mathbf{O}_2) = \mathbf{O}_1\mathbf{O}_2$ and $f_2(\mathbf{O}) = \mathbf{O}^{-1}$ are continuous. The space $\mathscr{O}_p$ is typically called the $p \times p$ *orthogonal group*. Because of these properties on $\mathscr{O}_p$, there exists a unique probability measure h_p, called the uniform or *Haar* measure, defined as follows.

Definition 10.1. The probability measure h_p defined on the Borel σ-field $\mathscr{B}_{op}$ of Borel subsets of $\mathcal{O}_p$ is called Haar measure if, for any Borel set $A \in \mathscr{B}_{op}$ and orthogonal matrix $\mathbf{O} \in \mathscr{O}_p$, $h_p(\mathbf{O}A) = h_p(A)$, where $\mathbf{O}A$ denotes the set of all $\mathbf{OA}$, $\mathbf{A} \in A$.

If a p-dimensional random orthogonal matrix $\mathbf{H}_p$ is distributed according to Haar measure h_p, then it is called a p-dimensional Haar matrix.

(Haar measures defined on general topological groups can be found in Halmos [145].) It is remarked here that the definition of Haar measure is equivalent to $h_p(A\mathbf{O}) = h_p(A)$, with $A\mathbf{O}$ analogously defined (Halmos [145]).

Now, we quote some simple properties of Haar matrices.

Property 1. *If* $\mathbf{H}_p$ *is* h_p*-distributed, then for any unit* p*-vector* $\mathbf{x}_p$*,* $\mathbf{y}_p = \mathbf{H}_p\mathbf{x}_p$ *is uniformly distributed on the unit* p*-sphere.*

Proof. For any orthogonal $p \times p$ matrix $\mathbf{O}$, $\mathbf{O}\mathbf{y}_p = \mathbf{O}\mathbf{H}\mathbf{x}_p \stackrel{\mathscr{D}}{=} \mathbf{H}\mathbf{x}_p = \mathbf{y}_p$. Thus we have the distribution of $\mathbf{y}_p$ invariant under orthogonal transformations. Using the fact that this uniquely characterizes the uniform distribution on the unit p-sphere (see, for example, Silverstein [265]), we get our result.

Property 2. *If* $\mathbf{H}_p$ *is* h_p*-distributed, then* $\mathbf{H}_p'$*is also* h_p*-distributed.*

Proof. Let $\mathbf{O} \in \mathcal{O}_p$, A be a Borel subset of $\mathcal{O}_p$, and A' denote the set of all transposes of elements in A. Then

$$\mathrm{P}(\mathbf{O}\mathbf{H}_p' \in A) = \mathrm{P}(\mathbf{H}_p \in A'\mathbf{O}) = h_p(A') = \mathrm{P}(\mathbf{H}_p \in A') = \mathrm{P}(\mathbf{H}_p' \in A),$$

which implies $\mathbf{H}_p'$ is Haar-distributed.

Property 3. *If* $\mathbf{Z}$ *is a* $p \times p$ *matrix with entries iid* $N(0,1)$*, then* $\mathbf{U} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1/2}$ *and* $\mathbf{V} = (\mathbf{Z}\mathbf{Z}')^{-1/2}\mathbf{Z}$ *are* h_p*-distributed.*

The proof for $\mathbf{U}$ follows from the fact that, for any orthogonal matrix $\mathbf{O}$, $\mathbf{O}\mathbf{U} = \mathbf{O}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1/2} = \mathbf{O}\mathbf{Z}((\mathbf{O}\mathbf{Z})'\mathbf{O}\mathbf{Z})^{-1/2} \stackrel{\mathscr{D}}{=} \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1/2} = \mathbf{U}$. The proof for $\mathbf{V}$ is similar.

Property 4. *Assume that on a common probability space, for each* p*,* $\mathbf{H}_p$ *is* h_p*-distributed and* $\mathbf{x}_p$ *is a unit* p*-vector. Let* $\mathbf{y}_p = (y_1, \cdots, y_p)' = \mathbf{H}_p'\mathbf{x}_p$ *and* f *a bounded continuous function. Then, as* $p \to \infty$*,*

$$\frac{1}{p}\sum_{j=1}^{p} f(\sqrt{p}y_j) \to \int f(x)\varphi(x)dx, \ \ a.s., \tag{10.1.1}$$

where $\varphi(x)$ *is the density of* $N(0,1)$*.*

Proof. Let Φ denote the standard normal distribution function. By Properties 1 and 2, $\mathbf{y}_p$ is uniformly distributed over the unit p-sphere, and hence its distribution is the same as $\mathbf{z}_p/\|\mathbf{z}_p\|$, where $\mathbf{z}_p = (z_1, \cdots, z_p)'$, whose entries are iid $N(0,1)$. We may assume $\mathbf{y}_p = \mathbf{z}_p/\|\mathbf{z}_p\|$. Consider the empirical distribution function of the entries of $\sqrt{p}\mathbf{y}_p$,

$$F_p(x) = \frac{1}{p}\sum_{i=1}^{p} I_{(-\infty,x]}(\sqrt{p}y_i) = \frac{1}{p}\sum_{i=1}^{p} I_{(-\infty,(\|\mathbf{z}_p\|/\sqrt{p})x]}(z_i),$$

I_A denoting the indicator function on the set A. By the strong law of large numbers, we have $\|\mathbf{z}_p\|/\sqrt{p} \xrightarrow{\text{a.s.}} 1$. Therefore, for any $\epsilon > 0$, we have with probability 1

$$\liminf_p \frac{1}{p}\sum_{i=1}^p I_{(-\infty,x-\epsilon]}(z_i) \le \liminf_p F_p(x)$$
$$\le \limsup_p F_p(x) \le \limsup_p \frac{1}{p}\sum_{i=1}^p I_{(-\infty,x+\epsilon]}(z_i).$$

By the strong law of large numbers, the two extremes are equal almost surely to $\Phi(x-\epsilon)$ and $\Phi(x+\epsilon)$, respectively. Since ϵ is arbitrary, we get

$$F_p \xrightarrow{\mathscr{D}} \Phi, \quad \text{a.s., as } p \to \infty,$$

where $\xrightarrow{\mathscr{D}}$ denotes weak convergence on probability measures on $\mathbb{R}$. The result follows.

Property 5. *Let $D([0,1])$ denote the space of functions on $[0,1]$ with discontinuities of the first kind (right-continuous with left-hand limits, abbreviated to rcll), endowed with the Skorohod metric. If $\mathbf{y}_p$ is defined as in Property 4 for an arbitrary unit $\mathbf{x}_p \in \mathbb{R}^p$, then as $p \to \infty$, the random element*

$$X_p(t) = \sqrt{\frac{p}{2}}\sum_{j=1}^{[pt]}\left(|y_j^2| - \frac{1}{p}\right) \xrightarrow{\mathscr{D}} W_0, \tag{10.1.2}$$

where $\xrightarrow{\mathscr{D}}$ denotes weak convergence on $D[0,1]$, $[a]$ is the integer part of a, and W_0 is a Brownian bridge (also called tied down Brownian motion).

Proof. As in the previous property, we can assume $\mathbf{y}_p = \mathbf{z}_p/\|\mathbf{z}_p\|$, the entries of $\mathbf{z}_p$ being iid $N(0,1)$. Therefore,

$$X_p(t) = \sqrt{\frac{p}{2}}\sum_{j=1}^{[pt]}\left(\frac{z_j^2}{\|\mathbf{z}\|^2} - \frac{1}{p}\right)$$
$$= \frac{p}{\|\mathbf{z}\|^2}\frac{1}{\sqrt{2}}\frac{1}{\sqrt{p}}\left(\sum_{j=1}^{[pt]}(z_j^2-1) - \frac{[pt]}{p}\sum_{k=1}^{p}(z_k^2-1)\right).$$

Using Donsker's theorem and consequences of measurable mappings on $D[0,1]$, along with the fact that $\|\mathbf{z}\|^2/p \to 1,\ a.s.$, with W denoting standard Brownian motion, we get $X_p \xrightarrow{\mathscr{D}} W(t) - tW(1) = W_0(t)$.

Property 6. *Let $n\mathbf{S}_p$ be a $p \times p$ standard Wishart matrix with degrees of freedom n and $\mathbf{O}_p$ be the eigenmatrix of $\mathbf{S}_p$. Assume also that the signs of the first row of $\mathbf{O}_p$ are iid, each symmetrically distributed. Then $\mathbf{O}_p$ is h_p-distributed.*

We refer the reader to Anderson [5].

10.1.2 Universality

We then see from the last property that when the entries making up $\mathbf{S}_p$ are $N(0,1)$, $\mathbf{O}_p$, the eigenmatrix of $\mathbf{S}_p$, is Haar-distributed. Thus, for large p, in contrast to the eigenvalues of $\mathbf{S}_p$ behaving in a deterministic manner, the eigenvectors display completely chaotic behavior from realization to realization. Indeed, $\mathbf{H}_p$ is equally likely to be contained in any ball in $\mathcal{O}_p$ having the same radius (since any ball can be transformed to any other ball with the same radius by multiplying each element of the first ball with an orthogonal matrix).

As is seen for the eigenvalues of large dimensional random matrices, their limiting properties under certain moment conditions made on the underlying distributions are the same as when the entries are normally distributed. Thus, it is conceivable that the same is true in some sense for eigenmatrices of real sample covariance matrices as $p/n \to y > 0$. We shall address the possibility that, for large p, and x_{11} not N(0,1), ν_p (the distribution of $\mathbf{O}_p$) and h_p are in some way "close" to each other. We use the term *asymptotic Haar* or *Haar conjecture* to characterize this imprecise definition on sequences $\{\mu_p\}$ where, for each p, μ_p is a Borel probability measure on $\mathcal{O}_p$.

Formal definitions of asymptotic Haar on $\{\mu_p\}$ can certainly be made. For example, we could require for all $\epsilon > 0$ that we have, for all p sufficiently large, $|\mu_p(A) - h_p(A)| < \epsilon$ for every Borel set $A \subset \mathcal{O}_p$. However, because atomic (discrete) measures would not satisfy this definition, it would eliminate all μ_p arising from discrete x_{11}. Let $S_{p,o}$ denote the collection of all open balls in $\mathcal{O}_p$. Then an appropriate definition of asymptotic Haar that would not immediately eliminate atomic μ_p could be the following: for every $\epsilon > 0$, we have for all p sufficiently large $|\mu_p(A) - h_p(A)| < \epsilon$ for every $A \in S_{p,o}$.

In view of Properties 1 and 2, as an indication that $\{\mu_p\}$ is asymptotic Haar, one may consider the vector $\mathbf{y}_p = \mathbf{O}_p'\mathbf{x}_p$, $\|\mathbf{x}_p\| = 1$, when $\mathbf{O}_p$ is μ_p-distributed, and define asymptotic uniformity on the unit p-sphere, for example, by requiring for any $\epsilon > 0$ and open ball A of the unit p-sphere that we have $|\mathrm{P}(\mathbf{y}_p \in A) - s_p(A)| < \epsilon$ for all sufficiently large p, where s_p is the uniform probability measure on the unit p-sphere.

Instead of seeking a particular definition of asymptotic Haar or its consequences, attention should be drawn to the properties the sequence $\{h_p\}$ possesses, which are listed in the previous section, in particular Properties 4 and 5. They involve sequences of mappings, which have potential analytical tractability from $\mathbf{S}_p$, and map the $\mathcal{O}_p$'s into a common space. Property 4 maps $\mathcal{O}_p$ to the real line and considers nonrandom limit behavior. Simulations strongly suggest that the result holds for more general $\{\nu_p\}$, but a strategy to prove it has not been made.

The remainder of this chapter will focus on Property 5. There the mappings from the $\mathcal{O}_p$'s go into $D[0,1]$ and we consider distributional behavior. We convert this property to one on general $\{\mu_p\}$. We say that $\{\mu_p\}$ satisfies Property 5$'$ if for $\mathbf{O}_p$ μ_p-distributed we have for any sequence $\{\mathbf{x}_p\}$, $\mathbf{x}_p \in \mathbb{R}^p$ of unit vectors, with X_p defined as in Property 5 from the entries of $\mathbf{y}_p = \mathbf{O}'_p \mathbf{x}_p$, as $p \to \infty$,

$$X_p \xrightarrow{\mathscr{D}} W_0.$$

10.2 A Necessary Condition for Property 5′

It seems reasonable to consider Property 5′ as a necessary condition for asymptotic Haar. However, when investigating the eigenvectors of the sample convariance matrix $\mathbf{S}_p$, we get the following somewhat surprising result.

Theorem 10.2. *Assume that* $\mathrm{E}(x_{11}^4) < \infty$. *If* $\{\nu_p\}$ *satisfies Property* $5'$, *then* $\mathrm{E}(x_{11}^4) = 3$.

Proof. Since we are dealing with distributional behavior, we may assume the x_{ij} stem from a double array of iid random variables. With $\lambda_{\max}$ denoting the largest eigenvalue of $\mathbf{S}_p$, we then have from Theorem 5.8

$$\lim_{p\to\infty} \lambda_{\max} = (1+\sqrt{y})^2 \quad \text{a.s.} \tag{10.2.1}$$

We present here a review of the essentials of the Skorohod topologies on two metric spaces of rcll functions without explicitly defining the metric. On $D[0,1]$, elements x_n in this set converge to x in this set if and only if there exist functions $\lambda_n : [0,1] \to [0,1]$, continuous and strictly increasing with $\lambda_n(0) = 0$, $\lambda_n(1) = 1$, such that

$$\lim_{n\to\infty} \sup_{t\in[0,1]} \max(|x_n(\lambda_n(t)) - x(t)|, |\lambda_n(t) - t|) = 0.$$

Billingsley [57] introduces $D[0,1]$ in detail.

Let $D_0[0,\infty)$ denote the space of rcll functions $x(t)$ defined on $[0,\infty)$ such that $\lim_{t\to\infty} x(t)$ exists and is finite. Elements x_n converge to x in its topology if and only if there exist functions $\lambda_n : [0,\infty) \to [0,\infty)$, continuous and strictly increasing with $\lambda_n(0) = 0$, $\lim_{t\to\infty} \lambda_n(t) = \infty$, such that

$$\lim_{n\to\infty} \sup_{t\in[0,\infty)} \max(|x_n(\lambda_n(t)) - x(t)|, |\lambda_n(t) - t|) = 0$$

(see Lindvall [198]).

For both $D[0,1]$ and $D_0[0,\infty)$, it is straightforward to verify that when the limiting x is continuous, convergence in the Skorohod topology is equivalent to uniform convergence.

We let $\mathscr{D}[0,1]$ and $\mathscr{D}_0[0,\infty)$ denote the respective σ-fields of Borel sets of $D[0,1]$ and $D_0[0,\infty)$.

From Theorem 3.6, we know that $F^{\mathbf{S}_p}$, the ESD of $\mathbf{S}_p$, converges a.s. in distribution to $F_y(x)$ defined in (3.1.1) (with $\sigma^2 = 1$), the standard M-P law.

Since the limit is continuous on $[0, \infty)$, this convergence is uniform:

$$\sup_{t\in[0,\infty)} |F^{\mathbf{S}_p}(t) - F_y(t)| \xrightarrow{\text{a.s.}} 0 \quad \text{as } p \to \infty.$$

It follows then that

$$F^{\mathbf{S}_p} \xrightarrow{\text{a.s.}} F_y \quad \text{as } p \to \infty \text{ in } D_0[0,\infty). \tag{10.2.2}$$

Let $\underline{D}_0[0,\infty)$ denote the collection of all subprobability distribution functions on $[0,\infty)$. Clearly $\underline{D}_0[0,\infty) \subset D_0[0,\infty)$ and is closed in the Skorohod topology. Let $\underline{\mathscr{D}}_0[0,\infty)$ denote its σ-field of Borel sets.

Assume now that $\{\nu_p\}$ satisfies Property 5′. We have that $(X_p, F^{\mathbf{S}_p})$ are random elements of the product space $D[0,1] \times \underline{D}_0[0,\infty)$. Since the limit of the $F^{\mathbf{S}_p}$ is nonrandom, we have from (10.2.2) and Theorem 4.4 of Billingsley [57] that

$$(X_p, F^{\mathbf{S}_p}) \xrightarrow{\mathscr{D}} (W_0, F_y) \quad \text{as } p \to \infty. \tag{10.2.3}$$

The mapping $\psi : D[0,1] \times \underline{D}_0[0,\infty) \to D_0[0,\infty)$ defined by $\psi(x,\varphi) = x \circ \varphi$ is well defined, and using the same argument as in Billingsley [57], p. 232, it is measurable; that is, $\psi^{-1}\mathscr{D}_0[0,\infty) \subset \mathscr{D}[0,1] \times \underline{\mathscr{D}}_0[0,\infty)$. (Note that the proof of measurability relies on the fact that the mappings $\pi_{t_1,\ldots,t_k} : D_0[0,\infty) \to \mathbb{R}^k$ defined by $\pi_{t_1,\ldots,t_k}(x) = (x(t_1),\ldots,x(t_k))$ are measurable for all k and nonnegative $t_1,\ldots,t_k$, and their inverse images over all Borel sets in $\mathbb{R}^k$ generate $\mathscr{D}_0$; see Lindvall [198] p. 117.) Using the same argument as in Billingsley [57], p. 145, we see that the mapping ψ is continuous whenever both x and φ are continuous.

For a positive integer $m \le p$, let $D(p,m)$ denote the $p \times p$ matrix containing zeros, except for 1's in its first m diagonal positions. We then have

$$\psi(X_p, F^{\mathbf{S}_p})(x) = X_p(F^{\mathbf{S}_p}(x)) = \frac{\sqrt{p}}{\sqrt{2}}(\mathbf{x}_p'\mathbf{O}_p D(p,[pF^{\mathbf{S}_p}(x)])\mathbf{O}_p'\mathbf{x}_p - F^{\mathbf{S}_p}(x)).$$

We see that $\mathbf{x}_p'\mathbf{O}_p D(p,[pF^{\mathbf{S}_p}(x)])\mathbf{O}_p'\mathbf{x}_p$ is a (random) probability distribution function, the mass points being the eigenvalues of $\mathbf{S}_p$, while the mass values are the squares of the components of $\mathbf{O}_p'\mathbf{x}_p$.

Since $\mathrm{P}(W_0 \in C[0,1]) = 1$, where $C[0,1]$ denotes the space of continuous functions on $[0,1]$ (Billingsley [57]) and F_y is continuous, we get from Corollary 1 to Theorem 5.1 of Billingsley

$$\begin{aligned} X_p(F^{\mathbf{S}_p}) &\xrightarrow{\mathscr{D}} W_0(F_y(x)) \\ &\equiv W_x^y \quad \text{in } D_0[0,\infty) \text{ as } p \to \infty. \end{aligned} \tag{10.2.4}$$

For every positive integer r, using integration by parts, we have

$$\frac{\sqrt{p}}{\sqrt{2}}(\mathbf{x}_p'\mathbf{S}_p^r\mathbf{x}_p - (1/p)\mathrm{tr}\mathbf{S}_p^r) = \int_0^\infty x^r dX_p(F^{\mathbf{S}_p}(x))$$

$$= -\int_0^\infty rx^{r-1} X_p(F^{\mathbf{S}_p}(x))dx, \quad (10.2.5)$$

where we have used the fact that, with probability 1, $X_p(F^{\mathbf{S}_p}(x))$ is zero outside a bounded set.

It is straightforward to verify that, for any $b > 0$, the mapping that takes $\phi \in D_0[0, \infty)$ to $\int_0^b rx^{r-1} y\phi(x)dx$ is continuous. Therefore, from (10.2.4) and Corollary 1 to Theorem 5.1 of Billingsley [57], we have

$$\int_0^b rx^{r-1} X_p(F^{\mathbf{S}_p}(x))dx \xrightarrow{\mathscr{D}} \int_0^b rx^{r-1} W_x^y dx \qquad p \to \infty. \qquad (10.2.6)$$

From (10.2.1), we see that when $b > (1 + \sqrt{y})^2$

$$\int_0^\infty rx^{r-1} X_p(F^{\mathbf{S}_p}(x))dx - \int_0^b rx^{r-1} X_p(F^{\mathbf{S}_p}(x))dx \xrightarrow{\text{a.s.}} 0 \qquad \text{as } p \to \infty.$$

This, (10.2.5), and (10.2.6) yield

$$\frac{\sqrt{p}}{\sqrt{2}}(\mathbf{x}_p' \mathbf{S}_p^r \mathbf{x}_p - (1/p)\text{tr}\mathbf{S}_p^r) \xrightarrow{\mathscr{D}} -\int_0^b rx^{r-1} W_x^y dx = -\int_{(1-\sqrt{y})^2}^{(1+\sqrt{y})^2} rx^{r-1} W_x^y dx \qquad (10.2.7)$$

as $p \to \infty$. The limiting distribution, being the limit of Riemann sums, each sum being Gaussian, must necessarily be Gaussian, with mean 0 and covariance

$$\sigma^2_{y,r_1,r_2} = \iint_{(1-\sqrt{y})^2}^{(1+\sqrt{y})^2} r_1 r_2 s^{r_1-1} t^{r_2-1} [F_y(s \wedge t) - F_y(s)F_y(t)]dsdt,$$

where F_y is the distribution function of the M-P law with index y. By the extended Hoeffding lemma,[1] we conclude that

---

[1] The Hoeffding [150] lemma says that

$$\text{Cov}(X, Y) = \iint [\text{P}(X \le x, Y \le y) - \text{P}(X \le x)\text{P}(Y \le y)]dxdy.$$

From this, for any square integrable differentiable functions f, g, if both are increasing, by letting $x \to f(x), y \to g(y)$,

$$\begin{aligned}
&\text{Cov}(f(X), g(Y)) \\
&= \iint [\text{P}(f(X) \le x, g(Y) \le y) - \text{P}(f(X) \le x)\text{P}(g(Y) \le y)]dxdy \\
&= \iint f'(x)g'(y)[\text{P}(X \le x, Y \le y) - \text{P}(X \le x)\text{P}(Y \le y)]dxdy.
\end{aligned}$$

For functions of bounded variation, the equality above can be proved by writing f and g as differences of two increasing functions.

$$\sigma^2_{y,r_1,r_2} = \mathrm{Cov}(M_y^{r_1}, M_y^{r_2}) = \beta_{r_1+r_2} - \beta_{r_1}\beta_{r_2}, \tag{10.2.8}$$

where M_y is the random variable distributed according to the M-P law and

$$\beta_r = \sum_{j=0}^{r-1} \frac{1}{j+1}\binom{r}{j}\binom{r-1}{j} y^j.$$

Especially, when $r = 1$, $\sigma^2_{y,1,1} = \beta_2 - \beta_1^2 = y$. Hence, we have

$$-\int_{(1-\sqrt{y})^2}^{(1+\sqrt{y})^2} W_x^y dx = N(0, y). \tag{10.2.9}$$

With $\mathbf{x}_p = (1, 0, \ldots, 0)'$, by ordinary CLT, we have

$$\frac{\sqrt{p}}{\sqrt{2}}(\mathbf{x}_p'\mathbf{S}_p\mathbf{x}_p - 1)\sqrt{p/n}\frac{1}{\sqrt{2n}}\sum_j (x_{1j}^2 - 1) \xrightarrow{\mathscr{D}} N(0, (y/2)(\mathrm{E}(x_{11}^4) - 1)).$$

Because Property 5′ requires that the limits be independent of the choice of $\mathbf{x}_p$, we conclude that $\frac{y}{2}(\mathrm{E}x_{11}^4 - 1) = y$, which implies $\mathrm{E}x_{11}^4 = 3$. The proof of the theorem is complete.

Thus property 5′ requires, under the assumption of a finite fourth moment, the first, second, and fourth moments of x_{11} to be identical to those of a Gaussian variable. This suggests the possibility that only Wishart $\mathbf{S}_p$ will satisfy this property, which at present has not been established.

We see from the proof that Property 5′ and $\mathrm{E}(x_{11}^4) < \infty$ yield (10.2.7), which can be viewed as the convergence of moments of $X_p(F^{\mathbf{S}_p})$. It is worthwhile to consider whether this limiting result is true, partly for its own contribution in displaying eigenvector behavior but mainly because it will be shown later to be important in verifying weak convergence of X_p under certain assumptions on $\mathbf{x}_p$ and the distribution of x_{11}. The next section gives a complete analysis of (10.2.7).

10.3 Moments of $X_p(F^{\mathbf{S}_p})$

In this section, we will prove the following theorem.

Theorem 10.3. *Assume no a priori conditions on the moments of x_{11}.*

(*a*) *We have*

$$\{\sqrt{p/2}(\mathbf{x}_p'\mathbf{S}_p^r\mathbf{x}_p - (1/p)\mathrm{tr}S^r)\}_{r=1}^\infty \xrightarrow{\mathscr{D}} \left\{-\int_{(1-\sqrt{y})^2}^{(1+\sqrt{y})^2} rx^{r-1}W_x^y dx\right\}_{r=1}^\infty \tag{10.3.1}$$

in $\mathbb{R}^\infty$ as $p \to \infty$ for every sequence $\mathbf{x}_p$, $\mathbf{x}_p \in \mathbb{R}^p$ $\|\mathbf{x}_p\| = 1$ if and only if

$$\mathrm{E}(x_{11}) = 0, \qquad \mathrm{E}(x_{11}^2) = 1, \qquad \mathrm{E}(x_{11}^4) = 3. \tag{10.3.2}$$

(b) If $\int_0^\infty x dX_p(F_p(x))$ is to converge in distribution to a random variable for each of the $\mathbf{x}_p$ sequences

$$\{(1, 0, \ldots, 0)'\}, \quad \{(1/\sqrt{p}, \ldots, 1/\sqrt{p})'\},$$

then necessarily $\mathrm{E}(x_{11}^4) < \infty$ and $\mathrm{E}(x_{11}) = 0$.

(c) If $\mathrm{E}(x_{11}^4) < \infty$ but

$$\frac{\mathrm{E}(x_{11} - \mathrm{E}(x_{11}))^4}{\mathrm{Var}^2(x_{11})} \neq 3,$$

then there exist sequences $\{\mathbf{x}_p\}$ of unit vectors for which

$$\left(\int_0^\infty x dX_p(F_p(x)), \int_0^\infty x^2 dX_p(F_p(x)) \right)$$

fails to converge in distribution.

The proof will be given in the following subsections.

10.3.1 Proof of (10.3.1) ⇒ (10.3.2)

Assume (10.3.1). By choosing $\mathbf{x}_p = (1, 0, \cdots, 0)'$, by (10.3.1) with $r = 1$ and (10.2.9), we have

$$\frac{(p-1)}{\sqrt{2pn}} \sum_{j=1}^n x_{1j}^2 - \frac{1}{\sqrt{2pn}} \sum_{i=2}^p \sum_{j=1}^n x_{ij}^2 \xrightarrow{\mathscr{D}} N(0, y). \tag{10.3.3}$$

By necessary and sufficient conditions for the CLT of sums of independent random variables (see Loève [200], Section 23.5, p. 238), we conclude that

$$\frac{(p-1)^2}{2np} \mathrm{Var}(x_{11}^2 I(|x_{11}^2| < \sqrt{n})) + \frac{(p-1)}{2pn} \mathrm{Var}(X_{11}^2 I(|x_{11}^2| < \sqrt{p}n)) \to y. \tag{10.3.4}$$

Noting $\frac{(p-1)^2}{2np} \to \frac{y}{2}$, the limit above implies that $\mathrm{Var}(X_{11}^2) = 2$ and then $\mathrm{E}x_{11}^4 < \infty$.

Next, consider the convergence for $\mathbf{x}_p = \mathbf{1}/\sqrt{p}$. We will have

$$\frac{1}{\sqrt{2pn}} \sum_{i \neq k \leq p} \sum_{j=1}^n (x_{ij} x_{kj}) \xrightarrow{\mathscr{D}} N(0, y). \tag{10.3.5}$$

The expectation and variance of the left-hand side are $\frac{\sqrt{p}(p-1)}{\sqrt{2}}(\mathrm{E}x_{11})^2$ and $\frac{p-1}{n}\mathrm{Var}(x_{11}x_{12})$, which imply $\mathrm{E}x_{11}=0$ for otherwise the LHS tends to infinity.

Finally, consider the convergence for $\mathbf{x}_p = \frac{1}{\sqrt{2}}(1,1,0,\cdots,0)'$. Then, by (10.3.1), we have

$$\frac{1}{\sqrt{2pn}}\left[\frac{p-2}{2}\sum_{\substack{i=1,2\\ j\le n}}(x_{ij}^2-\mathrm{E}x_{11})-\sum_{\substack{3\le i\le p\\ j\le n}}(x_{ij}^2-\mathrm{E}x_{11}^2)+p\sum_{j=1}^{n}x_{1j}x_{2j}\right]\xrightarrow{\mathscr{D}}N(0,y).$$

On the other hand, by the CLT, we know that the LHS tends to normal with mean zero and variance $\frac{y}{2}(1+(\mathrm{E}x_{11}^2)^2)$. Equating it to y, we obtain $\mathrm{E}x_{11}^2=1$. Then, by $\mathrm{Var}(x_{11}^2)=2$ shown before, we obtain $\mathrm{E}x_{11}^4=3$. This completes the proof of (10.3.2).

10.3.2 Proof of (b)

When $\int_0^\infty x dX_p(F_p(x))$ converges in distribution to a random variable for $\mathbf{x}_p=(1,0,\ldots,0)'$, we conclude that the LHS of (10.3.3) tends to a random variable in distribution that must be an infinitely divisible law. By Section 23.4, p. 323, of Loève [200], we conclude that the LHS of (10.3.4) tends to a nonnegative constant. By the same reasoning as argued in the last subsection, we conclude that $\mathrm{E}x_{11}^4<\infty$.

When $\int_0^\infty x dX_p(F_p(x))$ with $\mathbf{x}_p=\mathbf{1}/\sqrt{p}$ converges in distribution to a random variable, we conclude that the LHS of (10.3.5) tends to a random variable in distribution. Similarly, by considering its mean and variance, we obtain $\mathrm{E}x_{11}=0$. This completes the proof of (b).

10.3.3 Proof of (10.3.2) ⇒ (10.3.1)

Assume (10.3.2). As in Theorem 10.2, we can assume (10.2.1). We begin by truncating and centralizing the x_{ij}. Following the truncation given in Subsection 5.2.1, we may select a sequence $\delta=\delta_n\to 0$ and let $\hat{x}_{ij}=\hat{x}_{ij}(p)=x_{ij}I(|x_{ij}|\le\delta\sqrt{p})$ and $\widehat{\mathbf{S}}_p=(1/n)\widehat{\mathbf{X}}_p\widehat{\mathbf{X}}_p'$, where $\widehat{\mathbf{X}}_p=(\hat{x}_{ij})$. Then we can prove that with probability 1, for all large p, $\mathbf{S}_p=\widehat{\mathbf{S}}_p$ (that is, for any measurable function f_p on $p\times p$ matrices),

$$|f_p(\mathbf{S}_p)-f_p(\widehat{\mathbf{S}}_p)|\xrightarrow{\text{a.s.}}0 \quad \text{as } p\to\infty. \tag{10.3.6}$$

Let $\tilde{\mathbf{S}}_p=(1/n)(\widehat{\mathbf{X}}_p-\mathrm{E}(\hat{x}_{11})\mathbf{1}_p\mathbf{1}_n')(\widehat{\mathbf{X}}_p-\mathrm{E}(\hat{x}_{11})\mathbf{1}_p\mathbf{1}_n')'$, where $\mathbf{1}_m$ denotes the m-dimensional vector consisting of 1's. By Theorem A.46, we have

$$\max_{j \le p} \sqrt{p} |\lambda_j^{1/2}(\widehat{\mathbf{S}}_p) - \lambda_j^{1/2}(\tilde{\mathbf{S}}_p)| \le \|(\sqrt{p/n}) \mathrm{E}(\hat{x}_{11}) \mathbf{1}_p \mathbf{1}'_n\|$$

$$= p|\mathrm{E}(\hat{x}_{11})| \to 0 \quad \text{as } p \to \infty, \tag{10.3.7}$$

where we have used the fact that $\mathrm{E}(x_{11}) = 0$, $\mathrm{E}(x_{11}^4) < \infty$ implies $\mathrm{E}(x_{11}) = o(p^{-3/2})$. Consequently, $\lambda_{\max}(\tilde{\mathbf{S}}_p) \xrightarrow{\text{a.s.}} (1 + \sqrt{y})^2$ as $p \to \infty$.

It is straightforward to show for any $p \times p$ matrices $\mathbf{A}$, $\mathbf{B}$, and integer $r \ge 1$ that $\|(\mathbf{A} + \mathbf{B})^r - \mathbf{B}^r\| \le r\|\mathbf{A}\|(\|\mathbf{A}\| + \|\mathbf{B}\|)^{r-1}$. Therefore, with probability 1, for all large p,

$$\begin{aligned}
&\sqrt{p}|\mathbf{x}'_p(\tilde{\mathbf{S}}_p)^r \mathbf{x}_p - \mathbf{x}'_p(\widehat{\mathbf{S}}_p)^r \mathbf{x}_p| \le \sqrt{p}\|(\tilde{\mathbf{S}}_p)^r - (\widehat{\mathbf{S}}'_p)^r\| \\
&\le \sqrt{p} r \|\tilde{\mathbf{S}}_p - \widehat{\mathbf{S}}_p\|(\lambda_{\max}(\tilde{\mathbf{S}}_p) + 2\lambda_{\max}(\widehat{\mathbf{S}}_p))^{r-1} \\
&\le \sqrt{p} r 3^{r-1}(1 + \sqrt{y})^{2r-2}\|\tilde{\mathbf{S}}_p - \widehat{\mathbf{S}}_p\|.
\end{aligned}$$

We also have for any $p \times n$ matrices $\mathbf{A}$, $\mathbf{B}$ of the same dimension $\|\mathbf{A}\mathbf{A}' - \mathbf{B}\mathbf{B}'\| \le \|\mathbf{A} - \mathbf{B}\|(\|\mathbf{A}\| + \|\mathbf{B}\|)$. Therefore,

$$\sqrt{p}\|\tilde{\mathbf{S}}_p - \widehat{\mathbf{S}}_p\| \le \sqrt{p}\|(1/\sqrt{n})\mathrm{E}(\hat{x}_{11})\mathbf{1}_p\mathbf{1}'_n\|(\lambda_{\max}^{1/2}(\tilde{\mathbf{S}}_p) + \lambda_{\max}^{1/2}(\widehat{\mathbf{S}}_p))$$

$$= p|\mathrm{E}(\hat{x}_{11})|(\lambda_{\max}^{1/2}(\tilde{\mathbf{S}}_p) + \lambda_{\max}^{1/2}(\widehat{\mathbf{S}}_p)) \xrightarrow{\text{a.s.}} 0 \quad \text{as } p \to \infty.$$

Therefore

$$\sqrt{p}|\mathbf{x}'_p(\tilde{\mathbf{S}}_p)^r \mathbf{x}_p - \mathbf{x}'_p(\widehat{\mathbf{S}}_p)^r \mathbf{x}_p| \xrightarrow{\text{a.s.}} 0 \quad \text{as } p \to \infty. \tag{10.3.8}$$

Let $\tilde{\lambda}_i$, $\hat{\lambda}_i$ denote the respective eigenvalues of $\tilde{\mathbf{S}}_p$, $\widehat{\mathbf{S}}_p$, arranged in nondecreasing order. We have

$$\sqrt{p}|(1/p)\mathrm{tr}(\tilde{\mathbf{S}}_p)^r - (1/p)\mathrm{tr}(\widehat{\mathbf{S}}_p)^r| \le \sqrt{p}\max_{i \le p}|\tilde{\lambda}_i^r - \hat{\lambda}_i^r|$$

$$\le 2r\sqrt{p}\max_{i \le p}|\tilde{\lambda}_i^{1/2} - \hat{\lambda}_i^{1/2}|\Big(\max(\lambda_{\max}(\tilde{\mathbf{S}}_p), \lambda_{\max}(\widehat{\mathbf{S}}_p))\Big)^{\frac{1}{2}(2r-1)} \xrightarrow{\text{a.s.}} 0 \tag{10.3.9}$$

as $p \to \infty$. Therefore, by (10.3.6), (10.3.7), (10.3.8), and (10.3.3), we have for each integer $r \ge 1$,

$$|\sqrt{p/2}(\mathbf{x}'_p\mathbf{S}_p^r\mathbf{x}_p - (1/p)\mathrm{tr}(\mathbf{S}_p^r)) - \sqrt{p/2}(\mathbf{x}'_p\tilde{\mathbf{S}}_p^r\mathbf{x}_p - (1/p)\mathrm{tr}(\tilde{\mathbf{S}}_p^r))| \xrightarrow{\text{a.s.}} 0$$

as $p \to \infty$. We see then that, returning to the original notation, it is sufficient to prove (10.3.2) assuming $\mathrm{E}(x_{11}) = 0$, $|x_{11}| \le \delta\sqrt{p}$, $\mathrm{E}(x_{11}^2) \to 1$, $\mathrm{E}(x_{11}^4) \to 3$, as $p \to \infty$. By the existence of the fourth moment, we also have

$$\mathrm{E}(|x_{11}|^\ell) = o(p^{(\ell-4)/2}) \quad \text{for } \ell > 4. \tag{10.3.10}$$

We proceed with verifying two lemmas.

Lemma 10.4. *After truncation, for any integer* $r \geq 1$, $p^{-1/2}(\mathrm{tr}(\mathbf{S}_p^r) - \mathrm{E}(\mathrm{tr}(\mathbf{S}_p^r))) \xrightarrow{i.p.} 0$ *as* $p \to \infty$.

Proof. After truncation, by (3.2.4), we have $\mathrm{E}|\mathrm{tr}(\mathbf{S}_p^r) - \mathrm{E}(\mathrm{tr}(\mathbf{S}_p^r))|^4 = o(p^2)$, which proves the lemma.

Lemma 10.5. *For any integer* $r \geq 1$, $\sqrt{p}(\mathrm{E}(\mathbf{x}_p'\mathbf{S}_p^r\mathbf{x}_p) - \mathrm{E}((1/n)\mathrm{tr}(\mathbf{S}_p^r))) \to 0$ *as* $p \to \infty$.

Proof. Using the fact that the diagonal elements of $\mathbf{S}_p^r$ are identically distributed, we have

$$n^r\left(\mathrm{E}(\mathbf{x}_p'\mathbf{S}_p^r\mathbf{x}_p) - \mathrm{E}\left(\frac{1}{p}\mathrm{tr}(\mathbf{S}_p^r)\right)\right) = \sum_{i\neq j} x_i x_j \sum_{\substack{i_2,\ldots,i_r \\ k_1,\ldots,k_r}} \mathrm{E}(x_{ik_1}x_{i_2k_1}\cdots x_{jk_r}). \tag{10.3.11}$$

In accordance with Subsection 3.1.2, we draw a *chain graph* of $2r$ edges, $i_t \to k_t$ and $k_t \to i_{t+1}$, where $i_1 = i$, $i_{r+1} = j$, and $t = 1, \cdots, r$. An example of such a graph is shown in Fig. 10.1.

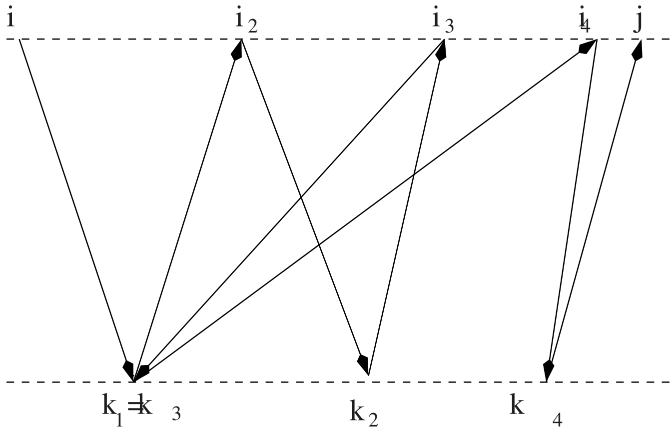

Fig. 10.1 A chain graph.

If there is a single edge, the corresponding term is 0. Therefore, we need only consider graphs that have no single edges. Suppose there are ℓ noncoincident edges with multiplicities $\nu_1, \cdots, \nu_\ell$. We have the constraint that $\nu_t \geq 2$ and $\nu_1 + \cdots + \nu_\ell = 2r$. Because the vertices i and j are the initial and end vertices of the chain and $i \neq j$, the degrees of the vertices i and j must be odd. Hence, there is at least one noncoincident edge connecting each of the vertices i and j and having a multiplicity ≥ 3. Therefore, the term is bounded by

$$B(\delta\sqrt{n})^{2r-2\ell-2} \le Bn^{r-\ell-1}.$$

Because the graph is connected, the number of noncoincident vertices is not greater than $\ell+1$ (including i and j).

Noting that $|\sum_{i\neq j} x_i x_j| \le p-1$, the RHS of (10.3.11) is bounded by

$$Bn^{-r}n^{r-\ell-1}n^{\ell-1}\sum_{i\neq j}|x_i x_j| \le O(n^{-1}).$$

From this the lemma follows.

Because of Lemmas 10.4 and 10.5, we see that (10.3.1) is equivalent to

$$\{\sqrt{p/2}(\mathbf{x}_p'\mathbf{S}_p^r\mathbf{x}_p - \mathrm{E}(\mathbf{x}_p'\mathbf{S}_p^r\mathbf{x}_p))\}_{r=1}^{\infty} \xrightarrow{\mathscr{D}} \{N_r\}_{r=1}^{\infty} \tag{10.3.12}$$

in $\mathbb{R}^\infty$ as $p \to \infty$, where $\{N_r\}$ are jointly normally distributed with mean 0 and covariance σ^2_{y,r_1,r_2} given in (10.2.8). We will use a multidimensional version of the method of moments (see Section B.1) to show that all mixed moments of the entries in (10.3.12) are bounded and that any asymptotic behavior depends solely on $\mathrm{E}(x_{11})$, $\mathrm{E}(x_{11}^2)$, and $\mathrm{E}(x_{11}^4)$. We know that (10.3.1) is true when x_{11} is $N(0,1)$ and, because of the two lemmas, (10.3.12) holds as well. Bounded mixed moments will imply, when x_{11} is $N(0,1)$, that the mixed moments of (10.3.12) converge to their proper values. The dependence of the limiting behavior of the mixed moments on $\mathrm{E}(x_{11})$, $\mathrm{E}(x_{11}^2)$, and $\mathrm{E}(x_{11}^2)$ implies that the moments in general will converge to the same values. The fact that a multivariate normal distribution is uniquely determined by its moments will then imply (10.3.12).

To apply the moment convergence theorem, we need a second step of truncation and centralization. Let $\tilde{x}_{ij} = x_{ij}I(|x_{ij}| < \log p) - \mathrm{E}x_{ij}I(|x_{ij}| < \log p)$ and write $\widetilde{\mathbf{S}}_p = n^{-1}\sum_{k=1}^n \tilde{x}_{ik}\tilde{x}_{jk}$. To this end, we need the following lemma.

Select index sets

$$\begin{aligned}
I &= \{i_1^1, i_2^1, \cdots, i_1^m, i_2^m\},\\
J &= \{j_2^1, \cdots, j_{r_1}^1, \cdots, j_2^m, \cdots, j_{r_m}^m\},\\
K &= \{k_1^1, \cdots, k_{r_1}^1, \cdots, k_1^m, \cdots, k_{r_m}^m\},
\end{aligned}$$

where $r_1, \cdots, r_m$ and the indices are positive integers. For each $t = 1, \cdots, m$, construct a chain graph G_t with vertices $\{i_1^t, i_2^t, j_2^t, \cdots, j_{r_t}^t, k_1^t, \cdots, k_{r_t}^t\}$ and $2r_t$ edges:

$$\{(i_1^t, k_1^t), (k_1^t, j_2^t), (j_2^t, k_2^t), \cdots, (j_{r_t}^t, k_{r_t}^t), (k_{r_t}^t, i_2^t)\}.$$

Combine the chain graphs $G = \bigcup_{t=1}^m G_t$. An example of G with $m=2$ is shown in Fig. 10.2. The indices are called I-, J-, and K-indices in accordance with the index set they belong to. A noncoincident vertex is called an L-vertex if it consists of only one I-index and some J-indices. A noncoincident vertex is called a J-(or K-)vertex if it consists of only J-(K- correspondingly) indices. A vertex is called a D-vertex if it is a J- or K-vertex. Denote the numbers of D- or L-vertices by d and l, respectively. We also denote by r' the number of noncoincident edges and write $r = r_1 + \cdots + r_m$. Let ι_α denote the number of noncoincident edges of multiplicity α. Then we have the following lemma.

Lemma 10.6. *If G does not have single edges and no subgraph G_t is separated from all others by edges (i.e., having at least one edge coincident with edges of other subgraphs), we have*

$$d \leq \begin{cases} r - \frac{3}{4}l - \frac{1}{2}m - \frac{1}{2}g = 2r' - \frac{3}{4}l - \frac{1}{2}(m + 2\iota_2 + \iota_3), & \textit{if } m \leq 2, \\ r - \frac{3}{4}l - \frac{1}{2}m - \frac{1}{4}g, & \textit{for any } m > 2, \end{cases}$$

where $g = \iota_5 + 2\iota_6 + \cdots$.

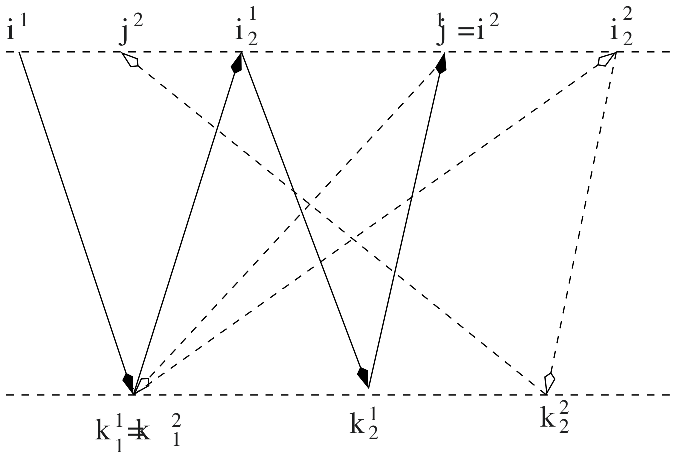

Fig. 10.2 A chain graph. The solid arrows form one chain and the broken arrows form another chain.

Proof. Consider the graph $\widetilde{G}$ of noncoincident edges and their vertices of G. A subgraph Γ of $\widetilde{G}$ is called a regular subtree if (i) it is a tree, (ii) all its edges have multiplicity 2, (iii) all its vertices consist of I_d-indices, and (iv) only one root connects to the rest of $\widetilde{G}$. A regular subtree is called maximal if it is not a proper subgraph of another regular subtree. Note that all edges of a regular subtree must come from one subgraph G_t. If a maximal regular subtree of μ edges is removed from $\widetilde{G}$, what is left is a graph combined from m subgraphs of sizes $r_1, \cdots, r_{t-1}, r_t - \mu, r_{t+1}, \cdots, r_m$. Now, remove all maximal regular subtrees. Suppose the total number of edges of all maximal regular subtrees is ν_1. In the remaining graph, the numbers r, r', d, and ι_2 are reduced by ν_1, and other numbers do not change.

Next, we consider the remaining graph Γ. If there is a root of multiplicity 2, its end vertex must be a K-vertex of subgraph G_t and the suspending vertex must consist of an I-index and a J-index, both of which belong to G_t, for otherwise it would have been removed if both are J-indices or the subgraph G_t is separated from others by edges if both are I-indices. Such a root is called an irregular root. If we remove this root and relabel the J-index as the removed I-index, the resulting graph is still a graph of the same kind with r_t, d, and ι_2 reduced by 1 and other numbers remain unchanged. Denote the number of irregular roots by ν_2. After removing all irregular roots, in the final remaining graph Γ, the numbers r, r', d, and ι_2 will be further reduced by ν_2 and other numbers remain unchanged.

In the remaining graph, if an I_d-vertex is an end vertex, the multiplicity of the edge connecting the I_d-vertex is larger than or equal to 4. Otherwise, the I_d-vertex connects at least two noncoincident edges. In both cases, the degree of the I_d-vertex is not less than 4. Because the degree of an I_d-vertex must be even, the number of noncoincident edges of odd multiplicities must be even. Now, we first consider the d_k K-vertices. If a K-vertex K_i connects $\iota_\alpha(i)$ noncoincident edges of multiplicity α, then its degree is

$$\vartheta_i = \sum_{\alpha \ge 2} \alpha \iota_\alpha(i) \ge 4 + \sum_{\substack{\alpha \ge 3 \\ \text{odd}}} (\alpha - 2)\iota_\alpha(i) + \sum_{\substack{\alpha \ge 4 \\ \text{even}}} (\alpha - 4)\iota_\alpha(i).$$

Summing these inequalities, we obtain

$$\begin{aligned} 2\tilde{r} &\ge 4d_k + \sum_{\substack{\alpha \ge 3 \\ \text{odd}}} (\alpha - 2)\iota_\alpha + \sum_{\substack{\alpha \ge 4 \\ \text{even}}} (\alpha - 4)\iota_\alpha \\ &\ge 4d_k + l + g, \end{aligned}$$

where we have used the fact that each I-vertex must connect at least one noncoincident edge of odd multiplicity. Therefore, we obtain

$$\tilde{r} \ge 2d_k + \frac{1}{2}(l + g).$$

Next, we consider the $\tilde{r} - m$ J-indices. There are at least l J-indices coincident with the l L-vertices, and thus we obtain

$$\tilde{r} - m \ge 2d_j + l,$$

where d_j is the number of J-vertices. Combining the two inequalities above and noting that $\tilde{d} = d_k + d_j$, we obtain

$$\tilde{d} \le \tilde{r} - \frac{3}{4}l - \frac{1}{2}m - \frac{1}{4}g.$$

This proves the case where $m > 2$. If $m \leq 2$, there are at most four I-indices. In this case, an edge of multiplicity $\alpha > 4$ always has a vertex consisting of at least $(\alpha - 4)/2$ J-indices. If the vertex is an L-vertex, then it consists of $(\alpha - 1)/2$ J-indices. Therefore, the second inequality becomes

$$\tilde{r} - m \geq 2d_j + l + g/2.$$

Then

$$\tilde{d} \leq \tilde{r} - \frac{3}{4}l - \frac{1}{2}(m + g).$$

The lemma then follows by noting that $r = \tilde{r} + \nu_1 + \nu_2$ and $d = \tilde{d} + \nu_1 + \nu_2$.

Second step of truncation and centralization

Expanding both $\mathbf{x}_p'\mathbf{S}_p\mathbf{x}_p - \mathrm{E}\mathbf{x}_p'\widetilde{\mathbf{S}}_p\mathbf{x}_p$, we obtain

$$\begin{aligned}
&p\mathrm{E}[(\mathbf{x}_p'\mathbf{S}_p^r\mathbf{x}_p - \mathrm{E}(\mathbf{x}_p'\mathbf{S}_p^r\mathbf{x}_p)) - (\mathbf{x}_p'\widetilde{\mathbf{S}}_p^r\mathbf{x}_p - \mathrm{E}(\mathbf{x}_p'\widetilde{\mathbf{S}}_p^r\mathbf{x}_p))]^2 \\
&= pn^{-2r}\sum\nolimits^* x_{i_1^1}x_{i_2^1}x_{i_1^2}x_{i_2^2}\mathrm{E}\Bigg[\Big(x_{i_1^1k_1^1}x_{j_2^1k_1^1}x_{j_2^1k_2^1}\cdots x_{j_{r_1}^1k_r^1}x_{i_2^1k_r^1} \\
&\quad -\mathrm{E}(x_{i_1^1k_1^1}x_{j_2^1k_1^1}x_{j_2^1k_2^1}\cdots x_{j_r^1k_r^1}x_{i_2^1k_r^1}) + \tilde{x}_{i_1^1k_1^1}\tilde{x}_{j_2^1k_1^1}\tilde{x}_{j_2^1k_2^1}\cdots \tilde{x}_{j_{r_1}^1k_r^1}\tilde{x}_{i_2^1k_r^1} \\
&\quad -\mathrm{E}(\tilde{x}_{i_1^1k_1^1}\tilde{x}_{j_2^1k_1^1}\tilde{x}_{j_2^1k_2^1}\cdots \tilde{x}_{j_r^1k_r^1}\tilde{x}_{i_2^1k_r^1})\Big)\Big(x_{i_1^2k_1^2}x_{j_2^2k_1^2}x_{j_2^2k_2^2}\cdots x_{j_{r_1}^2k_r^2}x_{i_2^2k_r^2} \\
&\quad -\mathrm{E}(x_{i_1^2k_1^2}x_{j_2^2k_1^2}x_{j_2^2k_2^2}\cdots x_{j_r^2k_r^2}x_{i_2^2k_r^2}) - \tilde{x}_{i_1^2k_1^2}\tilde{x}_{j_2^2k_1^2}\tilde{x}_{j_2^2k_2^2}\cdots \tilde{x}_{j_r^2k_r^2}\tilde{x}_{i_2^2k_r^1} \\
&\quad +\mathrm{E}(\tilde{x}_{i_1^2k_1^2}\tilde{x}_{j_2^2k_1^2}\tilde{x}_{j_2^2k_2^2}\cdots \tilde{x}_{j_r^2k_r^2}\tilde{x}_{i_2^2k_r^2})\Big)\Bigg],
\end{aligned} \qquad (10.3.13)$$

where the summation $\sum^*$ is taken for

$$\begin{aligned}
i_1^1, i_2^1, j_2^1, \ldots, j_r^1 \leq p,\ k_1^1, \ldots, k_r^1 \leq n, \\
i_1^2, i_2^2, j_2^2, \ldots, j_r^2 \leq p,\ k_1^m, \ldots, k_r^m \leq n.
\end{aligned}$$

Using these indices, we construct graphs G_1, G_2, and the combined G and use the notation defined in Lemma 10.6.

The absolute value of the sum of terms corresponding to a graph with numbers d, l, and g of the RHS of (10.3.13) is less than

$$Cpn^{-2r}p^{d+l/2}(\delta_p\sqrt{p})^g,$$

where we have used the inequality $\sum |x_i| \leq \sqrt{p}$. By Lemma 10.6, the sum tends to 0 if $g > 0$ or $l > 0$ or $d < 2r - 1$. When $g = l = 0$ and $d = 2r - 1$, there are two cases: either $\iota_2 = 2r$ or $\iota_2 = 2r - 2$ and $\iota_4 = 1$. That means the expansion of the expectation in (10.3.13) contains only the second and fourth moments of x_{11} and $\tilde{x}_{11}$. Because of the truncation, both the second and fourth moments of x_{11} and $\tilde{x}_{11}$ will tend to the same corresponding value. Thus we conclude that the absolute value of the expectation in (10.3.13) tends to 0 and thus the LHS of (10.3.13) tends to 0.

Completion of the proof of (10.3.1)

We shall complete the proof of (10.3.1) by showing (10.3.12) under the assumption $|x_{ij}| \le \log p$.

Any mixed moment can be written as

$$p^{m/2}\mathrm{E}[(\mathbf{x}_p'\mathbf{S}_p^{r_1}\mathbf{x}_p - \mathrm{E}(\mathbf{x}_p'\mathbf{S}_p^{r_1}\mathbf{x}_p))\cdots(\mathbf{x}_p'\mathbf{S}_p^{r_m}\mathbf{x}_p - \mathrm{E}(\mathbf{x}_p'\mathbf{S}_p^{r_m}\mathbf{x}_p))], \quad (10.3.14)$$

where the integer $m \ge 2$ and positive integers $r_1, \ldots, r_m$ are arbitrary. Expanding further, we have

$$\begin{aligned}
&(n^r p^{-m/2}) \times (10.3.14) \\
&= \sum\nolimits^{**} x_{i_1^1} x_{i_2^1} \cdots x_{i_1^m} x_{i_2^m} \mathrm{E}\Big[\Big(x_{i_1^1 k_1^1} x_{j_2^1 k_1^1} x_{j_2^1 k_2^1} \cdots x_{j_{r_1}^1 k_{r_1}^1} x_{i_2^1 k_{r_1}^1} \\
&\quad -\mathrm{E}(x_{i_1^1 k_1^1} x_{j_2^1 k_1^1} x_{j_2^1 k_2^1} \cdots x_{j_{r_1}^1 k_{r_1}^1} x_{i_2^1 k_{r_1}^1})\Big) \\
&\quad \cdots \Big(x_{i_1^m k_1^m} x_{j_2^m k_1^m} x_{j_2^m k_2^m} \cdots x_{j_{r_m}^m k_{r_m}^m} x_{i_2^m k_{r_m}^m} \\
&\quad -\mathrm{E}(x_{i_1^m k_1^m} x_{j_2^m k_1^m} x_{j_2^m k_2^m} \cdots x_{j_{r_m}^m k_{r_m}^m} x_{i_2^m k_{r_m}^m})\Big)\Big], \qquad (10.3.15)
\end{aligned}$$

where the summation $\sum^{**}$ is taken for

$$\begin{gathered}
i_1^1, i_2^1, j_2^1, \ldots, j_{r_1}^1 \le p, \; k_1^1, \ldots, k_{r_1}^1 \le n \\
\vdots \\
i_1^m, i_2^m, j_2^m, \ldots, j_{r_m}^m \le p, \; k_1^m, \ldots, k_{r_m}^m \le n.
\end{gathered}$$

Using the notation of Lemma 10.6, we use the indices $i_1^t, i_2^t, j_2^t, \ldots, j_{r_t}^t (\le p)$, $k_1^t, \ldots, k_{r_t}^t (\le n)$ to construct a graph G_t and let $G = G_1 \cup \cdots \cup G_m$.

We see a zero term if in the corresponding graph
(1) there is a single edge in G, or
(2) there is a graph G_t that does not have any coincident edges with another graph $G_{t'}$, $t' \ne t$.

Then the contribution to (10.3.15) of those terms associated with such a graph G is bounded in absolute value by

$$Kp^{(l/2)+d}\mathrm{E}(|x_{i^1 k_1^1} \cdots x_{j^1 k_{r_1}^1} \cdots x_{i^m k_1^m} \cdots x_{j^m k_{r_m}^m}|). \quad (10.3.16)$$

Here we have used the fact that $|\sum x_i| \le p^{1/2}$.

The expectation is bounded by

$$C(\log p)^{\sum_{\alpha=5}^{2r}(\alpha-4)\iota_\alpha} \le \begin{cases} C(\log p)^g \le (\log p)^r & \text{if } \; g > 0, \\ C & \text{otherwise.} \end{cases} \quad (10.3.17)$$

By (10.3.16), (10.3.17), and Lemma 10.6, we conclude that the sum of all terms in the expansion of (10.3.14) corresponding to a graph with $g > 0$ or $l > 0$ or $d < r - m/2$ will tend to 0. When $d = r - m/2$, $l = 0$, and $g = 0$, the limit of (10.3.14) will only depend on $\mathrm{E}x_{11}^2$ and $\mathrm{E}x_{11}^4$ and the powers $r_1, \cdots, r_m$.

Hence the proof of (10.3.1) is complete.

10.3.4 Proof of (c)

To verify (c), we see that because of (b) we can assume $\mathrm{E}(x_{11}) = 0$ and without loss of generality we can assume $\mathrm{E}(x_{11}^2) = 1$. We expand

$$\mathrm{E}((\sqrt{p/2}\mathbf{x}_p'\mathbf{S}_p\mathbf{x}_p - \mathrm{E}(\sqrt{p/2}\mathbf{x}_p'\mathbf{S}_p\mathbf{x}_p))(\sqrt{p/2}\mathbf{x}_p'\mathbf{S}_p^2\mathbf{x}_p - \mathrm{E}(\sqrt{p/2}\mathbf{x}_p'\mathbf{S}_p^2\mathbf{x}_p)))$$

$$\sim \left(\sum_{i \neq j} x_i^2 x_j^2\right)(2y + y^2) + \left(\sum_i x_i^4\right)(\mathrm{E}(x_{11}^4) - 1)(y + (1/2)y^2)$$

$$= (2y + y^2) + \left(\sum_i x_i^4\right)(\mathrm{E}(x_{11}^4) - 1)(y + (1/2)y^2 - (2y + y^2)). \quad (10.3.18)$$

The coefficient of $\sum_i x_i^4$ is zero if and only if $\mathrm{E}(x_{11}^4) = 3$. If $\mathrm{E}(x_{11}^4) \neq 3$, then since $\sum_i x_i^4$ can range between $1/p$ and 1, sequences $\{\mathbf{x}_p\}$ can be formed where (10.3.18) will not converge. Since we have shown, after truncation, that all mixed moments are bounded, for these sequences the ordered pair of variables in (c) will not converge in distribution. Therefore, (c) follows.

10.4 An Example of Weak Convergence

We see now that, when $\mathrm{E}(x_{11}^4) < \infty$, the condition $\mathrm{E}(x_{11}^4) = 3$, which is necessary (because of Theorem 10.2) for Property $5'$ to hold, is enough for the moments of the process $X_p(F^{\mathbf{S}_p})$ to converge weakly to those of W_a^y. Theorem 10.3 could be viewed as a display of similarity between $\{\nu_p\}$ and $\{h_p\}$ when the first, second, and fourth moments of x_{11} match those of a Gaussian. But its importance will be demonstrated in the main theorem presented in this section, which is a partial solution to the question of whether $\{\nu_p\}$ satisfies Property $5'$.

Theorem 10.7. *Assume* x_{11} *is symmetric (that is, symmetrically distributed about* 0*) and* $\mathrm{E}(x_{11}^4) < \infty$*. Then, when* $\mathbf{x}_p = (\pm\frac{1}{\sqrt{p}}, \pm\frac{1}{\sqrt{p}}, \cdots, \pm\frac{1}{\sqrt{p}})'$, $\mathbf{O}_p$ *is* ν*-distributed, and* X_p *is defined as in the equality of (10.1.2), then the limit of (10.1.2) holds.*

From the theorem, one can easily argue other choices of $\mathbf{x}_p$ for which the limit (10.1.2) holds, namely vectors close enough to those in the theorem so that the resulting X_p approaches in the Skorohod metric random functions satisfying (10.1.2). It will become apparent that the techniques used in the proof of Theorem 10.7 cannot easily be extended to $\mathbf{x}_p$ having more variability in the magnitude of its components, while the symmetry requirement may be weakened with a deeper analysis. At present, the possibility exists that only for the x_{11} mean-zero Gaussian will (10.1.2) be satisfied for all $\{\mathbf{x}_p\}$.

Theorem 10.7 adds another possible way of classifying the distribution of $\mathbf{O}_p$ as to its closeness to Haar measure. The eigenvectors of $\mathbf{S}_p$ with x_{11} symmetric and fourth moment finite display a certain amount of uniform behavior, and $\mathbf{O}_p$ can possibly be even more closely related to Haar measure if $\mathrm{E}(v_{11}^4) = 3$, due to Theorem 10.3.

For the proof of Theorem 10.7, we first recall in the proof of Theorem 10.2 that it is shown that (10.1.2), (10.2.1), and (10.2.2) imply (10.2.4). The proof of Theorem 10.7 verifies the truth of the implication in the other direction and then the truth of (10.2.4). The proof of Theorem 10.3 will be modified to show (10.3.1) still holds for the $\mathbf{x}_p$'s and x_{11} assumed in Theorem 10.7 and without a condition on the fourth moment of x_{11} other than its being finite. It will be seen that (10.3.1) yields uniqueness of weakly converging subsequences whose limits are continuous functions. With the assumptions made on $\mathbf{x}_p$ and x_{11}, tightness of $\{X_p(F^{\mathbf{S}_p})\}$ and the continuity of weakly convergent subsequences can be proven. This is the main issue for whether (10.1.2) holds more generally, due to Theorem 10.3 and parts of the proof that hold in a general setting.

The proof will be carried out in the next three subsections. Subsection 10.4.1 presents a formal description of $\mathbf{O}_p$ to account for the ambiguities mentioned at the beginning, followed by a result that converts the problem to one of showing weak convergence of $X_p(F^{\mathbf{S}_p})$ on $D[0,\infty)$, the space of rcll functions on $[0,\infty)$. Subsection 10.4.2 contains results on random elements in $D[0,b]$ for any $b > 0$ that are extensions of certain criteria for weak convergence given in Billingsley [57]. In Subsection 10.4.3, the proof is completed by showing the conditions in Subsection 10.4.2 are met. Some of the results will be stated more generally than presently needed to render them applicable for future use.

Throughout the remainder of this section, we let F_p denote $F^{\mathbf{S}_p}$.

10.4.1 Converting to $D[0,\infty)$

Let us first give a more detailed description of the distribution of $\mathbf{O}_p$ that will lead us to a concrete construction of $\mathbf{y}_p \equiv \mathbf{O}_p'\mathbf{x}_p$. For an eigenvalue λ of $\mathbf{S}_p$ with multiplicity r, we assume the corresponding r columns of $\mathbf{O}_p$ to be generated uniformly; that is, its distribution is the same as $\mathbf{O}_{p,r}\mathbf{O}_r$, where

$\mathbf{O}_{p,r}$ is $p \times r$ containing r orthonormal columns from the eigenspace of λ, and $\mathbf{O}_r \in \mathscr{O}_r$ is Haar-distributed, independent of $\mathbf{S}_p$. The $\mathbf{O}_r$'s corresponding to distinct eigenvalues are also assumed to be independent. Thus we have a natural way of constructing the random orthogonal matrix of eigenvectors of $\mathbf{S}_p$, resulting in a unique measure ν_p on $\mathscr{O}_p$.

The coordinates of $\mathbf{y}_p$ corresponding to λ are then of the form

$$(\mathbf{O}_{p,r}\mathbf{O}_r)'\mathbf{x}_p = \mathbf{O}_r'\mathbf{O}_{p,r}'\mathbf{x}_p = \|\mathbf{O}_{p,r}'\mathbf{x}_p\|\mathbf{w}_r,$$

where $\mathbf{w}_r$ is uniformly distributed on the unit sphere in $\mathbb{R}^r$. We will use the fact that the distribution of $\mathbf{w}_r$ is the same as that of a normalized vector of iid mean-zero Gaussian components. Notice that $\|\mathbf{O}_{p,r}'\mathbf{x}_p\|$ is the length of the projection of $\mathbf{x}_p$ on the eigenspace of λ. Thus, $\mathbf{y}_p$ can be represented as follows.

Enlarge the sample space defining $\mathbf{S}_p$ to allow the construction of $z_1, z_2, \ldots, z_n$, iid $N(0,1)$ random variables independent of $\mathbf{S}_p$. For a given $\mathbf{S}_p$, let $\lambda_{(1)} < \lambda_{(2)} < \cdots < \lambda_{(t)}$ be the t distinct eigenvalues with multiplicities $m_1, m_2, \ldots, m_t$. For $i = 1, 2, \ldots, t$, let a_i be the length of the projection of $\mathbf{x}_p$ on the eigenspace of $\lambda_{(i)}$. Define $m_0 = 0$. Then, for each i, we define the coordinates

$$(y_{m_1+\cdots+m_{i-1}+1}, y_{m_1+\cdots+m_{i-1}+2}, \ldots, y_{m_1+\cdots+m_i})'$$

of $\mathbf{y}_p$ to be the respective coordinates of

$$a_i \cdot \frac{(z_{m_1+\cdots+m_{i-1}+1}, z_{m_1+\cdots+m_{i-1}+2}, \ldots, z_{m_1+\cdots+m_i})'}{\sqrt{\sum_{k=1}^{m_i} z^2_{m_1+\cdots+m_{i-1}+k}}}. \tag{10.4.1}$$

We are now in a position to prove the following theorem

Theorem 10.8. *If* $X_p(F_p) \xrightarrow{\mathscr{D}} W_x^y$ *in* $D[0,\infty)$, $F_p \xrightarrow{i.p.} F_y$, *and* $\lambda_{\max} \xrightarrow{i.p.} (1+\sqrt{y})^2$, *then we have* $X_p \xrightarrow{\mathscr{D}} W_0$.

Proof. By the extended Skorohod theorem (see the footnote on page 68), we may assume that convergence of F_p and $\lambda_{\max}$ is a.s. We will continue to rely on basic results in Billingsley [57] showing weak convergence of random elements of a metric space (most notably Theorems 4.1 and 4.4 and Corollary 1 to Theorem 5.1), in particular the results on the function spaces $D[0,1]$ and $C[0,1]$. For the topology and conditions of weak convergence in $D[0,\infty)$, see Lindvall [198]. For our purposes, the only information needed regarding $D[0,\infty)$ beyond that of Billingsley [57] is the fact that weak convergence of a sequence of random functions on $D[0,\infty)$ is equivalent to the following: for every $B > 0$, there exists a constant $b > B$ such that the sequence on $D[0,b]$ (under the natural projection) converges weakly. Let ρ denote the sup metric used on $C[0,1]$ and $D[0,1]$ (used only in the latter when limiting distributions lie in $C[0,1]$ with probability 1), that is, for $x, y \in D[0,1]$,

$$\rho(x, y) = \sup_{t \in [0,1]} |x(t) - y(t)|.$$

Similar to the proof in Theorem 10.2, we need one further general result on weak convergence, which is an extension of the material on pp. 144–145 in Billingsley [57] concerning random changes of time. Let

$$\underline{D}[0, 1] = \{x \in D[0, 1] : x \text{ is nonnegative and nondecreasing}\}.$$

Since it is a closed subset of $D[0, 1]$, we take the topology of $\underline{D}[0, 1]$ to be the Skorohod topology of $D[0, 1]$ relativized to it. The mapping

$$h : D[0, \infty) \times \underline{D}[0, 1] \longrightarrow D[0, 1]$$

defined by $h(x, \varphi) = x \circ \varphi$ is measurable (the same argument as in Billingsley [57], p. 232, except the range of the integer i in (39) is now extended to all natural numbers). It is a simple matter to show that h is continuous for each

$$(x, \varphi) \in C[0, \infty) \times (C[0, 1] \cap \underline{D}[0, 1]).$$

Therefore, we have (by Corollary 1 to Theorem 5.1 of Billingsley [57])

$$\begin{aligned} &(Y_n, \Phi_n) \xrightarrow{\mathscr{D}} (Y, \Phi) \text{ in } D[0, \infty) \times \underline{D}[0, 1] \\ &\mathrm{P}(Y \in C[0, \infty)) = \mathrm{P}\big(\Phi \in C[0, 1]\big) = 1 \\ &\Rightarrow Y_n \circ \Phi_n \xrightarrow{\mathscr{D}} Y \circ \Phi \text{ in } D[0, 1]. \end{aligned} \tag{10.4.2}$$

We can now proceed with the proof of the theorem. For $t \in [0, 1]$, let $F_p^{-1}(t)$ = largest λ_j such that $F_p(\lambda_j) \le t$ (0 for $t < F_p(0)$). We have $X_p(F_p(F_p^{-1}(t))) = X_p(t)$ (although $F_p(F_p^{-1}(t)) \ne t$) except on intervals $[m/n, (m + 1)/n)$, where $\lambda_m = \lambda_{m+1}$. Let $F_y^{-1}(t)$ be the inverse of $F_y(x)$ for $x \in \big((1 - \sqrt{y})^2, (1 + \sqrt{y})^2\big]$.

We consider first the case $y \le 1$. Let $F_y^{-1}(0) = (1 - \sqrt{y})^2$. It is straightforward to show, for all $t \in (0, 1]$, $F_p^{-1}(t) \xrightarrow{\text{a.s.}} F_y^{-1}(t)$. Let $\tilde{F}_p^{-1}(t) = \max\big((1 - \sqrt{y})^2, F_p^{-1}(t)\big)$. Then, for all $t \in [0, 1]$, $\tilde{F}_p^{-1}(t) \to F_y^{-1}(t)$, and since $\lambda_{\max} \xrightarrow{\text{a.s.}} (1 + \sqrt{y})^2$, we have $\rho(\tilde{F}_p^{-1}, F_y^{-1}) \xrightarrow{\text{a.s.}} 0$. Therefore, from (10.4.2) (and Theorem 4.4 of Billingsley [57]) we have

$$X_p\big(F_p(\tilde{F}_p^{-1})\big) \xrightarrow{\mathscr{D}} W^y_{F_y^{-1}} = W_0(F_y(F_y^{-1})) = W_0, \text{ in } D[0, 1].$$

Since $F_y(x) = 0$ for $x \in [0, (1 - \sqrt{y})^2]$, we have $X_p(F_p) \xrightarrow{\mathscr{D}} 0$ in $D[0, (1 - \sqrt{y})^2]$, which implies $X_p(F_p) \xrightarrow{i.p.} 0$ in $D[0, (1 - \sqrt{y})^2]$, and since the zero function lies in $C[0, (1 - \sqrt{y})^2]$, we conclude that

$$\sup_{x\in[0,(1-\sqrt{y})^2]} |X_p(F_p(x))| \xrightarrow{i.p.} 0.$$

We then have

$$\rho\Big(X_p\big(F_p(F_p^{-1})\big), X_p\big(F_p(\tilde{F}_p^{-1})\big)\Big) \le 2 \times \sup_{x\in[0,(1-\sqrt{y})^2]} |X_p(F_p(x))| \xrightarrow{i.p.} 0.$$

Therefore, we have (by Theorem 4.1 of Billingsley [57])

$$X_p(F_p(F_p^{-1})) \xrightarrow{\mathscr{D}} W_0 \text{ in } D[0,1].$$

Notice that if x_{11} has a density, then we would be done with this case of the proof since for $p \le n$ the eigenvalues would be distinct with probability 1, so that $X_p(F_p(F_p^{-1})) = X_p$ almost surely. However, for more general x_{11}, the multiplicities of the eigenvalues need to be accounted for.

For each $\mathbf{S}_p$, let $\lambda_{(1)} < \lambda_{(2)} < \cdots < \lambda_{(\nu)}$, $(m_1, m_2, \ldots, m_\nu)$, and $(a_1, a_2, \ldots, a_\nu)$ be defined above (10.4.1). We have from (10.4.1) that

$$\rho(X_p, X_p(F_p(F_p^{-1}))) = \max_{\substack{1\le i\le \nu \\ 1\le j\le m_i}} \sqrt{\frac{p}{2}} \left| \frac{\sum_{\ell=1}^{j} z^2_{m_1+\cdots+m_{i-1}+\ell}}{\sum_{k=1}^{m_i} z^2_{m_1+\cdots+m_{i-1}+k}} - \frac{j}{p} \right|. \quad (10.4.3)$$

The measurable function h on $D[0,1]$ defined by

$$h(x) = \rho(x(\cdot), x(\cdot - 0))$$

is continuous on $C[0,1]$ (note that $h(x) = \lim_{\delta\downarrow 0} w(x,\delta)$, where $w(x,\delta)$ is the modulus of continuity of x) and is identically zero on $C[0,1]$. Therefore (using Corollary 1 to Theorem 5.1 of Billingsley [57]) $h\big(X_p(F_p(F_p^{-1}(\cdot)))\big) \xrightarrow{\mathscr{D}} 0$, which is equivalent to

$$\max_{1\le i\le\nu} \sqrt{\frac{p}{2}} \left| a_i^2 - \frac{m_i}{p} \right| \xrightarrow{i.p.} 0. \quad (10.4.4)$$

For each $i \le \nu$ and $j \le m_i$, we have

$$\sqrt{\frac{p}{2}} \left(a_i^2 \frac{\sum_{\ell=1}^{j} z^2_{m_1+\cdots+m_{i-1}+\ell}}{\sum_{k=1}^{m_i} z^2_{m_1+\cdots+m_{i-1}+k}} - \frac{j}{p} \right)$$

$$= \sqrt{\frac{p}{2}} \left(a_i^2 - \frac{m_i}{p} \right) \frac{\sum_{\ell=1}^{j} z^2_{m_1+\cdots+m_{i-1}+\ell}}{\sum_{k=1}^{m_i} z^2_{m_1+\cdots+m_{i-1}+k}} \quad (a)$$

$$+ \sqrt{\frac{p}{2}} \frac{m_i}{p} \left(a_i^2 \frac{\sum_{\ell=1}^{j} z^2_{m_1+\cdots+m_{i-1}+\ell}}{\sum_{k=1}^{m_i} z^2_{m_1+\cdots+m_{i-1}+k}} - \frac{j}{m_i} \right). \quad (b)$$

From (10.4.4), we have that the maximum of the absolute value of (a) over $1 \le i \le \nu$ converges in probability to zero. For the maximum of (b), we see that the ratio of chi-square random variables is beta-distributed with parameters $p = j/2$, $q = (m_i - j)/2$. Such a random variable with $p = r/2$, $q = (m - r)/2$ has mean r/m and fourth central moment bounded by Cr^2/m^4, where C does not depend on r and m. Let $b_{m_i,j}$ represent the expression in parentheses in (b). Let $\epsilon > 0$ be arbitrary. We use Theorem 12.2 of Billingsley after making the following associations: $S_j = \sqrt{m_i} b_{m_i,j}$, $m = m_i$, $u_\ell = \sqrt{C}/m_i$, $\gamma = 4$, $\alpha = 2$, and $\lambda = \epsilon\sqrt{2p/m_i}$. We then have the existence of $C' > 0$ for which

$$\mathrm{P}\left(\max_{1 \le j \le m_i} \left|\sqrt{\frac{p}{2}} \frac{m_i}{p} b_{m_i,j}\right| > \epsilon \;\middle|\mathbf{S}_p\right) \le \frac{C' m_i^2}{4p^2\epsilon^4}.$$

By Boole's inequality, we have

$$\mathrm{P}\left(\max_{\substack{1 \le i \le \nu \\ 1 \le j \le m_i}} \left|\sqrt{\frac{p}{2}} \frac{m_i}{p} b_{m_i,j}\right| > \epsilon \;\middle|\mathbf{S}_p\right) \le \frac{C'}{4\epsilon^4} \max_{1 \le i \le \nu} \frac{m_i}{p}.$$

Therefore

$$\mathrm{P}\left(\max_{\substack{1 \le i \le \nu \\ 1 \le j \le m_i}} \left|\sqrt{\frac{p}{2}} \frac{m_i}{p} b_{m_i,j}\right| > \epsilon\right) \le \frac{C'}{4\epsilon^4} \mathrm{E}\left(\max_{1 \le i \le \nu} \frac{m_i}{p}\right). \tag{10.4.5}$$

Because F_y is continuous on $(-\infty, \infty)$, we have $F_p(x) \xrightarrow{\text{a.s.}} F_y(x) \Rightarrow \sup_{x \in [0,\infty)} |F_p(x) - F_y(x)| \xrightarrow{\text{a.s.}} 0 \Rightarrow \sup_{x \in [0,\infty)} |F_p(x) - F_p(x-0)| \xrightarrow{\text{a.s.}} 0$, which is equivalent to $\max_{1 \le i \le \nu} m_i/p \xrightarrow{\text{a.s.}} 0$. Therefore, by the dominated convergence theorem, we have the LHS of (10.4.5) $\to 0$. We therefore have (10.4.3) $\xrightarrow{i.p.} 0$, and we conclude (again from Theorem 4.1 of Billingsley [57]) that $X_p \xrightarrow{\mathscr{D}} W_0$ in $D[0,1]$.

For $y > 1$, we assume p is sufficiently large that $p/n > 1$. Then $F_p(0) = m_1/p \ge 1 - (n/p) > 0$. For $t \in [0, 1 - (1/y)]$, define $F_y^{-1}(t) = (1 - \sqrt{y})^2$. For $t \in (1 - (1/y), 1]$, we have $F_p^{-1}(t) \xrightarrow{\text{a.s.}} F_y^{-1}(t)$. Define as before $\tilde{F}_p^{-1}(t) = \max((1-\sqrt{y})^2, F_p^{-1}(t))$. Again, $\rho(\tilde{F}_p^{-1}, F_y^{-1}) \xrightarrow{\text{a.s.}} 0$, and from (10.4.2) (and Theorem 4.4 of Billingsley) we have

$$X_p(F_p(\tilde{F}_p^{-1})) \xrightarrow{\mathscr{D}} W_0(F_y(F_y^{-1})) = \begin{cases} W_0(1 - (1/y)), & \text{for } t \in [0, 1-(1/y)], \\ W_0(t), & \text{for } t \in [1 - (1/y), 1], \end{cases}$$

in $D[0,1]$.

Since the mapping h defined on $D[0,b]$ by $h(x) = \sup_{t \in [0,b]} |x(t) - x(b)|$ is continuous for all $x \in C[0,b]$, we have by Theorem 5.1 of Billingsley [57]

$$\rho(X_p(F_p(F_p^{-1})), X_p(F_p(\tilde{F}_p^{-1})))$$

$$= \sup_{x\in[0,(1-\sqrt{y})^2]} |X_p(F_p(x)\big) - X_p(F_p((1-\sqrt{y})^2))|$$

$$\xrightarrow{\mathscr{D}} \sup_{x\in[0,(1-\sqrt{y})^2]} |W_0(F_y(x)) - W_0(F_y((1-\sqrt{y})^2))| = 0,$$

which implies

$$\rho(X_p\big(F_p(F_p^{-1})\big), X_p(F_p(\tilde{F}_p^{-1}))) \xrightarrow{i.p.} 0.$$

Therefore (by Theorem 4.1 of Billingsley [57])

$$X_p(F_p(F_p^{-1})) \xrightarrow{\mathscr{D}} W_0(F_y(F_y^{-1})).$$

For $t < F_p(0) + \frac{1}{p}$,

$$X_p(t) = \sqrt{\frac{p}{2}}\left(a_1^2 \, \frac{\sum_{i=1}^{[pt]} z_i^2}{\sum_{\ell=1}^{pF_p(0)} z_\ell^2} - \frac{[pt]}{p}\right)$$

$$= \frac{a_1^2}{\sqrt{F_p(0)}}\sqrt{\frac{pF_p(0)}{2}}\left(\frac{\sum_{i=1}^{[pt]} z_i^2}{\sum_{\ell=1}^{pF_p(0)} z_\ell^2} - \frac{[pt]}{pF_p(0)}\right) + \frac{[pt]}{pF_p(0)}\sqrt{\frac{p}{2}}(a_1^2 - F_p(0)).$$

Notice that $\sqrt{\frac{p}{2}}(a_1^2 - F_p(0)) = X_p(F_p(0))$.

For $t \in [0,1]$, let $\varphi_p(t) = \min(t/F_p(0), 1)$, $\varphi(t) = \min(t/(1-(1/y)), 1)$, and

$$Y_p(t) = \sqrt{\frac{p}{2}}\left(\frac{\sum_{i=1}^{[pt]} z_i^2}{\sum_{\ell=1}^{p} z_\ell^2} - \frac{[pt]}{p}\right).$$

Then $\varphi_n \xrightarrow{i.p.} \varphi$ in $D_0 \equiv \{x \in \underline{D}[0,1] : x(1) \le 1\}$ (see Billingsley [57], p. 144), and for $t < F_p(0) + \frac{1}{p}$

$$Y_{pF_p(0)}(\varphi_p(t)) = \sqrt{\frac{pF_p(0)}{2}}\left(\frac{\sum_{i=1}^{[pt]} z_i^2}{\sum_{\ell=1}^{pF_p(0)} z_\ell^2} - \frac{[pt]}{pF_p(0)}\right).$$

For all $t \in [0,1]$, let

$$H_p(t) = \frac{a_1^2}{\sqrt{F_p(0)}} Y_{pF_p(0)}(\varphi_p(t))$$

$$+X_p(F_p(0))\left(\frac{[pF_p(0)\varphi_p(t)]}{pF_p(0)} - 1\right) + X_p(F_p(F_p^{-1}(t))).$$

Then $H_p(t) = X_p(t)$ except on intervals $[m/p, (m+1)/p)$, where $0 < \lambda_m = \lambda_{m+1}$. We will show $H_p \xrightarrow{\mathscr{D}} W_0$ in $D[0,1]$.

Let $\psi_p(t) = F_p(0)t$, $\psi(t) = (1-(1/y))t$, and

$$V_p(t) = \frac{1}{\sqrt{2p}} \sum_{i=1}^{[pt]} (z_i^2 - 1).$$

Then $\psi_p \xrightarrow{i.p.} \psi$ in D_0 and

$$Y_p(t) = \frac{V_p(t) - ([pt]/p)V_p(1)}{1 + \sqrt{2/p}V_p(1)}. \tag{10.4.6}$$

Since $X_p(F_p(F_p^{-1}))$ and V_p are independent, we have (using Theorems 4.4 and 16.1 of Billingsley [57])

$$(X_p(F_p(F_p^{-1})), V_p, \varphi_p, \psi_p) \xrightarrow{\mathscr{D}} (W_0(F_y(F_y^{-1})), \overline{W}, \varphi, \psi),$$

where $\overline{W}$ is a Weiner process, independent of W_0. We immediately get (Billingsley [57], p. 145)

$$(X_p(F_p(F_p^{-1})), V_p \circ \psi_p, \varphi_p) \xrightarrow{\mathscr{D}} (W_0(F_y(F_y^{-1})), \overline{W} \circ \psi, \varphi).$$

Since $V_p(\psi_p(t)) = \sqrt{F_p(0)} V_{pF_p(0)}(t)$, we have

$$\rho(V_p \circ \psi_p, \sqrt{1-(1/y)} V_{pF_p(0)}) = \left| \sqrt{F_p(0)} - \sqrt{1-(1/y)} \right| \sup_{t \in [0,1]} |V_{pF_p(0)}(t)| \xrightarrow{i.p.} 0.$$

Therefore

$$(X_p(F_p(F_p^{-1})), V_{pF_p(0)}, \varphi_p) \xrightarrow{\mathscr{D}} \left(W_0(F_y(F_y^{-1})), \frac{1}{\sqrt{1-(1/y)}} \overline{W} \circ \psi, \varphi \right). \tag{10.4.7}$$

Notice that $\frac{1}{\sqrt{1-(1/y)}} \overline{W} \circ \psi$ is again a Weiner process, independent of W_0.

From (10.4.6), we have

$$Y_p(t) - (V_p(t) - tV_p(1)) = V_p(t) \frac{t - [pt]/p + \sqrt{2/p}(tV_p(1) - V_p(t))}{1 + \sqrt{2/p}V_p(1)}.$$

Therefore

$$\rho(Y_{pF_p(0)}(t), V_{pF_p(0)}(t) - tV_{pF_p(0)}(1)) \xrightarrow{i.p.} 0. \tag{10.4.8}$$

From (10.4.7), (10.4.1), and the fact that $\overline{W}(t) - t\overline{W}(1)$ is a Brownian bridge, it follows that

$$(X_p(F_p(F_p^{-1})), Y_{pF_p(0)}, \varphi_p) \xrightarrow{\mathscr{D}} (W_0(F_y(F_y^{-1})), \widehat{W}_0, \varphi),$$

where $\widehat{W}_0$ is another Brownian bridge, independent of W_0.

The mapping $h : D[0,1] \times D[0,1] \times D_0 \to D[0,1]$ defined by

$$h(x_1, x_2, z) = \sqrt{1-(1/y)}x_2 \circ z + x_1(0)(z-1) + x_1$$

is measurable and continuous on $C[0,1] \times C[0,1] \times (D_0 \cap C[0,1])$. Also, from (10.4.4) we have $a_1^2 \xrightarrow{i.p.} 1 - (1/y)$. Finally, it is easy to verify

$$\frac{[pF_p(0)\varphi_p]}{pF_p(0)} \xrightarrow{i.p.} \varphi \quad \text{in } D_0.$$

Therefore, we can conclude (using Theorem 4.1 and Corollary 1 of Theorem 5.1 of Billingsley [57]) that

$$H_p \xrightarrow{\mathscr{D}} \sqrt{1-(1/y)}\,\widehat{W}_0 \circ \varphi + W_0(1-(1/y))(\varphi - 1) + W_0(F_y(F_y^{-1})) \equiv H.$$

It is immediately clear that H is a mean 0 Gaussian process lying in $C[0,1]$. It is a routine matter to verify for $0 \le s \le t \le 1$ that

$$\mathrm{E}(H_s H_t) = s(1-t).$$

Therefore, H is a Brownian bridge.

We see that $\rho(X_p, H_p)$ is the same as the RHS of (10.4.3) except $i = 1$ is excluded. The arguments leading to (10.4.4) and (10.4.5) ($2 \le i \le t$) are exactly the same as before. The fact that $\max_{2 \le i \le t} m_i/p \xrightarrow{i.p.} 0$ follows from the case $y \le 1$ since the nonzero eigenvalues (including multiplicities) of AA' and $A'A$ are identical for any rectangular A. Thus

$$\rho(X_p, H_p) \xrightarrow{i.p.} 0$$

and we have X_p converging weakly to a Brownian bridge.

10.4.2 A New Condition for Weak Convergence

In this section, we establish two results on random elements of $D[0,b]$ needed for the proof of Theorem 10.7. In the following, we denote the modulus of continuity of $x \in D[0,b]$ by $w(x,\cdot)$:

$$w(x,\delta) = \sup_{|s-t|<\delta} |x(s) - x(t)|, \quad \delta \in (0,b].$$

To simplify the analysis, we assume, for now, $b = 1$.

Theorem 10.9. *Let* $\{X_p\}$ *be a sequence of random elements of* $D[0,1]$ *whose probability measures satisfy the assumptions of Theorem 15.5 of Billingsley [57]; that is,* $\{X_p(0)\}$ *is tight, and for every positive* ϵ *and* η*, there exists a* $\delta \in (0,1)$ *and an integer* n_0*, such that, for all* $n > n_0$, $\mathrm{P}(w(X_p,\delta) \ge \epsilon) \le \eta$*. If there exists a random element* X *with* $\mathrm{P}(X \in C[0,1]) = 1$ *and such that*

$$\left\{\int_0^1 t^r X_p(t)dt\right\}_{r=0}^{\infty} \xrightarrow{\mathscr{D}} \left\{\int_0^1 t^r X(t)dt\right\}_{r=0}^{\infty} \quad as\ n \to \infty \tag{10.4.9}$$

(($\mathscr{D}$) in (10.4.9) denoting convergence in distribution on $\mathbb{R}^\infty$), then $X_p \xrightarrow{\mathscr{D}} X$.

Proof. Note that the mappings

$$x \to \int_0^1 t^r x(t)dt$$

are continuous in $D[0,1]$. Therefore, by Theorems 5.1 and 15.5 of Billingsley [57], $X_p \xrightarrow{\mathscr{D}} X$ will follow if we can show that the distribution of X is uniquely determined by the distribution of

$$\left\{\int_0^1 t^r X(t)dt\right\}_{r=0}^{\infty}. \tag{10.4.10}$$

Since the finite-dimensional distributions of X uniquely determine the distribution of X, it suffices to show for any integer m and numbers a_i, t_i, $i = 0, 1, \ldots, m$ with $0 = t_0 < t_1 < \cdots < t_m = 1$ that the distribution of

$$\sum_{i=0}^{m} a_i X(t_i) \tag{10.4.11}$$

is uniquely determined by the distribution of (10.4.10).

Let $\{f_n\}, f$ be uniformly bounded measurable functions on [0,1] such that $f_n \to f$ pointwise as $n \to \infty$. Using the dominated convergence theorem, we have

$$\int_0^1 f_n(t)X(t)dt \to \int_0^1 f(t)X(t)dt \quad \text{as } n \to \infty. \tag{10.4.12}$$

Let $\epsilon > 0$ be any number less than half the minimum distance between the t_i's. Notice that for the indicator function $I([a,b])$ we have the sequence of continuous "ramp" functions $\{R_n(t)\}$ with

$$R_n(t) = \begin{cases} 1 & t \in [a,b], \\ 0 & t \in [a-1/n, b+1/n]^c \end{cases}$$

and linear on each of the sets $[a-1/n, a]$, $[b, b+1/n]$, satisfying $R_n \downarrow I([a,b])$ as $n \to \infty$. Notice also that we can approximate any ramp function uniformly on [0,1] by polynomials. Therefore, using (10.4.12) for polynomials appropriately chosen, we find that the distribution of

$$\sum_{i=0}^{m-1} a_i \int_{t_i}^{t_i+\epsilon} X(t)dt + a_m \int_{1-\epsilon}^{1} X(t)dt \tag{10.4.13}$$

is uniquely determined by the distribution of (10.4.10).

Dividing (10.4.13) by ϵ and letting $\epsilon \to 0$, we get a.s. convergence to (10.4.11) (since $X \in C[0,1]$ with probability 1) and we are done.

Theorem 10.10. *Let X be a random element of $D[0,1]$. Suppose there exist constants $B > 0$, $\gamma \geq 0$, $\alpha > 1$, and a random nondecreasing, right-continuous function $F : [0,1] \to [0,B]$ such that, for all $0 \leq t_1 \leq t_2 \leq 1$ and $\lambda > 0$,*

$$\mathrm{P}(|X(t_2) - X(t_1)| \geq \lambda) \leq \frac{1}{\lambda^\gamma}\mathrm{E}[(F(t_2) - F(t_1))^\alpha]. \tag{10.4.14}$$

Then, for every $\epsilon > 0$ and δ an inverse of a positive integer, we have

$$\mathrm{P}(w(X,\delta) \geq 3\epsilon) \leq \frac{KB}{\epsilon^\gamma}\mathrm{E}\left[\max_{j<\delta^{-1}}(F((j+1)\delta) - F(j\delta))^{\alpha-1}\right], \tag{10.4.15}$$

where j ranges on positive integers and K depends only on γ and α.

The theorem is proven by modifying the proofs of the first three theorems in Section 12 of Billingsley [57]. It is essentially an extension of part of a result contained in Theorem 12.3 of Billingsley [57]. The original arguments, for the most part, remain unchanged. We will indicate only the specific changes and refer the reader to Billingsley [57] for details. The extensions of two of the theorems in Billingsley [57] will be given below as lemmas. However, some definitions must first be given.

Let $\xi_1, \ldots, \xi_m$ be random variables, and $S_k = \xi_1 + \cdots + \xi_k$ $(S_0 = 0)$. Let

$$M_m = \max_{0 \leq k \leq m} |S_k|,$$

$$M'_m = \max_{0 \leq k \leq m} \min(|S_k|, |S_m - S_k|).$$

Lemma 10.11. (Extension to Theorem 12.1 of Billingsley [57]). *Suppose $u_1, \ldots, u_m$ are nonnegative random variables such that*

$$\mathrm{P}(|S_j - S_i| \geq \lambda, |S_k - S_j| \geq \lambda) \leq \frac{1}{\lambda^{2\gamma}}\mathrm{E}\left[\left(\sum_{i<\ell\leq k} u_\ell\right)^{2\alpha}\right] < \infty,$$

$$0 \leq i \leq j \leq k \leq m$$

for some $\alpha > \frac{1}{2}$, $\gamma \geq 0$, and for all $\lambda > 0$. Then, for all $\lambda > 0$,

$$\mathrm{P}(M'_m \geq \lambda) \leq \frac{K}{\lambda^{2\gamma}}\mathrm{E}[(u_1 + \cdots + u_m)^{2\alpha}], \tag{10.4.16}$$

where $K = K_{\gamma,\alpha}$ depends only on γ and α.

Proof. We follow Billingsley [57], p. 91. The constant K is chosen in the same way, and the proof proceeds by induction on m. The arguments for $m = 1$

and 2 are the same except that for the latter $(u_1 + u_2)^{2\alpha}$ is replaced by $\mathrm{E}(u_1 + u_2)^{2\alpha}$. Assuming (10.4.16) is true for all integers less than m, we find an integer h, $1 \le h \le m$, such that

$$\frac{\mathrm{E}[(u_1 + \cdots + u_{h-1})^{2\alpha}]}{\mathrm{E}[(u_1 + \cdots + u_m)^{2\alpha}]} \le \frac{1}{2} \le \frac{\mathrm{E}[(u_1 + \cdots + u_h)^{2\alpha}]}{\mathrm{E}[(u_1 + \cdots + u_m)^{2\alpha}]},$$

the sum on the left-hand side being 0 if $h = 1$.

Since $2\alpha > 1$, we have for all nonnegative x and y

$$x^{2\alpha} + y^{2\alpha} \le (x + y)^{2\alpha}.$$

We then have

$$\begin{aligned}\mathrm{E}[(u_{h+1} + \cdots + u_m)^{2\alpha}] &\le \mathrm{E}[(u_1 + \cdots + u_m)^{2\alpha}] - \mathrm{E}[(u_1 + \cdots + u_h)^{2\alpha}] \\ &\le \mathrm{E}\left[(u_1 + \cdots + u_m)^{2\alpha}\right]\left(1 - \frac{1}{2}\right) = \frac{1}{2}\mathrm{E}[(u_1 + \cdots + u_m)^{2\alpha}].\end{aligned}$$

Therefore, defining U_1, U_2, D_1, D_2 as in Billingsley [57], we get the same inequalities as in (12.30)–(12.33) of Billingsley [57], p. 92, with $u^{2\alpha}$ replaced by $\mathrm{E}[(u_1 + \cdots + u_m)^{2\alpha}]$. The rest of the proof follows exactly.

Lemma 10.12. (Extension to Theorem 12.2 of Billingsley [57]). *If, for random nonnegative u_ℓ, there exist $\alpha > 1$ and $\gamma \ge 0$ such that, for all $\lambda > 0$,*

$$\mathrm{P}(|S_j - S_i| \ge \lambda) \le \frac{1}{\lambda^\gamma}\mathrm{E}\left[\left(\sum_{i<\ell\le j} u_\ell\right)^{2\alpha}\right] < \infty, \quad 0 \le i \le j \le m,$$

then

$$\mathrm{P}(M_n \ge \lambda) \le \frac{K'_{\gamma,\alpha}}{\lambda^\gamma}\mathrm{E}[(u_1 + \cdots + u_m)^{2\alpha}], \qquad K'_{\gamma,\alpha} = 2^\gamma(1 + K_{\gamma/2,\alpha/2}).$$

Proof. Following Billingsley [57], we have for $0 \le i \le j \le k \le m$

$$\begin{aligned}\mathrm{P}(|S_j - S_i| \ge \lambda, |S_k - S_j| \ge \lambda) &\le \mathrm{P}^{\frac{1}{2}}(|S_j - S_i| \ge \lambda)\mathrm{P}^{\frac{1}{2}}(|S_k - S_j| \ge \lambda) \\ &\le \frac{1}{\lambda\gamma}\mathrm{E}\left[\left(\sum_{i<\ell\le k} u_\ell\right)^{2\alpha}\right],\end{aligned}$$

so Lemma 10.4.9 is satisfied with constants $\gamma/2, \alpha/2$. The rest follows exactly as in Billingsley [57], p. 94, with $(u_1 + \cdots + u_m)^\alpha$ in (12.46) and (12.47) replaced by the expected value of the same quantity.

We can now proceed with the proof of Theorem 10.10. Following the proof of Theorem 12.3 of Billingsley [57], we fix positive integers $j < \delta^{-1}$ and m and define

$$\xi_i = X\left(j\delta + \frac{i}{m}\delta\right) - X\left(j\delta + \frac{i-1}{m}\delta\right), \qquad i = 1, 2, \ldots, m.$$

The partial sums of the ξ_i's satisfy Lemma 10.12 with

$$u_i = F\left(j\delta + \frac{i}{m}\delta\right) - F\left(j\delta + \frac{i-1}{m}\delta\right).$$

Therefore

$$\mathrm{P}\left(\max_{1\le i\le m}\left|X\left(j\delta + \frac{i}{m}\delta\right) - X\left(j\delta\right)\right| \ge \epsilon\right) \le \frac{K}{\epsilon^\gamma}\mathrm{E}[(F((j+1)\delta) - F(j\delta))^\alpha]$$

with $K = K'_{\gamma,\alpha}$.

Since $X \in D[0,1]$, we have

$$\mathrm{P}\left(\sup_{j\delta\le s\le (j+1)\delta}|X(s) - X(j\delta)| > \epsilon\right)$$

$$= \mathrm{P}\left(\max_{1\le i\le m}\left|X\left(j\delta + \frac{i}{m}\delta\right) - X\left(j\delta\right)\right| > \epsilon \text{ for all } m \text{ sufficiently large }\right)$$

$$\le \liminf_m \mathrm{P}\left(\max_{1\le i\le m}\left|X\left(j\delta + \frac{i}{m}\delta\right) - X\left(j\delta\right)\right| \ge \epsilon\right)$$

$$\le \frac{K}{\epsilon^\gamma}\mathrm{E}[(F((j+1)\delta) - F(j\delta))^\alpha].$$

By considering a sequence of numbers approaching ϵ from below, we get from the continuity theorem on probability measures

$$\mathrm{P}\left(\sup_{j\delta\le s\le (j+1)\delta}|X(s) - X(j\delta)| \ge \epsilon\right) \le \frac{K}{\epsilon^\gamma}\mathrm{E}[(F((j+1)\delta) - F(j\delta))^\alpha]. \tag{10.4.17}$$

Summing both sides of (10.4.17) over all $j < \delta^{-1}$ and using the corollary to Theorem 8.3 of Billingsley [57], we get

$$\mathrm{P}(w(X,\delta) \ge 3\epsilon) \le \frac{K}{\epsilon^\gamma}\mathrm{E}\left[\sum_{j<\delta^{-1}}(F((j+1)\delta) - F(j\delta))^\alpha\right]$$

$$\le \frac{K}{\epsilon^\gamma}\mathrm{E}\left[\max_{j<\delta^{-1}}(F((j+1)\delta) - F(j\delta))^{\alpha-1}(F(1) - F(0))\right]$$

$$\le \frac{KB}{\epsilon^\gamma}\mathrm{E}\left[\max_{j<\delta^{-1}}(F((j+1)\delta) - F(j\delta))^{\alpha-1}\right],$$

and we are done.

For general $D[0,b]$, we simply replace (10.4.9) by

$$\left\{\int_0^b t^r X_p(t)dt\right\}_{r=0}^{\infty} \xrightarrow{\mathscr{D}} \left\{\int_0^b t^r X(t)dt\right\}_{r=0}^{\infty} \quad \text{as } n \to \infty \tag{10.4.18}$$

and (10.4.15) by

$$\mathrm{P}(w(X, b\delta) \geq 3\epsilon) \leq \frac{KB}{\epsilon^\gamma}\mathrm{E}\left[\max_{j<\delta^{-1}} (F(b(j+1)\delta) - F(bj\delta))^{\alpha-1}\right], \tag{10.4.19}$$

j and δ^{-1} still positive integers.

10.4.3 Completing the Proof

We finish up by verifying the conditions of Theorem 10.9.

Theorem 10.13. *Let* $\mathrm{E}(x_{11}) = 0$, $\mathrm{E}(x_{11}^2) = 1$, *and* $\mathrm{E}(x_{11}^4) < \infty$. *Suppose the sequence of vectors* $\{\mathbf{x}_p\}$, $\mathbf{x}_p = (x_{p1}, x_{p2}, \ldots, x_{pp})'$, $\|\mathbf{x}_p\| = 1$ *satisfies*

$$\sum_{i=1}^p x_{pi}^4 \to 0 \quad \textit{as } n \to \infty. \tag{10.4.20}$$

Then (10.3.1) holds.

Proof. We return to the proof of Theorem 10.3 and consider the sum of nonnegligible terms contributing to (10.3.14) in the limit. If there is an I_1-vertex consisting of four or more I_1-indices, such terms are negligible because of condition (10.4.20). Thus, all I_1-indices must be pairwise matched. Also, by Lemma 10.6, the nonnegligible terms should satisfy $d = r - m/2$. This is impossible if m is odd. Thus, the limit of the mixed moment is 0 if m is odd.

Now, let us consider the case where m is even. Due to the fact that (2) is avoided, the graph G consists of at most $m/2$ connected subgraphs. Also, a graph of r' noncoincident edges has at most $r' + m/2$ noncoincident vertices. Because there are exactly m noncoincident vertices, each of which matches with two I_1-indices, we have $d \leq r' - m/2$. Therefore $r = r'$, which implies that all noncoincident edges have multiplicity 2. We conclude that the asymptotic behavior of (10.3.14) depends only on $\mathrm{E}(x_{11}^2)$, and we are done.

Let $\mathbb{R}_+ = [0, \infty)$ and $\mathscr{B}_+$, $\mathscr{B}_+^4$ denote the Borel σ-fields on, respectively, $\mathbb{R}_+$ and $\mathbb{R}_+^4$. For any $p \times p$ symmetric, nonnegative definite matrix $\mathbf{B}_p$ and any $A \in \mathscr{B}_+$, let $\mathbf{P}^{\mathbf{B}_p}(A)$ denote the projection matrix on the subspace of $\mathbb{R}^p$ spanned by the eigenvectors of $\mathbf{B}_p$ having eigenvalues in A (the collection of projections $\{\mathbf{P}^{\mathbf{B}_p}((-\infty, a]) : a \in \mathbb{R}\}$ is usually referred to as the spectral family of $\mathbf{B}_p$). We have $\mathrm{tr}\mathbf{P}^{\mathbf{B}_p}(A)$ equal to the number of eigenvalues of $\mathbf{B}_p$ contained in A. If $\mathbf{B}_p$ is random, then it is straightforward to verify the following facts.

Fact 1. For every $\mathbf{x}_p \in \mathbb{R}^p$, $\|\mathbf{x}_p\| = 1$, $\mathbf{x}_p' \mathbf{P}^{\mathbf{B}_p}(\cdot)\mathbf{x}_p$ is a random probability measure on $\mathbb{R}_+$ placing mass on the eigenvalues of $\mathbf{B}_p$.

Fact 2. For any four entries $P_{i_1 j_1}^{\mathbf{B}_p}(\cdot)$, $P_{i_2 j_2}^{\mathbf{B}_p}(\cdot)$, $P_{i_3 j_3}^{\mathbf{B}_p}(\cdot)$, $P_{i_4 j_4}^{\mathbf{B}_p}(\cdot)$ of $\mathbf{P}^{\mathbf{B}_p}(\cdot)$, the function defined on rectangles $A_1 \times A_2 \times A_3 \times A_4 \in \mathscr{B}_+^4$ by

$$\mathrm{E}\big(P_{i_1 j_1}^{\mathbf{B}_p}(A_1) P_{i_2 j_2}^{\mathbf{B}_p}(A_2) P_{i_3 j_3}^{\mathbf{B}_p}(A_3) P_{i_4 j_4}^{\mathbf{B}_p}(A_4)\big) \tag{10.4.21}$$

generates a signed measure $m_p^{\mathbf{B}_p} = m_p^{\mathbf{B}_p,(i_1,j_1,\ldots,i_4,j_4)}$ on $(\mathbb{R}_+^4, \mathscr{B}_+^4)$ such that $|m_n^{\mathbf{B}_p}(A)| \le 1$ for every $A \in \mathscr{B}_+^4$.

When $\mathbf{B}_p = \mathbf{S}_p$ we also have the following facts.

Fact 3. For any $A \in \mathscr{B}_+$, the distribution of $\mathbf{P}^{\mathbf{S}_p}(A)$ is invariant under permutation transformations; that is, $\mathbf{P}^{\mathbf{S}_p}(A) \overset{\mathscr{D}}{=} \mathbf{O}_p \mathbf{P}^{\mathbf{S}_p}(A)\mathbf{O}_p'$ for any permutation matrix $\mathbf{O}_p$ $\big($using the fact that $\mathbf{P}^{\mathbf{B}_p}(\cdot)$ is uniquely determined by $\{\mathbf{B}_p^r\}_{r=1}^\infty$ along with $\mathbf{O}_p\mathbf{P}^{\mathbf{B}_p}(\cdot)\mathbf{O}_p' = P^{\mathbf{O}_p\mathbf{B}_p\mathbf{O}_p'}(\cdot)$ and $\{\mathbf{S}_p^r\}_{r=1}^\infty \overset{\mathscr{D}}{=} \{(\mathbf{O}_p\mathbf{S}_p\mathbf{O}_p')^r\}_{r=1}^\infty\big)$.

Fact 4. For $0 \le x_1 \le x_2$,

$$\frac{1}{p}\mathrm{tr}\mathbf{P}^{\mathbf{S}_p}([0,x_1]) = F_p(x_1),$$

$$X_p(F_p(x_1)) = \sqrt{\frac{n}{2}}\Big(\mathbf{x}_p'\mathbf{P}^{\mathbf{S}_p}([0,x_1])\mathbf{x}_p - \frac{1}{p}\mathrm{tr}(\mathbf{P}^{\mathbf{S}_p}([0,x_1]))\Big),$$

and

$$X_p(F_p(x_2)) - X_p(F_p(x_1)) = \sqrt{\frac{p}{2}}\Big(\mathbf{x}_p'\mathbf{P}^{\mathbf{S}_p}((x_1,x_2])\mathbf{x}_p - \frac{1}{p}\mathrm{tr}(\mathbf{P}^{\mathbf{S}_p}((x_1,x_2]))\Big).$$

Lemma 10.14. *Assume x_{11} is symmetric. If one of the indices $i_1, j_1, \cdots, i_4, j_4$ appears an odd number of times, then $m_p^{\mathbf{S}_p} \equiv 0$.*

Proof. Let $\mathbf{O}_i$ be the diagonal matrix with diagonal entries 1 except the i-th, which is -1. Then, by assumption we have $\mathbf{S}_p \overset{\mathscr{D}}{=} \mathbf{O}_i\mathbf{S}_p\mathbf{O}_i$ and hence $m_p^{\mathbf{S}_p} = m_p^{\mathbf{O}_i\mathbf{S}_p\mathbf{O}_i}$. On the other hand, similar to Fact 3, one can show that $\mathbf{P}^{\mathbf{O}_i\mathbf{S}_p\mathbf{O}_i} = \mathbf{O}_i\mathbf{P}^{\mathbf{S}_p}\mathbf{O}_i$ and hence $m_p^{\mathbf{O}_i\mathbf{S}_p\mathbf{O}_i} = (-1)^\alpha m_p^{\mathbf{S}_p}$, where α is the frequency of i among $\{i_1, j_1, \cdots, i_4, j_4\}$. Thus the lemma follows.

Theorem 10.15. *Assume x_{11} is symmetric and $\mathbf{x}_p = (\pm 1/\sqrt{p}, \pm 1/\sqrt{p}, \cdots, \pm 1/\sqrt{p})'$. Let $G_p(x) = 4F_p(x)$. Then*

$$\mathrm{E}\big((X_p(F_p(0)))^4\big) \le \mathrm{E}\big((G_p(0))^2\big), \tag{10.4.22}$$

and for any $0 \le x_1 \le x_2$

$$\mathrm{E}\big((X_p(F_p(x_2)) - X_p(F_p(x_1)))^4\big) \le \mathrm{E}\big((G_p(x_2) - G_p(x_1))^2\big). \tag{10.4.23}$$

Proof. With $A = \{0\}$ in (10.4.22), and $A = (x_1, x_2]$ in (10.4.23), we use Fact 4 to find the LHS of (10.4.22) and (10.4.23) equal to

$$\frac{1}{4p^2}\mathrm{E}\left(\sum_{i\neq j}\gamma_{ij}P_{ij}^{\mathbf{S}_p}(A)\right)^4, \tag{10.4.24}$$

where $\gamma_{ij} = \mathrm{sgn}((\mathbf{x}_p)_i(\mathbf{x}_p)_j)$. For the remainder of the argument, we simplify the notation by suppressing the dependence of the projection matrix on $\mathbf{S}_p$ and A. Upon expanding (10.4.24), we use Fact 3 to combine identically distributed factors, and Lemma 10.14 to arrive at

$$\begin{aligned}(10.4.24) = \frac{(p-1)}{p}\Big(&12(p-2)\mathrm{E}(P_{12}^2P_{13}^2) + 3(p-2)(p-3)\mathrm{E}(P_{12}^2P_{34}^2)\\ &+12(p-2)(p-3)\mathrm{E}(P_{12}P_{23}P_{34}P_{14}) + 2\mathrm{E}(P_{12}^4)\Big).\end{aligned} \tag{10.4.25}$$

We can write the second and third expected values in (10.4.25) in terms of the first expected value and expected values involving P_{11}, P_{22}, and P_{12} by making further use of Fact 3 and the fact that $\mathbf{P}$ is a projection matrix (i.e., $\mathbf{P}^2 = \mathbf{P}$). For example, we take the expected value of both sides of the identity

$$P_{12}P_{23}\left(\sum_{j\geq 4}P_{3j}P_{1j} + P_{31}P_{11} + P_{32}P_{12} + P_{33}P_{13}\right) = P_{12}P_{23}P_{13}$$

and get

$$(p-3)\mathrm{E}(P_{12}P_{23}P_{34}P_{14}) + 2\mathrm{E}(P_{11}P_{12}P_{23}P_{31}) + \mathrm{E}(P_{12}^2P_{13}^2) = \mathrm{E}(P_{12}P_{23}P_{13}).$$

Proceeding in the same way, we find

$$(p-2)\mathrm{E}(P_{11}P_{12}P_{23}P_{31}) + \mathrm{E}(P_{11}^2P_{12}^2) + \mathrm{E}(P_{11}P_{22}P_{12}^2) = \mathrm{E}(P_{11}P_{12}^2)$$

and

$$(p-2)\mathrm{E}(P_{12}P_{23}P_{13}) + 2\mathrm{E}(P_{11}P_{12}^2) = \mathrm{E}(P_{12}^2).$$

Therefore,

$$\begin{aligned}&(p-2)(p-3)\mathrm{E}(P_{12}P_{23}P_{34}P_{14})\\ &= \mathrm{E}(P_{12}^2) + 2\mathrm{E}(P_{11}P_{22}P_{12}^2) + 2\mathrm{E}(P_{11}^2P_{12}^2) - (p-2)\mathrm{E}(P_{12}^2P_{13}^2) - 4\mathrm{E}(P_{11}P_{12}^2).\end{aligned}$$

Since $P_{11} \geq \max(P_{11}P_{22}, P_{11}^2)$ and $P_{12}^2 \leq P_{11}P_{22}$ (since P is nonnegative definite), we have

$$(p-2)(p-3)\mathrm{E}(P_{12}P_{23}P_{34}P_{14}) \leq \mathrm{E}(P_{11}P_{22}) - (p-2)\mathrm{E}(P_{12}^2P_{13}^2).$$

Similar arguments will yield

$$(p-3)\mathrm{E}(P_{12}^2P_{34}^2)+2\mathrm{E}(P_{12}^2P_{13}^2)+\mathrm{E}(P_{12}^2P_{33}^2)=\mathrm{E}(P_{12}^2P_{33}),$$

$$(p-2)\mathrm{E}(P_{12}^2P_{13}^2)+\mathrm{E}({}^4_{12})+\mathrm{E}(P_{11}^2P_{12}^2)=\mathrm{E}(P_{11}P_{12}^2).$$

After multiplying the first equation by $n-2$ and adding it to the second, we get

$$\begin{aligned}&(p-2)(p-3)\mathrm{E}(P_{12}^2P_{34}^2)+3(p-2)\mathrm{E}(P_{12}^2P_{13}^2)\\ &=(p-2)\mathrm{E}(P_{12}^2P_{33})-(p-2)\mathrm{E}(P_{12}^2P_{33}^2)+\mathrm{E}(P_{11}P_{12}^2)-\mathrm{E}(P_{11}^2P_{12}^2)-\mathrm{E}(P_{12}^4)\\ &=\mathrm{E}(P_{11}P_{22})+\mathrm{E}(P_{11}^2P_{22}^2)-2\mathrm{E}(P_{11}P_{22}^2)-\mathrm{E}(P_{12}^4)\\ &\le \mathrm{E}(P_{11}P_{22})-\mathrm{E}(P_{12}^4).\end{aligned}$$

Combining the expressions above, we obtain

$$(10.4.24)\ \le\ 15\frac{(p-1)}{p}\mathrm{E}(P_{11}P_{22}).$$

Therefore, using Facts 3 and 4, we get

$$(10.4.24)\le\frac{15}{p^2}\mathrm{E}\left(\sum_{i\neq j}P_{ii}P_{jj}\right)\le\mathrm{E}\left(\left(\frac{4}{p}\mathrm{tr}P\right)^2\right)$$

$$=\begin{cases}\mathrm{E}\big((G_p(0))^2\big) & \text{for } A=\{0\},\\ \mathrm{E}\big((G_p(x_2)-G_p(x_1))^2\big) & \text{for } A=(x_1,x_2],\end{cases}$$

and we are done.

We can now complete the proof of Theorem 10.7. Choose any $b>(1+\sqrt{y})^2$. As in the proof of Theorem 10.2, we may assume (10.2.1) and, by Theorem 10.13, (10.3.1), which implies

$$\left\{\int_0^b x^rX_p(F_p(x))dx\right\}_{r=0}^{\infty}\xrightarrow{\mathscr{D}}\left\{\int_0^b x^rW_x^ydx\right\}_{r=0}^{\infty}\qquad\text{as } n\to\infty,$$

so that (10.4.18) is satisfied. By Theorems 10.10 and 10.15, we have (10.4.19) with $X=X_p(F_p)$, $F=4F_p$, $B=4$, $\gamma=4$, and $\alpha=2$. From Theorem 3.6 and Theorem 5.1 of Billingsley [57], we have, for every $\delta\in(0,b]$,

$$w(F_p,\delta)\xrightarrow{i.p.}w(F_y,\delta)\qquad\text{as } p\to\infty.$$

Since F_y is continuous on $[0,\infty)$, we apply the dominated convergence theorem to the RHS of (10.4.19) and find that, for every $\epsilon>0$, $\mathrm{P}(w(X_p(F_p),\delta)\ge\epsilon)$ can be made arbitrarily small for all p sufficiently large by choosing δ appropriately. Therefore, by Theorem 10.4.9, $X_p(F_p)\xrightarrow{\mathscr{D}}W_x^y$ in $D[0,b]$, which implies $X_p(F_p(\cdot))\xrightarrow{\mathscr{D}}W_x^y$ in $D[0,\infty)$, and by Theorem 10.8 we conclude that $X_p\xrightarrow{\mathscr{D}}W^\circ$.

10.5 Extension of (10.2.6) to $\mathbf{B}_n = \mathbf{T}^{1/2}\mathbf{S}_p\mathbf{T}^{1/2}$

First, we point out that the random variables $X_p(F_p^{\mathbf{S}_p}(x))$ can be written as

$$X_p(F_p^{\mathbf{S}_p}(x)) = \sqrt{p/2}(F_*^{\mathbf{S}_p}(x) - F_p^{\mathbf{S}_p}(x)),$$

where $F_*^{\mathbf{S}_p}$ is another empirical spectral distribution that puts mass $|y_i|^2$ at the place λ_i. We shall call it the eigenvector ESD or simply VESD. Throughout this section, we shall consider the linear functionals of $X_p(F_p^{\mathbf{B}_n})$ associated with the matrix $\mathbf{B}_n = \mathbf{T}^{1/2}\mathbf{S}_p\mathbf{T}^{1/2}$, where the random variables may be complex.

10.5.1 First-Order Limit

Theorem 10.16. *Suppose:*

(1) *For each p $x_{ij} = x_{ij}^{(p)}, i, j = 1, 2, \cdots$, are iid complex random variables with* $\mathrm{E}x_{11} = 0$ *and* $\mathrm{E}|x_{11}^2| = 1$.

(2) $\mathbf{x}_p \in \mathbb{C}_1^p = \{\mathbf{x} \in \mathbb{C}^p, \|\mathbf{x}\| = 1\}$ *and* $\lim_{p\to\infty} \frac{p}{n} = y \in (0, \infty)$.

(3) *$\mathbf{T}$ is $p \times p$ nonrandom Hermitian (or symmetric in the real case) nonnegative definite with its spectral norm bounded in p, with $H_p = F^{\mathbf{T}} \xrightarrow{\mathscr{D}} H$ a proper distribution function and with $\mathbf{x}_p^*(\mathbf{T} - z\mathbf{I})^{-1}\mathbf{x}_p \to s_{F^H}(z)$, where $s_{F^H}(z)$ denotes the Stieltjes transform of $H(t)$.*

Then, it holds that

$$F_*^{\mathbf{B}_p}(x) \to F^{y,H}(x) \quad a.s.$$

Remark 10.17. If $\mathbf{T} = b\mathbf{I}$ for some positive constant b or more generally $\lambda_{\max}(\mathbf{T}) - \lambda_{\min}(\mathbf{T}) \to 0$, the condition $\mathbf{x}_p^*(\mathbf{T} - z\mathbf{I})^{-1}\mathbf{x}_p \to m_{F^H}(z)$ holds uniformly for all $\mathbf{x}_p \in \mathbb{C}_1^p$. In other cases, this condition may not hold for all $\mathbf{x}_p \in \mathbb{C}_1^p$. However, there always exists some $\mathbf{x}_p \in \mathbb{C}_1^p$ such that this condition holds, say $\mathbf{x}_p = (\mathbf{u}_1 + \cdots + \mathbf{u}_p)/\sqrt{p}$, where $\mathbf{u}_1, \cdots, \mathbf{u}_n$ are the orthonormal eigenvectors in the spectral decomposition of $\mathbf{T}$.

Applying Theorem 10.16, we get the following interesting results.

Corollary 10.18. *Let $(\mathbf{B}_p^m)_{ii}, m = 1, 2, \cdots$, denote the i-th diagonal elements of matrices $\mathbf{B}_p^m$. Under the conditions of Theorem 10.16 for $\mathbf{x}_p = \mathbf{e}_{pi}$, we have, for any fixed m,*

$$\lim_{p\to\infty} \left|(\mathbf{B}_p^m)_{ii} - \int x^m dF^{y,H}(x)\right| \to 0 \quad a.s.,$$

where $\mathbf{e}_{pi}$ is the p-vector with i-th element 1 and others 0.

Remark 10.19. If $\mathbf{T} = b\mathbf{I}$ for some positive constant b or more generally $\lambda_{\max}(\mathbf{T}) - \lambda_{\min}(\mathbf{T}) \to 0$, there is a better result,

$$\lim_{n\to\infty} \max_i \left|(\mathbf{B}_p^m)_{ii} - \int x^m dF^{y,H}(x)\right| \to 0 \quad \text{a.s.} \tag{10.5.26}$$

(The proof of this corollary follows easily from the uniform convergence of condition (3) of Theorem *10.16* for all $\mathbf{x}_p \in \mathbb{C}_1^p$, with careful checking of the proof of Theorem 10.16.)

More generally, we have the following corollary.

Corollary 10.20. *If $f(x)$ is a bounded function and the assumptions of Theorem 10.16 are satisfied, then*

$$\sum_{j=1}^p |y_j^2| f(\lambda_j) - \frac{1}{p}\sum_{j=1}^p f(\lambda_j) \to 0 \quad a.s.$$

10.5.2 CLT of Linear Functionals of B_p

Theorem 10.21. *In addition to the conditions of Theorem 10.16, we assume that* $\mathrm{E}|x_{11}^4| < \infty$ *and*

(4) $g_1, \cdots, g_k$ *are defined and analytic on an open region $\mathcal{D}$ of the complex plane that contains the real interval*

$$\left[\liminf_p \lambda_{\min}^{\mathbf{T}} I_{(0,1)}(y)(1-\sqrt{y})^2, \limsup_p \lambda_{\max}^{\mathbf{T}}(1+\sqrt{y})^2\right]. \tag{10.5.27}$$

(5) $\sup_z \sqrt{n}\left|\mathbf{x}_p^*(\underline{s}_{F^{y_n,H_n}}(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{x}_p - \int \frac{1}{\underline{s}_{F^{y_n,H_n}}(z)t+1} dH_n(t)\right| \to 0$ *as* $n \to \infty$.

Then the following conclusions hold:

(a) *The random vectors*

$$\left(\int g_1(x)dG_n(x), \cdots, \int g_k(x)dG_n(x)\right) \tag{10.5.28}$$

form a tight sequence.

(b) *If x_{11} and $\mathbf{T}$ are real and* $\mathrm{E}x_{11}^4 = 3$, *the random vector above converges weakly to a Gaussian vector $X_{g_1}, \cdots, X_{g_k}$, with zero means and covariance function*

$$\begin{aligned}&\mathrm{Cov}(X_{g_1}, X_{g_2}) \qquad (10.5.29)\\&= -\frac{1}{2\pi^2}\int_{\mathcal{C}_1}\int_{\mathcal{C}_2} g_1(z_1)g_2(z_2)\frac{(z_2\underline{m}(z_2)-z_1\underline{m}(z_1))^2}{y^2 z_1 z_2(z_2-z_1)(\underline{m}(z_2)-\underline{m}(z_1))}dz_1dz_2.\end{aligned}$$

The contours $\mathcal{C}_1$ and $\mathcal{C}_2$ in the equality above are disjoint, both contained in the analytic region for the functions $(g_1, \cdots, g_k)$ and both enclosing the support of F^{y_n,H_n} for all large p.

(c) *If x_{11} is complex, with $\mathrm{E}x_{11}^2 = 0$ and $\mathrm{E}|x_{11}|^4 = 2$, then the conclusions (a) and (b) still hold but the covariance function reduces to half of the quantity given in (10.5.29).*

Remark 10.22. If $\mathbf{T} = b\mathbf{I}$ for some positive constant b or more generally $\sqrt{p}(\lambda_{\max}(\mathbf{T}) - \lambda_{\min}(\mathbf{T})) \to 0$, then condition (5) holds uniformly for all $\mathbf{x}_p \in \mathbb{C}_1^p$.

Theorem 10.23. *Besides the assumptions of Theorem 10.21, $H(x)$ satisfies*

$$\int \frac{dH(t)}{(1+t\underline{m}(z_1))(1+t\underline{s}(z_2))} - \int \frac{dH(t)}{(1+t\underline{s}(z_1))}\int \frac{dH(t)}{(1+t\underline{s}(z_2))} = 0 \quad (10.5.30)$$

Then all results of Theorem 10.21 remain true. Moreover, formula (10.5.29) can be simplified to

$$\begin{aligned}&\mathrm{Cov}(X_{g_1}, X_{g_2}) \qquad (10.5.31)\\&= \frac{2}{y}\left(\int g_1(x)g_2(x)dF^{y,H}(x) - \int g_1(x_1)dF^{y,H}(x_1)\int g_2(x_2)dF^{y,H}(x_2)\right).\end{aligned}$$

Remark 10.24. Obviously, (*10.5.30*) holds when $\mathbf{T} = b\mathbf{I}$. Actually, (10.5.30) holds if and only if $H(x)$ is a degenerate distribution. To see it, one only needs to choose z_2 to be the complex conjugate of z_1.

10.6 Proof of Theorem 10.16

Without loss of generality, we assume that $\|\mathbf{T}\| \le 1$, where $\|\cdot\|$ denotes the spectral norm on the matrices; i.e., the largest singular values.

For $C > 0$, let $\tilde{x}_{ij} = x_{ij}I(|x_{ij}| \le C) - \mathrm{E}x_{ij}I(|x_{ij}| \le C)$ and $\tilde{\mathbf{B}}_p = \frac{1}{n}\mathbf{T}^{\frac12}\tilde{\mathbf{X}}_p\tilde{\mathbf{X}}_p^*\mathbf{T}^{\frac12}$, where $\tilde{\mathbf{X}}_p = (\tilde{x}_{ij})$.

Let $v = \Im z > 0$. Since $x_{ij} - \tilde{x}_{ij} = x_{ij}I(|x_{ij}| > C) - \mathrm{E}x_{ij}I(|x_{ij}| > C)$ and $\|(\mathbf{B}_p - z\mathbf{I})^{-1}\|$ is bounded by $\frac{1}{v}$, by Theorem 5.8 we have

$$\begin{aligned}&|\mathbf{x}_p^*(\mathbf{B}_p - z\mathbf{I})^{-1}\mathbf{x}_p - \mathbf{x}_p^*(\tilde{\mathbf{B}}_p - z\mathbf{I})^{-1}\mathbf{x}_p|\\&\le \|(\mathbf{B}_p - z\mathbf{I})^{-1}\|\|(\mathbf{B}_p - \bar{\mathbf{B}}_p)\|\|(\tilde{\mathbf{B}}_p - z\mathbf{I})^{-1}\|\\&\le \frac{1}{nv^2}\|\mathbf{X}_p - \tilde{\mathbf{X}}_p\|\|\mathbf{X}_p^*\| + \|\tilde{\mathbf{X}}_p\|\|\mathbf{X}_p^* - \tilde{\mathbf{X}}_p^*\| \qquad (10.6.1)\end{aligned}$$

$$\to \frac{(1+\sqrt{y})^2}{v^2}[\mathrm{E}^{1/2}|x_{11}-\tilde{x}_{11}|^2(\mathrm{E}^{1/2}|\tilde{x}_{11}|^2+\mathrm{E}^{1/2}|x_{11}|^2)] \quad \text{a.s.}$$
$$\leq \frac{2(1+\sqrt{y})^2}{v^2}\mathrm{E}^{1/2}|x_{11}|^2\mathbf{I}(|x_{11}|>C).$$

The bound above can be made arbitrarily small by choosing C sufficiently large. Since $\lim_{C\to\infty}\mathrm{E}|\bar{x}_{11}|^2=1$, after proper rescaling of $\tilde{x}_{ij}$, the difference can still be made arbitrarily small. Hence, in what follows, it is enough to assume $|x_{ij}|\leq C, \mathrm{E}x_{11}=0$, and $\mathrm{E}|x_{11}|^2=1$.

Next, we will show that

$$\mathbf{x}_p^*(\mathbf{B}_p-z\mathbf{I})^{-1}\mathbf{x}_p-\mathbf{x}_p^*\mathrm{E}(\mathbf{B}_p-z\mathbf{I})^{-1}\mathbf{x}_p\to 0 \quad \text{a.s.} \tag{10.6.2}$$

Let $\mathbf{r}_j$ denote the j-th column of $\frac{1}{\sqrt{n}}\mathbf{T}^{\frac{1}{2}}\mathbf{X}_p$, $\mathbf{D}(z)=\mathbf{B}_p-z\mathbf{I}$, $\mathbf{D}_j(z)=\mathbf{D}(z)-\mathbf{r}_j\mathbf{r}_j^*$,

$$\alpha_j(z)=\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*(\mathrm{E}\underline{s}_p(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{r}_j-\frac{1}{n}\mathbf{x}_p^*(\mathrm{E}\underline{s}_p(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_j^{-1}(z)\mathbf{x}_p,$$

$$\xi_j(z)=\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j-\frac{1}{n}\mathrm{tr}\mathbf{T}\mathbf{D}_j^{-1}(z),$$

$$\gamma_j=\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j-\frac{1}{n}\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{T}\mathbf{D}_j^{-1}(z)\mathbf{x}_p \tag{10.6.3}$$

and

$$\beta_j(z)=\frac{1}{1+\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j}, \quad b_j(z)=\frac{1}{1+n^{-1}\mathrm{tr}\mathbf{T}\mathbf{D}_j^{-1}(z)}.$$

Noting that $|\beta_j(z)|\leq|z|/v$, $\|\mathbf{D}_j^{-1}(z)\|\leq 1/v$, by Lemma B.26, we have

$$\mathrm{E}|\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j|^r=O\left(\frac{1}{n^r}\right),\ \mathrm{E}|\xi_j(z)|^r=O\left(\frac{1}{n^{r/2}}\right). \tag{10.6.4}$$

Define the σ-field $\mathscr{F}_j=\sigma(\mathbf{r}_1,\cdots,\mathbf{r}_j)$, and let $\mathrm{E}_j(\cdot)$ denote conditional expectation with respect to the σ-field $\mathscr{F}_j$ and $\mathrm{E}_0(\cdot)$ denote unconditional expectation. Note that

$$\mathbf{x}_p^*(\mathbf{B}_p-z\mathbf{I})^{-1}\mathbf{x}_p-\mathbf{x}_p^*\mathrm{E}(\mathbf{B}_p-z\mathbf{I})^{-1}\mathbf{x}_p \tag{10.6.5}$$
$$=\sum_{j=1}^n\mathbf{x}_p^*\mathbf{E}_j\mathbf{D}^{-1}(z)\mathbf{x}_p-\mathbf{x}_p^*\mathrm{E}_{j-1}\mathbf{D}^{-1}(z)\mathbf{x}_p$$
$$=\sum_{j=1}^n\mathbf{x}_p^*\mathrm{E}_j(\mathbf{D}^{-1}(z)-\mathbf{D}_j^{-1}(z))\mathbf{x}_p-\mathbf{x}_p^*\mathrm{E}_{j-1}(\mathbf{D}^{-1}(z)-\mathbf{D}_j^{-1}(z))\mathbf{x}_p$$
$$=-\sum_{j=1}^n(\mathrm{E}_j-\mathrm{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j$$

$$= -\sum_{j=1}^{n}[\mathrm{E}_j b_j(z)\gamma_j(z) - (\mathrm{E}_j - \mathrm{E}_{j-1})\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j\beta_j(z)b_j(z)\xi_j(z).$$

By the fact that $|\frac{1}{1+\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j}| \le \frac{|z|}{v}$ and making use of the Burkholder inequality, (10.6.4), and the martingale expression (10.6.5), we have

$$\begin{aligned}
&E|\mathbf{x}_p^*(\mathbf{B}_p - z\mathbf{I})^{-1}\mathbf{x}_p - \mathbf{x}_p^*\mathrm{E}(\mathbf{B}_p - z\mathbf{I})^{-1}\mathbf{x}_p|^r \\
&\le E\left[\sum_{j=1}^{n}\mathrm{E}_{j-1}|(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j|^2\right]^{\frac{r}{2}} \\
&\quad +\mathrm{E}\sum_{j=1}^{n}|(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j|^r \\
&\le E\left[\sum_{j=1}^{n}\frac{K|z|^2}{v^2}\mathrm{E}_{j-1}|\gamma_j(z)|^2 + \mathrm{E}_{j-1}|\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j\xi_j(z)|^2\right]^{\frac{r}{2}} \\
&\quad +\sum_{j=1}^{n}\frac{K|z|^r}{v^r}\mathrm{E}|\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j|^r \\
&\le K[p^{-\frac{r}{2}} + p^{-r+1}].
\end{aligned}$$

Thus, (10.6.2) follows from the Borel-Cantelli lemma, by taking $r > 2$.

Write

$$\mathbf{D}(z) - (-z\mathrm{E}\underline{s}_p(z)\mathbf{T} - z\mathbf{I}) = \sum_{j=1}^{n}\mathbf{r}_j\mathbf{r}_j^* - (-z\mathrm{E}\underline{s}_p(z))\mathbf{T}.$$

Using equalities

$$\mathbf{r}_j^*\mathbf{D}^{-1}(z) = \beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)$$

and

$$\underline{s}_p(z) = -\frac{1}{zn}\sum_{j=1}^{n}\beta_j(z) \tag{10.6.6}$$

(see (6.2.4)), we obtain

$$\begin{aligned}
&\mathrm{E}\mathbf{D}^{-1}(z) - (-z\mathrm{E}\underline{s}_p(z)\mathbf{T} - z\mathbf{I})^{-1} \\
&= (z\mathrm{E}\underline{s}_p(z)\mathbf{T} + z\mathbf{I})^{-1}\mathrm{E}\left[\sum_{j=1}^{n}\mathbf{r}_j\mathbf{r}_j^* - (-z\mathrm{E}\underline{s}_p(z))\mathbf{T}\mathbf{D}^{-1}(z)\right] \\
&= \frac{1}{z}\sum_{j=1}^{n}\mathrm{E}\beta_j(z)\left[(\mathrm{E}\underline{s}_p(z)\mathbf{T} + \mathbf{I})^{-1}\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z) - \frac{1}{n}(\mathrm{E}\underline{s}_p(z)\mathbf{T} + \mathbf{I})^{-1}\mathbf{T}\mathrm{E}\mathbf{D}^{-1}(z)\right].
\end{aligned}$$

Multiplying by $\mathbf{x}_p^*$ on the left and $\mathbf{x}_p$ on the right, we have

$$\begin{aligned}
&\mathbf{x}_p^*\mathrm{E}\mathbf{D}^{-1}(z)\mathbf{x}_p - \mathbf{x}_p^*(-z\mathrm{E}\underline{s}_p(z)\mathbf{T} - z\mathbf{I})^{-1}\mathbf{x}_p \\
&= n\frac{1}{z}\mathrm{E}\beta_1(z)[\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*(\mathrm{E}\underline{s}_p(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{r}_1 \qquad (10.6.7)\\
&\quad -\frac{1}{n}\mathbf{x}_p^*(\mathrm{E}\underline{s}_p(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathrm{E}\mathbf{D}^{-1}(z)\mathbf{x}_p] \\
&\triangleq \delta_1+\delta_2+\delta_3,
\end{aligned}$$

where

$$\begin{aligned}
\delta_1 &= \frac{n}{z}\mathrm{E}\beta_1(z)\alpha_1(z),\\
\delta_2 &= \frac{1}{z}\mathrm{E}[\beta_1(z)\mathbf{x}_p^*(\mathrm{E}\underline{s}_p(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}(\mathbf{D}_1^{-1}(z)-\mathbf{D}^{-1}(z))\mathbf{x}_p],\\
\delta_3 &= \frac{1}{z}\mathrm{E}[\beta_1(z)\mathbf{x}_p^*(\mathrm{E}\underline{s}_p(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}(\mathbf{D}^{-1}(z)-\mathrm{E}\mathbf{D}^{-1}(z))\mathbf{x}_p].
\end{aligned}$$

Similar to (10.6.4), by Lemma B.26, for $r \ge 2$, we have

$$\mathrm{E}|\alpha_j(z)|^r = O\left(\frac{1}{n^r}\right).$$

Therefore,

$$\delta_1 = \frac{n}{z}\mathrm{E}b_1(z)\beta_1(z)\xi_1(z)\alpha_1(z) = O(n^{-1/2}).$$

It follows that

$$\begin{aligned}
|\delta_2| &= \frac{1}{|z|}|\mathrm{E}[\beta_1^2(z)\mathbf{x}_p^*(\mathrm{E}\underline{s}_p(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_1^{-1}(z)\mathbf{r}_1\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{x}_p]|\\
&\le K\left(\mathrm{E}|\mathbf{x}_p^*(\mathrm{E}\underline{s}_p(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_1^{-1}(z)\mathbf{r}_1|^2\mathrm{E}|\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{x}_p|^2\right)^{1/2}\\
&= O(n^{-1})
\end{aligned}$$

and

$$\begin{aligned}
|\delta_3| &= \frac{1}{|z|}|\mathrm{E}[\beta_1(z)b_1(z)\xi_1(z)\mathbf{x}_p^*(\mathrm{E}\underline{s}_p(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}(\mathbf{D}^{-1}(z)-\mathrm{E}\mathbf{D}^{-1}(z))\mathbf{x}_p]|\\
&\le K\mathrm{E}|\xi_1(z)| = O(n^{-1/2}).
\end{aligned}$$

Combining the three bounds above and (10.6.7), we can conclude that

$$\mathbf{x}_p^*\mathrm{E}\mathbf{D}^{-1}(z)\mathbf{x}_p - \mathbf{x}_p^*(-z\mathrm{E}\underline{s}_p(z)\mathbf{T} - zI)^{-1}\mathbf{x}_p \to 0. \qquad (10.6.8)$$

It has been proven in Section 9.11 that, under the conditions of Theorem 10.16, $\mathrm{E}\underline{s}_p(z) \to \underline{s}(z)$, which is the solution to equation (9.7.1). We then conclude that

$$\mathbf{x}_p^*\mathrm{E}\mathbf{D}^{-1}(z)\mathbf{x}_p - \mathbf{x}_p^*(-z\underline{s}(z)\mathbf{T} - zI)^{-1}\mathbf{x}_p \to 0.$$

By condition (3) of Theorem 10.16, we finally obtain

$$\mathbf{x}_p^* \mathrm{E}\mathbf{D}^{-1}(z)\mathbf{x}_p \to \int \frac{dH(t)}{-z\underline{s}t - z},$$

which completes the proof of Theorem 10.16.

10.7 Proof of Theorem 10.21

The proof of Theorem 10.21 will be separated into several subsections.

10.7.1 An Intermediate Lemma

To complete the proof of Theorem 10.21, we need the intermediate Lemma 10.25 below.

Write

$$M_p(z) = \sqrt{n}(s_{F_*^{\mathbf{B}_p}}(z) - s_{F^{y_p, H_p}}(z)),$$

which is defined on a contour $\mathcal{C}$ in the complex plane, where $\mathcal{C}$ and the numbers u_r, u_l, μ_1, μ_2, and $v_0 > 0$ are the same as defined in Section 9.8.

Similar to Section 9.8, we consider $M_p^*(z)$, a truncated version of $M_p(z)$. Choose a sequence of positive numbers $\{\delta_p\}$ such that for $0 < \rho < 1$

$$\delta_p \downarrow 0, \quad \delta_p \geq p^{-\rho}. \tag{10.7.1}$$

Write

$$M_p^*(z) = \begin{cases} M_p(z) \\ \quad \text{if} \quad z \in \mathcal{C}_0 \cup \bar{\mathcal{C}}_0 \\ \frac{pv+\delta_p}{2\delta_p} M_p(u_r + in^{-1}\delta_p) + \frac{\delta_p - pv}{2\delta_p} M_p(u_r - ip^{-1}\delta_p) \\ \quad \text{if} \quad u = u_r, v \in [-p^{-1}\delta_p, p^{-1}\delta_p] \\ \frac{pv+\delta_p}{2\delta_p} M_p(u_l + ip^{-1}\delta_p) + \frac{\delta_p - pv}{2\delta_p} M_p(u_l - ip^{-1}\delta_p) \\ \quad \text{if} \quad u = u_l > 0, v \in [-p^{-1}\delta_p, p^{-1}\delta_p]. \end{cases}$$

$M_p^*(z)$ can be viewed as a random element in the metric space $C(\mathcal{C}, \mathbb{R}^2)$ of continuous functions from $\mathcal{C}$ to $\mathbb{R}^2$. We shall prove the following lemma.

Lemma 10.25. *Under the assumptions of Theorem 10.16 and (4) and (5) of Theorem 10.21, $M_p^*(z)$ forms a tight sequence on $\mathcal{C}$. Furthermore, when the conditions in (b) and (c) of Theorem 10.21 on x_{11} hold, for $z \in \mathcal{C}$, $M_p^*(z)$ converges to a Gaussian process $M(\cdot)$ with zero mean and for $z_1, z_2 \in \mathcal{C}$, under the assumptions in (b),*

$$\mathrm{Cov}(M(z_1), M(z_2)) = \frac{2(z_2\underline{s}(z_2) - z_1\underline{s}(z_1))^2}{y^2 z_1 z_2 (z_2 - z_1)(\underline{s}(z_2) - \underline{s}(z_1))}, \tag{10.7.2}$$

while under the assumptions in (c), a covariance function similar to (10.7.2) is half of the value of (10.7.2).

To prove Theorem 10.21, it suffices to prove Lemma 10.25. Before proving the lemma, we first truncate and recentralize the variables x_{ij}. Choose $\eta_p \to 0$ and such that $\mathrm{E}|x_{11}^4|I(|x_{11}| > \eta_p\sqrt{p}) = o(\eta_n^4)$. Truncate the variables x_{ij} at $\eta_p p^{1/2}$ and recentralize them. Similar to Subsection 10.3.3, one can prove that the truncation and recentralization do not affect the limiting result. Therefore, we may assume that the following additional conditions hold:

$$|x_{ij}| \le \eta_p\sqrt{p}, \mathrm{E}x_{11} = 0, \mathrm{E}|x_{11}|^2 = 1 + o(p^{-1})$$

and

$$\begin{cases} \mathrm{E}|x_{11}|^4 = 3 + o(1), & \text{for the real case,} \\ \mathrm{E}x_{11}^2 = o(p^{-1}), \mathrm{E}|x_{11}|^4 = 2 + o(1), & \text{for the complex case.} \end{cases}$$

The proof of Lemma 10.25 will be given in the next two subsections.

10.7.2 Convergence of the Finite-Dimensional Distributions

For $z \in \mathcal{C}_0$, let

$$M_p^1(z) = \sqrt{n}(s_{F_*^{\mathbf{B}_p}}(z) - \mathrm{E}s_{F_*^{\mathbf{B}_p}}(z))$$

and

$$M_p^2(z) = \sqrt{n}(\mathrm{E}s_{F_*^{\mathbf{B}_p}}(z) - s_{F^{y_p,H_p}}(z)).$$

Then

$$M_p(z) = M_p^1(z) + M_p^2(z).$$

In this part, for any positive integer r and complex numbers $a_1, \cdots, a_r$, we will show that

$$\sum_{i=1}^r a_i M_p^1(z_i) \qquad (\Im z_i \ne 0)$$

converges in distribution to a Gaussian random variable and will derive the covariance function (10.7.2).

Before proceeding with the proofs, we first recall some known facts and results. For any nonrandom matrices $\mathbf{C}$ and $\mathbf{Q}$ and positive constant $2 \le \ell \le 8\log p$, by using Lemma 9.1 for some constant K, we have

$$\mathrm{E}|\mathbf{r}_1^*\mathbf{C}\mathbf{r}_1 - n^{-1}\mathrm{tr}\mathbf{T}\mathbf{C}|^\ell \le K^\ell \|\mathbf{C}\|^\ell \eta_p^{2\ell-4} p^{-1}, \tag{10.7.3}$$

$$\mathrm{E}|\mathbf{r}_1^*\mathbf{C}\mathbf{x}_p\mathbf{x}_p^*\mathbf{Q}\mathbf{r}_1 - n^{-1}\mathbf{x}_p^*\mathbf{Q}\mathbf{T}\mathbf{C}\mathbf{x}_p|^\ell \leq K^\ell\|\mathbf{C}\|^\ell\|\mathbf{Q}\|^\ell\eta_p^{2\ell-4}p^{-2}, \quad (10.7.4)$$

$$\mathrm{E}|\mathbf{r}_1^*\mathbf{C}\mathbf{x}_p\mathbf{x}_p^*\mathbf{Q}\mathbf{r}_1|^\ell \leq K^\ell\|\mathbf{C}\|^\ell\|\mathbf{Q}\|^\ell\eta_p^{2\ell-4}p^{-2}. \quad (10.7.5)$$

Let $v = \Im z$. To facilitate the analysis, we will assume $v > 0$. By (10.6.5), we have

$$\sqrt{n}(s_{F_*^{\mathbf{B}_p}}(z) - \mathrm{E}s_{F_*^{\mathbf{B}_p}}(z)) = -\sqrt{n}\sum_{j=1}^n(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j.$$

Since

$$\beta_j(z) = b_j(z) - \beta_j(z)b_j(z)\xi_j(z) = b_j(z) - b_j^2(z)\xi_j(z) + b_j^2(z)\beta_j(z)\xi_j^2(z),$$

we then get

$$\begin{aligned}
&(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j \\
&= \mathrm{E}_j b_j(z)\gamma_j(z) - \mathrm{E}_j\left(b_j^2(z)\xi_j(z)\frac{1}{n}\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{T}\mathbf{D}_j^{-1}(z)\mathbf{x}_p\right) \\
&\quad +(\mathrm{E}_j - \mathrm{E}_{j-1})(b_j^2(z)\beta_j(z)\xi_j^2(z)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j - b_j^2(z)\xi_j(z)\gamma_j(z)),
\end{aligned}$$

where $\gamma_j = \mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j - \frac{1}{n}\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{T}\mathbf{D}_j^{-1}(z)\mathbf{x}_p$.

Applying (10.7.3),

$$\mathrm{E}\left|\sqrt{n}\sum_{j=1}^n\mathrm{E}_j\left(b_j^2(z)\xi_j(z)\frac{1}{n}\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{T}\mathbf{D}_j^{-1}(z)\mathbf{x}_p\right)\right|^2$$

$$= \frac{1}{n}\sum_{j=1}^n\mathrm{E}|\mathrm{E}_j(b_j^2(z)\xi_j(z)\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{T}\mathbf{D}_j^{-1}(z)\mathbf{x}_p)|^2 \leq K\frac{|z|^4}{v^8}\mathrm{E}|\xi_1(z)|^2 = O(p^{-1}),$$

which implies that $\sqrt{n}\sum_{j=1}^n\mathrm{E}_j(b_j^2(z)\xi_j(z)\frac{1}{n}\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{T}\mathbf{D}_j^{-1}(z)\mathbf{x}_p) \xrightarrow{i.p.} 0$.

By (10.7.3), (10.7.5), and Hölder's inequality with $\ell = \log p$ and $l = \log p/(\log p - 1)$, we have

$$\begin{aligned}
&\mathrm{E}\left|\sqrt{n}\sum_{j=1}^n(\mathrm{E}_j - \mathrm{E}_{j-1})(b_j^2(z)\beta_j(z)\xi_j^2(z)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j\right|^2 \\
&\leq K\left(\frac{|z|}{v}\right)^6 n\sum_{j=1}^n\left(\mathrm{E}|\xi_j^4(z)||\gamma_j^2(z)| + \frac{1}{n^2}\mathrm{E}|\xi_j^4(z)||\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{T}\mathbf{D}_j^{-1}(z)\mathbf{x}_p|^2\right) \\
&\leq K\left(\frac{|z|}{v}\right)^6 n\sum_{j=1}^n\left(\mathrm{E}|\xi_j^{4\ell}(z)|\right)^{1/\ell}\left(\mathrm{E}|\gamma_j^{2l}(z)|\right)^{1/l} + O(p^{-1})
\end{aligned}$$

$$\leq Kn^2 \left(\eta_p^{8\ell-4} p^{-1}\right)^{1/\ell} \left(\eta_p^{4l-4} p^{-2}\right)^{1/l} + O(p^{-1})$$
$$= o(1),$$

which implies that

$$\sqrt{n}\sum_{j=1}^{n}(\mathrm{E}_j - \mathrm{E}_{j-1}) b_j^2(z)\beta_j(z)\xi_j^2(z)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z)\mathbf{r}_j \xrightarrow{i.p.} 0.$$

Using a similar argument, we have

$$\sqrt{n}\sum_{j=1}^{n}(\mathrm{E}_j - \mathrm{E}_{j-1}) b_j^2(z)\xi_j(z)\gamma_j(z) \xrightarrow{i.p.} 0.$$

The estimates (6.2.36), (9.9.20), and (10.7.4) yield

$$\mathrm{E}|(b_j(z) + z\underline{s}(z))\gamma_j(z)|^2 = \mathrm{E}[\mathrm{E}(|(b_j(z) + z\underline{s}(z))\gamma_j(z)|^2 | \mathscr{B}(\mathbf{r}_i, i \neq j))]$$
$$= \mathrm{E}[|b_j(z) + z\underline{s}(z)|^2 \mathrm{E}(|\gamma_j(z)|^2 | \mathscr{B}(\mathbf{r}_i, i \neq j))] = o(p^{-2}),$$

which gives us

$$\sqrt{n}\sum_{j=1}^{n}\mathrm{E}_j[(b_j(z) + z\underline{s}(z))\gamma_j(z)] \xrightarrow{i.p.} 0,$$

where $\mathscr{B}(\cdot)$ denotes the Borel field generated by the random variables indicated in the brackets.

Note that the results above also hold when $\Im z \leq -v_0$ by symmetry. Hence, for the finite dimensional convergence, we need only consider the sum

$$\sum_{i=1}^{r} a_i \sum_{j=1}^{n} Y_j(z_i) = \sum_{j=1}^{n}\sum_{i=1}^{r} a_i Y_j(z_i),$$

where $Y_j(z_i) = -\sqrt{n} z_i \underline{s}(z_i)\mathrm{E}_j\gamma_j(z_i)$ and γ_j is defined in (10.6.3).

Next, we will show that $Y_j(z_i)$ satisfies the Lindeberg condition; that is, for any $\varepsilon > 0$,

$$\sum_{j=1}^{n}\mathrm{E}|Y_j(z_i)|^2 I(|Y_j(z_i)| \geq \varepsilon) \to 0. \tag{10.7.6}$$

Write $\gamma_j(z_i) = \gamma_j^{(1)} + \gamma_j^{(2)} + \gamma_j^{(3)} + \gamma_j^{(4)}$, where

$$\gamma_j^{(1)} = \frac{1}{n}\sum_{k \neq l}\mathbf{e}_k'\mathbf{D}_j^{-1}(z_i)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_i)\mathbf{e}_l \bar{x}_{kj}x_{lj},$$
$$\gamma_j^{(2)} = \frac{1}{n}\sum_{k=1}^{p}\mathbf{e}_k'\mathbf{D}_j^{-1}(z_i)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_i)\mathbf{e}_k$$

$$\times[|x_{kj}^2|I(|x_{kj}^2|<\log p)-\mathrm{E}|x_{kj}^2|I(|x_{kj}^2|<\log p)],$$

$$\gamma_j^{(3)}=\frac{1}{n}\sum_{k=1}^p \mathbf{e}_k'\mathbf{D}_j^{-1}(z_i)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_i)\mathbf{e}_k$$
$$\times[|x_{kj}^2|I(|x_{kj}^2|\geq\log p)-\mathrm{E}|x_{kj}^2|I(|x_{kj}^2|\geq\log p)]$$

$$\gamma_j^{(4)}=\frac{1}{n}\sum_{k=1}^p \mathbf{e}_k'\mathbf{D}_j^{-1}(z_i)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_i)\mathbf{e}_k[\mathrm{E}|x_{kj}^2|-1]=O(p^{-1}).$$

Similar to the proof of Lemma B.26, we can prove that

$$\mathrm{E}|\gamma_j^{(1)}|^4=O(p^{-4}),\ \mathrm{E}|\gamma_j^{(2)}|^4=O(p^{-4}\log^2 p),\quad\text{and}\quad \mathrm{E}|\gamma_j^{(3)}|^2=o(p^{-2}), \tag{10.7.7}$$

where the $o(1)$ comes from the fact that $\mathrm{E}|x_{kj}^4||(|x_{kj}^2|\geq\log p)\to 0$. Consequently, (10.7.6) follows from the observation that

$$\sum_{j=1}^n \mathrm{E}|Y_j(z_i)|^2 I(|Y_j(z_i)|\geq\varepsilon)$$
$$\leq 4\sum_{j=1}^n\sum_{l=1}^4 \mathrm{E}\big|Y_j^{(l)}(z_i)\big|^2 I\big(|Y_j^{(l)}(z_i)|\geq\varepsilon/4\big)$$
$$\leq\frac{64}{\varepsilon^2}\sum_{j=1}^n\sum_{l=1}^2 \mathrm{E}\big|Y_j^{(l)}(z_i)\big|^4+\sum_{j=1}^n \mathrm{E}\big|Y_j^{(3)}(z_i)^2\big|\to 0,$$

where $Y_j^{(l)}(z_i)=-\sqrt{n}z_i\underline{s}(z_i)\gamma_j^{(l)}$, $l\leq 4$.

By Lemma 9.12, we only need to show that, for $z_1,z_2\in\mathbb{C}\backslash\mathbb{R}$,

$$\sum_{j=1}^n \mathrm{E}_{j-1}(Y_j(z_1)Y_j(z_2)) \tag{10.7.8}$$

converges in probability to a constant under the assumptions in (b) or (c). It is easy to verify that

$$|\mathrm{tr}\mathrm{E}_j(\mathbf{D}_j^{-1}(z_1)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1))\mathbf{T}\mathrm{E}_j(\mathbf{D}_j^{-1}(z_2)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_2)\mathbf{T})|\leq\frac{1}{|v_1v_2|^2}, \tag{10.7.9}$$

where $v_1=\Im(z_1)$ and $v_2=\Im(z_2)$. It follows that, for the complex case, applying (9.8.6), (10.7.8) now becomes

$$z_1z_2\underline{s}(z_1)\underline{s}(z_2)\frac{1}{n}\sum_{j=1}^n \mathrm{E}_{j-1}\mathrm{tr}\mathrm{E}_j(\mathbf{D}_j^{-1}(z_1)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1))\mathbf{T}$$
$$\times\mathrm{E}_j(\mathbf{D}_j^{-1}(z_2)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_2)\mathbf{T})+o_p(1)$$

$$= z_1 z_2 \underline{s}(z_1)\underline{s}(z_2)\frac{1}{n}\sum_{j=1}^{n} \mathrm{E}_{j-1}(\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p)$$
$$\times(\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{D}_j^{-1}(z_1)\mathbf{x}_p) + o_p(1), \qquad (10.7.10)$$

where $\breve{\mathbf{D}}_j^{-1}(z_2)$ is similarly defined as $\mathbf{D}_j^{-1}(z_2)$ by $(\mathbf{r}_1, \cdots, \mathbf{r}_{j-1}, \breve{\mathbf{r}}_{j+1}, \cdots, \breve{\mathbf{r}}_n)$, where $\breve{\mathbf{r}}_{j+1}, \cdots, \breve{\mathbf{r}}_N$ are iid copies of $\mathbf{r}_{j+1}, \cdots, \mathbf{r}_n$.

For the real case, (10.7.8) will be twice the amount of (10.7.10).

Define

$$\mathbf{D}_{ij}(z) = \mathbf{D}(z) - \mathbf{r}_i\mathbf{R}_i^* - \mathbf{r}_j\mathbf{r}_j^*, \quad \mathbf{H}^{-1}(z_1) = \left(z_1\mathbf{I} - \frac{n-1}{n}b_{p1}(z_1)\mathbf{T}\right)^{-1},$$

$$\beta_{ij}(z) = \frac{1}{1+\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z)\mathbf{r}_i}, \quad \text{and} \quad b_{p1}(z) = \frac{1}{1+n^{-1}\mathrm{Etr}\mathbf{T}\mathbf{D}_{12}^{-1}(z)}.$$

Write

$$\mathbf{x}_p^*(\mathbf{D}_j^{-1}(z_1) - \mathrm{E}_{j-1}\mathbf{D}_j^{-1}(z_1))\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p$$
$$= \sum_{t=j}^{n} \mathbf{x}_p^*(\mathrm{E}_t\mathbf{D}_j^{-1}(z_1) - \mathrm{E}_{t-1}\mathbf{D}_j^{-1}(z_1))\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p. \qquad (10.7.11)$$

By (10.7.11), we notice that

$$\mathrm{E}_{j-1}(\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p)(\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{D}_j^{-1}(z_1)\mathbf{x}_p) \qquad (10.7.12)$$
$$= \sum_{t=j}^{n} \mathrm{E}_{j-1}\mathbf{x}_p^*(\mathrm{E}_t\mathbf{D}_j^{-1}(z_1) - \mathrm{E}_{t-1}\mathbf{D}_j^{-1}(z_1))\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)$$
$$\times\mathbf{T}(\mathrm{E}_t\mathbf{D}_j^{-1}(z_1) - \mathrm{E}_{t-1}\mathbf{D}_j^{-1}(z_1))\mathbf{x}_p$$
$$+\mathrm{E}_{j-1}(\mathbf{x}_p^*(\mathrm{E}_{j-1}\mathbf{D}_j^{-1}(z_1)\mathbf{T})\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p)(\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}(\mathrm{E}_{j-1}\mathbf{D}_j^{-1}(z_1))\mathbf{x}_p)$$
$$= \mathrm{E}_{j-1}(\mathbf{x}_p^*(\mathrm{E}_{j-1}\mathbf{D}_j^{-1}(z_1)\mathbf{T})\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p)$$
$$\times(\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}(\mathrm{E}_{j-1}\mathbf{D}_j^{-1}(z_1))\mathbf{x}_p) + O(p^{-1}),$$

where we have used the fact that

$$|\mathrm{E}_{j-1}\mathbf{x}_p^*(\mathrm{E}_t\mathbf{D}_j^{-1}(z_1) - \mathrm{E}_{t-1}\mathbf{D}_j^{-1}(z_1))\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p\times$$
$$\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}(\mathrm{E}_t\mathbf{D}_j^{-1}(z_1) - E_{t-1}\mathbf{D}_j^{-1}(z_1))\mathbf{x}_p|$$
$$\le 4\left(\mathrm{E}_{j-1}\left|\beta_{tj}(z_1)\mathbf{x}_p^*(\mathbf{D}_{tj}^{-1}(z_1)\mathbf{r}_t\mathbf{r}_t^*(\mathbf{D}_{tj}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p\right|^2\right.$$
$$\left.\times\mathrm{E}_{j-1}\left|\beta_{tj}(z_1)\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}(\mathbf{D}_{tj}^{-1}(z_1)\mathbf{r}_t\mathbf{r}_t^*\mathbf{B}_j^{-1}(z_1))\mathbf{x}_p\right|^2\right)^{1/2} = O(p^{-2}).$$

Similarly, one can prove that

$$\begin{aligned}
&\mathrm{E}_{j-1}(\mathbf{x}_p^*(\mathrm{E}_{j-1}\mathbf{D}_j^{-1}(z_1)\mathbf{T})\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p)(\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}(\mathrm{E}_{j-1}\mathbf{D}_j^{-1}(z_1))\mathbf{x}_p)\\
&=\mathrm{E}_{j-1}(\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p)\mathrm{E}_{j-1}(\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{D}_j^{-1}(z_1)\mathbf{x}_p)+O(p^{-1}).
\end{aligned}$$

Then, using the decomposition (9.9.12), we obtain

$$\begin{aligned}
&\mathrm{E}_{j-1}(\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p)\mathrm{E}_{j-1}(\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{D}_j^{-1}(z_1)\mathbf{x}_p)\\
&=-\mathrm{E}_{j-1}(\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p)\mathrm{E}_{j-1}(\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{T}\mathbf{x}_p)\\
&\quad+\mathbf{A}(z_1,z_2)+\mathbf{B}(z_1,z_2)+\mathbf{C}(z_1,z_2),
\end{aligned}\tag{10.7.13}$$

where

$$\mathbf{A}(z_1,z_2)=b_{p1}(z_1)\mathrm{E}_{j-1}(\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p)\mathrm{E}_{j-1}(\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{A}(z_1)\mathbf{x}_p),$$

$$\mathbf{B}(z_1,z_2)=\mathrm{E}_{j-1}(\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p)\mathrm{E}_{j-1}(\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{B}(z_1)\mathbf{x}_p),$$

and

$$\mathbf{C}(z_1,z_2)=\mathrm{E}_{j-1}(\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p)(\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{C}(z_1)\mathbf{x}_p).$$

We next prove that

$$\mathrm{E}|\mathbf{B}(z_1,z_2)|=o(1)\quad\text{and}\quad\mathrm{E}|\mathbf{C}(z_1,z_2)|=o(1).\tag{10.7.14}$$

Note that although $\mathbf{B}$ and $\mathbf{C}$ depend on j implicitly, $\mathrm{E}|\mathbf{B}(z_1,z_2)|$ and $\mathrm{E}|\mathbf{C}(z_1,z_2)|$ are independent of j since the entries of $\mathbf{X}_p$ are iid.

Then, we have

$$\begin{aligned}
\mathrm{E}|\mathbf{B}(z_1,z_2)|&\le\frac{1}{|v_1v_2|}\mathrm{E}|\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{B}(z_1)\mathbf{x}_p|\\
&\le\frac{1}{|v_1v_2|}\sum_{i\neq j}(\mathrm{E}|\beta_{ij}(z_1)-b_{p1}(z_1)|^2\\
&\qquad\times\mathrm{E}|\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p\mathbf{x}_p^*(\breve{\mathbf{D}}_j^{-1}(z_2))\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{r}_i|^2)^{1/2}.
\end{aligned}$$

When $i>j$, $\mathbf{r}_i$ is independent of $\breve{\mathbf{D}}_j^{-1}(z_2)$. As the proof of (10.6.4), we have

$$\mathrm{E}|\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p\mathbf{x}_p^*(\breve{\mathbf{D}}_j^{-1}(z_2))\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{r}_i|^2=O(p^{-2}).\tag{10.7.15}$$

When $i<j$, substituting $\breve{\mathbf{D}}_j^{-1}(z_2)$ by $\breve{\mathbf{D}}_{ij}^{-1}(z_2)-\breve{\beta}_{ij}(z_2)\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)$, we can also obtain the inequality above. Noting that

$$\mathrm{E}|\beta_{ij}(z_1)-b_{p1}(z_1)|^2=\mathrm{E}|\beta_{ij}(z_1)b_{p1}(z_1)\xi_{ij}|^2=O(n^{-1}),\tag{10.7.16}$$

where $\xi_{ij}(z)=\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z)\mathbf{r}_i-\frac{1}{n}\mathbf{D}_{ij}^{-1}(z)$ and $\breve{\beta}_{ij}(z_2)$ is similarly defined as $\beta_{ij}(z_2)$, combining (10.7.15)–(10.7.16), we conclude that

$$\mathrm{E}|\mathbf{B}(z_1,z_2)| = o(1).$$

The argument for $\mathbf{C}(z_1,z_2)$ is similar to that of $\mathbf{B}(z_1,z_2)$, just simpler, and is therefore omitted. Hence (10.7.14) holds.

Next, write

$$\mathbf{A}(z_1,z_2) = \mathbf{A}_1(z_1,z_2) + \mathbf{A}_2(z_1,z_2) + \mathbf{A}_3(z_1,z_2), \tag{10.7.17}$$

where

$$\begin{aligned}\mathbf{A}_1(z_1,z_2) = \sum_{i<j} & b_{p1}(z_1)\mathrm{E}_{j-1}\mathbf{x}_p^*\beta_{ij}(z_1)\mathbf{D}_{ij}^{-1}(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p \\ & \times\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_i\mathbf{r}_i^* - n^{-1}\mathbf{T})\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p,\end{aligned}$$

$$\begin{aligned}\mathbf{A}_2(z_1,z_2) = \sum_{i<j} & b_{p1}(z_1)\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\breve{\beta}_{ij}(z_2)\mathbf{x}_p \\ & \times\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_i\mathbf{r}_i^* - n^{-1}\mathbf{T})\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p,\end{aligned}$$

and

$$\begin{aligned}\mathbf{A}_3(z_1,z_2) = \sum_{i<j} & b_{p1}(z_1)\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{x}_p \\ & \times\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_i\mathbf{r}_i^* - n^{-1}\mathbf{T})\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p.\end{aligned}$$

Splitting $\breve{\mathbf{D}}_j^{-1}(z_2)$ as the sum of $\breve{\mathbf{D}}_{ij}^{-1}(z_2)$ and $-\breve{\beta}_{ij}(z_2)\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)$ as in the proof of (10.7.16), one can show that

$$\begin{aligned}\mathrm{E}|\mathbf{A}_1(z_1,z_2)| &\le \sum_{i<j}|b_{p1}(z_1)|\Big(\mathrm{E}|\mathbf{x}_p^*\beta_{ij}(z_1)\mathbf{D}_{ij}^{-1}(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p|^2 \\ &\qquad \times \mathrm{E}|\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_i\mathbf{r}_i^* - n^{-1}\mathbf{T})\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p|^2\Big)^{1/2} \\ &= O(n^{-1/2}).\end{aligned}$$

By the same argument, we have

$$\mathrm{E}|\mathbf{A}_2(z_1,z_2)| = O(n^{-1}).$$

To deal with $\mathbf{A}_3(z_1,z_2)$, we again split $\breve{\mathbf{D}}_j^{-1}(z_2)$ into the sum of $\breve{\mathbf{D}}_{ij}^{-1}(z_2)$ and $-\breve{\beta}_{ij}(z_2)\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)$. We first show that

$$\begin{aligned}\mathbf{A}_{31}(z_1,z_2) = \sum_{i<j} & b_{p1}(z_1)\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{x}_p \\ & \times\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_i\mathbf{r}_i^* - n^{-1}\mathbf{T})\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p\end{aligned}$$

$$= o_p(1). \tag{10.7.18}$$

Further, we have

$$\begin{aligned}
&\mathrm{E}|\mathbf{A}_{31}(z_1,z_2)|^2\\
&= \sum_{i_1,i_2<j} |b_{p1}(z_1)|^2 \mathrm{E}\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{i_1j}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_{i_1j}^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{i_2j}^{-1}(\bar{z}_1)\mathbf{T}\breve{\mathbf{D}}_{i_2j}^{-1}(\bar{z}_2)\mathbf{x}_p\\
&\times\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_{i_1j}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_{i_1}\mathbf{r}_{i_1}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_1j}^{-1}(z_1)\mathbf{x}_p\mathbf{x}_p^*\breve{\mathbf{D}}_{i_2j}^{-1}(\bar{z}_2)\\
&\times\mathbf{T}\mathbf{H}^{-1}(\bar{z}_1)(\mathbf{r}_{i_2}\mathbf{r}_{i_2}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_2j}^{-1}(\bar{z}_1)\mathbf{x}_p.
\end{aligned}$$

When $i_1 = i_2$, the term in the expression above is bounded by

$$K\mathrm{E}|\mathbf{x}_p^*\breve{\mathbf{D}}_{i_1j}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_{i_1}\mathbf{r}_{i_1}^* - n^{-1}\mathbf{T})\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p|^2 = O(n^{-2}).$$

For $i_1 \neq i_2 < j$, define

$$\beta_{i_1i_2j}(z_1) = \frac{1}{1+\mathbf{r}_{i_2}^*\mathbf{D}_{i_1i_2j}^{-1}(z_1)\mathbf{r}_{i_2}},\ \mathbf{D}_{i_1i_2j}(z_1) = \mathbf{D}(z_1) - \mathbf{r}_{i_1}\mathbf{r}_{i_1}^* - \mathbf{r}_{i_2}\mathbf{r}_{i_2}^* - \mathbf{r}_j\mathbf{r}_j^*,$$

and similarly define $\breve{\beta}_{i_1,i_2,j}(z_2)$ and $\breve{\mathbf{D}}_{i_1i_2j}(z_2)$.

We have

$$\begin{aligned}
&|\mathrm{E}\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{i_1,i_2,j}^{-1}(z_1)\mathbf{r}_{i_2}\mathbf{r}_{i_2}^*\mathbf{D}_{i_1,i_2,j}^{-1}(z_1)\beta_{i_1,i_2,j}(z_1)\mathbf{T}\breve{\mathbf{D}}_{i_1j}^{-1}(z_2)\\
&\quad\times\mathbf{x}_p\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{i_2j}^{-1}(\bar{z}_1)\mathbf{T}\breve{\mathbf{D}}_{i_2j}^{-1}(\bar{z}_2)\mathbf{x}_p\\
&\times E_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_{i_1j}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_{i_1}\mathbf{r}_{i_1}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_1j}^{-1}(z_1)\mathbf{x}_p\\
&\quad\times\mathbf{x}_p^*\breve{\mathbf{D}}_{i_2j}^{-1}(\bar{z}_2)\mathbf{T}\mathbf{H}^{-1}(\bar{z}_1)(\mathbf{r}_{i_2}\mathbf{r}_{i_2}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_2j}^{-1}(\bar{z}_1)\mathbf{x}_p|\\
&\leq K(\mathrm{E}|\mathbf{x}_p^*\mathbf{D}_{i_1,i_2,j}^{-1}(z_1)\mathbf{r}_{i_2}\mathbf{r}_{i_2}^*\mathbf{D}_{i_1,i_2,j}^{-1}(z_1)\beta_{i_1,i_2,j}(z_1)\mathbf{T}\breve{\mathbf{D}}_{i_1j}^{-1}(z_2)\mathbf{x}_p|^2)^{1/2}\\
&\quad\times(\mathrm{E}|\mathbf{x}_p^*\breve{\mathbf{D}}_{i_1j}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_{i_1}\mathbf{r}_{i_1}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_1j}^{-1}(z_1)\mathbf{x}_p|^4)^{1/4}\\
&\quad\times(\mathrm{E}|\mathbf{x}_p^*\breve{\mathbf{D}}_{i_2j}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_{i_2}\mathbf{r}_{i_2}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_2j}^{-1}(z_1)\mathbf{x}_p|^4)^{1/4}\\
&= O(n^{-5/2}),
\end{aligned}$$

$$\begin{aligned}
&|\mathrm{E}\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{i_1,i_2,j}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\mathbf{r}_{i_2}\mathbf{r}_{i_2}^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\breve{\beta}_{i_1,i_2,j}(z_2)\mathbf{x}_p\\
&\quad\times\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{i_2j}^{-1}(\bar{z}_1)\mathbf{T}\breve{\mathbf{D}}_{i_2j}^{-1}(\bar{z}_2)\mathbf{x}_p\\
&\times\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_{i_1j}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_{i_1}\mathbf{r}_{i_1}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_1j}^{-1}(z_1)\mathbf{x}_p\\
&\quad\times\mathbf{x}_p^*\breve{\mathbf{D}}_{i_2j}^{-1}(\bar{z}_2)\mathbf{T}\mathbf{H}^{-1}(\bar{z}_1)(\mathbf{r}_{i_2}\mathbf{r}_{i_2}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_2j}^{-1}(\bar{z}_1)\mathbf{x}_p|\\
&\leq K(\mathrm{E}|\mathbf{x}_p^*\mathbf{D}_{i_1,i_2,j}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\mathbf{r}_{i_2}\mathbf{r}_{i_2}^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\breve{\beta}_{i_1,i_2,j}(z_2)\mathbf{x}_p|^2)^{1/2}\\
&\quad\times(\mathrm{E}|\mathbf{x}_p^*\breve{\mathbf{D}}_{i_1j}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_{i_1}\mathbf{r}_{i_1}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_1j}^{-1}(z_1)\mathbf{x}_p|^4)^{1/4}\\
&\quad\times(\mathrm{E}|\mathbf{x}_p^*\breve{\mathbf{D}}_{i_2j}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_{i_2}\mathbf{r}_{i_2}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_2j}^{-1}(z_1)\mathbf{x}_p|^4)^{1/4}
\end{aligned}$$

$$= O(n^{-5/2})$$

and, by (9.8.6),

$$\begin{aligned}
&|\mathrm{E}\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{i_1,i_2,j}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{i_2j}^{-1}(\bar{z}_1)\mathbf{T}\breve{\mathbf{D}}_{i_2j}^{-1}(\bar{z}_2)\mathbf{x}_p\\
&\times\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\mathbf{r}_{i_2}\mathbf{r}_{i_2}^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\breve{\beta}_{i_1,i_2,j}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_{i_1}\mathbf{r}_{i_1}^* - n^{-1}\mathbf{T})\\
&\quad\mathbf{D}_{i_1j}^{-1}(z_1)\mathbf{x}_p\mathbf{x}_p^*\breve{\mathbf{D}}_{i_2j}^{-1}(\bar{z}_2)\mathbf{T}\mathbf{H}^{-1}(\bar{z}_1)(\mathbf{r}_{i_2}\mathbf{r}_{i_2}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_2j}^{-1}(\bar{z}_1)\mathbf{x}_p|\\
&\le K(\mathrm{E}|\mathbf{x}_p^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\mathbf{r}_{i_2}\mathbf{r}_{i_2}^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\breve{\beta}_{i_1,i_2,j}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_{i_1}\mathbf{r}_{i_1}^* - n^{-1}\mathbf{T})\\
&\times\mathbf{D}_{i_1j}^{-1}(z_1)\mathbf{x}_p|^2)^{1/2}(\mathrm{E}|\mathbf{x}_p^*\breve{\mathbf{D}}_{i_2j}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_{i_2}\mathbf{r}_{i_2}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_2j}^{-1}(z_1)\mathbf{x}_p|^2)^{1/2}\\
&\le K(\mathrm{E}|\mathbf{x}_p^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\mathbf{r}_{i_2}\mathbf{r}_{i_2}^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\\
&\quad\times(\mathbf{r}_{i_1}\mathbf{r}_{i_1}^* - n^{-1}\mathbf{T})\mathbf{D}_{i_1j}^{-1}(z_1)\mathbf{x}_p|^2)^{1/2}\times O(n^{-1})\\
&\le K(\mathrm{E}\mathbf{x}_p^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\mathbf{r}_{i_2}\mathbf{r}_{i_2}^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{T}\mathbf{H}^{-1}(\bar{z}_1)\mathbf{T}\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(\bar{z}_2)\mathbf{r}_{i_2}\\
&\quad\times\mathbf{r}_{i_2}^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(\bar{z}_2)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_{i_1,j}^{-1}(\bar{z}_1)\mathbf{T}\mathbf{D}_{i_1j}^{-1}(z_1)\mathbf{x}_p)^{1/2}\times O(n^{-2})\\
&\le K(\mathrm{E}|\mathbf{x}_p^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\mathbf{r}_{i_2}\mathbf{r}_{i_2}^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(\bar{z}_2)\mathbf{x}_p|^2)^{1/4}\\
&\quad\times(\mathrm{E}|\mathbf{r}_{i_2}^*\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{T}\mathbf{H}^{-1}(\bar{z}_1)\mathbf{T}\breve{\mathbf{D}}_{i_1,i_2,j}^{-1}(\bar{z}_2)\mathbf{r}_{i_2}^2)^{1/4}\times O(n^{-2})\\
&= O(n^{-9/4}).
\end{aligned}$$

Then, the conclusion (10.7.18) follows from the three estimates above.

Therefore,

$$\begin{aligned}
\mathbf{A}_3(z_1,z_2) &= \mathbf{A}_{32}(z_1,z_2) + o_p(1)\\
\mathbf{A}_{32}(z_1,z_2) &= -\sum_{i<j} b_{p1}(z_1)\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{x}_p
\end{aligned}$$

$$\begin{aligned}
&\quad\times\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\breve{\beta}_{ij}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)(\mathbf{r}_i\mathbf{r}_i^* - n^{-1}\mathbf{T})\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p\\
&= -\sum_{i<j} b_{p1}(z_1)\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{x}_p\\
&\quad\times\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\breve{\beta}_{ij}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p + o_p(1).
\end{aligned}$$

By (6.2.36), (10.7.3), and (10.7.5), for $i<j$, we have

$$\begin{aligned}
&\mathrm{E}\Big|\mathbf{x}_p^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p\Big(\mathbf{r}_i^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\breve{\beta}_{ij}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{r}_i \qquad (10.7.19)\\
&\quad -n^{-1}b_{p1}(z_2)\mathrm{tr}\mathbf{T}\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\Big)\Big|\\
&\le (\mathrm{E}|\mathbf{x}_p^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p^2)^{1/2}\Big[(\mathrm{E}|\breve{\beta}_{ij}(z_2)|^2|\mathbf{r}_i^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{r}_i\\
&\quad -n^{-1}\mathrm{tr}\mathbf{T}\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)|^2)^{1/2}
\end{aligned}$$

$$+(\mathrm{E}|\breve{\beta}_{ij}(z_2)-b_{p1}(z_2)|^2|n^{-1}\mathrm{tr}\mathbf{T}\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)|^2)^{1/2}\Big]$$
$$=O(n^{-3/2}).$$

Collecting the proofs from (10.7.13) to (10.7.19), we have shown that

$$\mathbf{A}(z_1,z_2)=-b_{p1}(z_1)b_{n1}(z_2)\sum_{i<j}\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}$$
$$\times\left(\mathbf{x}_p^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p n^{-1}\mathrm{tr}\mathbf{T}\mathrm{tr}\mathbf{T}\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\right)+o_p(1).$$

Similar to the proof of (10.7.18), we may further replace $\mathbf{r}_i\mathbf{r}_i^*$ in the expression above by $n^{-1}\mathbf{T}$; that is

$$\mathbf{A}(z_1,z_2)=-b_{p1}(z_1)b_{p1}(z_2)n^{-2}\sum_{i<j}\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}$$
$$\times\left(\mathbf{x}_p^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{T}\mathbf{D}_{ij}^{-1}(z_1)\mathbf{x}_p\mathrm{tr}\mathbf{T}\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\right)+o_p(1).$$

Reversing the procedure above, one finds that we may also replace $\mathbf{D}_{ij}^{-1}(z_1)$ and $\breve{\mathbf{D}}_{ij}^{-1}(z_2)$ in $\mathbf{A}(z_1,z_2)$ by $\mathbf{D}_j^{-1}(z_1)$ and $\mathbf{D}_j^{-1}(z_2)$, respectively; that is,

$$\mathbf{A}(z_1,z_2)=-\frac{b_{n1}(z_1)b_{n1}(z_2)(j-1)}{n^2}\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}$$
$$\times\left(\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{D}_j^{-1}(z_1)\mathbf{x}_p\mathrm{tr}\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\right)+o_p(1).$$

Using the martingale decomposition (10.7.11), one can further show that

$$\mathbf{A}(z_1,z_2)=-\frac{b_{p1}(z_1)b_{p1}(z_2)(j-1)}{n^2}\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p$$
$$\times\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{D}_j^{-1}(z_1)\mathbf{x}_p\mathrm{E}_{j-1}\mathrm{tr}\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)+o_p(1).$$

It is easy to verify that

$$n^{-1}\mathrm{tr}(\mathbf{T}M(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1))=o_p(1)$$

when $M(z_2)$ is either $\breve{\mathbf{A}}(z_2)$, $\breve{\mathbf{B}}(z_2)$, or $\breve{\mathbf{C}}(z_2)$. Thus, substituting the decomposition (9.9.12) for $\breve{\mathbf{D}}_j^{-1}(z_2)$ in the approximation above for $\mathbf{A}(z_1,z_2)$, one finds that

$$\mathbf{A}(z_1,z_2)=\tfrac{b_{p1}(z_1)b_{p1}(z_2)(j-1)}{n^2}\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)$$
$$\times\mathbf{T}\mathbf{D}_j^{-1}(z_1)\mathbf{x}_p\mathrm{E}_{j-1}\mathrm{tr}\mathbf{T}\mathbf{H}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)+o_p(1). \qquad (10.7.20)$$

Finally, let us consider the first term of (10.7.13). Using the expression for $\mathbf{D}_j^{-1}(z_1)$ in (9.9.12), we get

$$-\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{x}_p$$
$$= W_1(z_1,z_2)+W_2(z_1,z_2)+W_3(z_1,z_2)+W_4(z_1,z_2), \qquad (10.7.21)$$

where

$$\begin{aligned}
W_1(z_1,z_2) &= \mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{H}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{x}_p,\\
W_2(z_1,z_2) &= -b_{p1}(z_1)\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{A}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p\\
&\quad\times\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{x}_p,\\
W_3(z_1,z_2) &= -\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{B}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{x}_p,
\end{aligned}$$

and

$$W_4(z_1,z_2) = -\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{C}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{x}_p.$$

By the same argument as in (10.7.14), one can get

$$\mathrm{E}|W_3(z_1,z_2)| = o(1) \quad \text{and} \quad \mathrm{E}|W_4(z_1,z_2)| = o(1). \qquad (10.7.22)$$

Furthermore, as when dealing with $\mathbf{A}(z_1,z_2)$, the first $\breve{\mathbf{D}}_j^{-1}(z_2)$ in $W_2(z_1,z_2)$ can be replaced by $-b_{p1}(z_2)\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)$; that is,

$$\begin{aligned}
&W_2(z_1,z_2)\\
&= b_{p1}(z_1)b_{p1}(z_2)\sum_{i<j}\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{H}^{-1}(z_1)(\mathbf{r}_i\mathbf{r}_i^*-n^{-1}\mathbf{T})\mathbf{D}_{ij}^{-1}(z_1)\mathbf{T}\\
&\quad\times\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{x}_p+o_p(1)\\
&= b_{p1}(z_1)b_{p1}(z_2)\sum_{i<j}\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{H}^{-1}(z_1)\mathbf{r}_i\mathbf{r}_i^*\mathbf{D}_{ij}^{-1}(z_1)\mathbf{T}\\
&\quad\times\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{r}_i\mathbf{r}_i^*\breve{\mathbf{D}}_{ij}^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{x}_p+o_p(1)\\
&= \frac{b_{p1}(z_1)b_{p1}(z_2)(j-1)}{n^2}\mathrm{E}_{j-1}\left(\mathbf{x}_p^*\mathbf{H}^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p\mathrm{tr}\mathbf{T}\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\right)\\
&\quad\times\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{x}_p+o_p(1).
\end{aligned}$$

It can also be verified that

$$\mathbf{x}_p^*M(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{x}_p = o_p(1)$$

when $M(z_2)$ takes $\breve{\mathbf{A}}(z_2)$, $\breve{\mathbf{B}}(z_2)$, or $\breve{\mathbf{C}}(z_2)$. Therefore, $W_2(z_1,z_2)$ can be further approximated by

$$\begin{aligned}
W_2(z_1,z_2) &= \frac{b_{p1}(z_1)b_{p1}(z_2)(j-1)}{n^2}\left(\mathbf{x}_p^*\mathbf{H}^{-1}(z_1)\mathbf{T}\mathbf{H}^{-1}(z_2)\mathbf{x}_p\right)^2\\
&\quad\times\mathrm{E}_{j-1}\mathrm{tr}(\mathbf{T}\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}^{-1}(z_2))+o_p(1). \qquad (10.7.23)
\end{aligned}$$

In (9.9.21), it is proven that

$$\begin{aligned}
&\mathrm{E}_{j-1}\mathrm{tr}(\mathbf{T}\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)) \\
&= \frac{\mathrm{tr}(\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{T}\mathbf{H}^{-1}(z_2)) + o_p(1)}{1 - \frac{j-1}{n^2} z_1 z_2 \underline{s}(z_1)\underline{s}(z_2)\mathrm{tr}(\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{T}\mathbf{H}^{-1}(z_2))}.
\end{aligned}$$

In the same way, $W_1(z_1, z_2)$ can be approximated by

$$\begin{aligned}
&W_1(z_1, z_2) \\
&= \mathbf{x}_p^*\mathbf{H}^{-1}(z_1)\mathbf{T}\mathbf{H}^{-1}(z_2)\mathbf{x}_p\mathbf{x}_p^*\mathbf{H}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{x}_p + o_p(1). \qquad (10.7.24)
\end{aligned}$$

Consequently, from (10.7.13)–(10.7.24), we obtain

$$\begin{aligned}
&\mathrm{E}_{j-1}\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{T}\mathrm{E}_j(\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{x}_p\mathrm{E}_{j-1}\mathbf{x}_p^*\breve{\mathbf{D}}_j^{-1}(z_2))\mathbf{T}\mathbf{D}_j^{-1}(z_1)\mathbf{x}_p \\
&\times\Big[1 - \frac{j-1}{n} b_{p1}(z_1)b_{p1}(z_2)\frac{1}{n}\mathrm{tr}\mathbf{H}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{T}\Big] \\
&= \mathbf{x}_p^*\mathbf{H}^{-1}(z_1)\mathbf{T}\mathbf{H}^{-1}(z_2)\mathbf{x}_p\mathbf{x}_p^*\mathbf{H}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{x}_p \\
&\Big(1 + \frac{j-1}{n} b_{p1}(z_1)b_{p1}(z_2)\frac{1}{n}\mathrm{E}_{j-1}\mathrm{tr}(\mathbf{D}_j^{-1}(z_1)\mathbf{T}\breve{\mathbf{D}}_j^{-1}(z_2)\mathbf{T})\Big) + o_p(1). \qquad (10.7.25)
\end{aligned}$$

Recall that $b_{p1}(z) \to -z\underline{s}(z)$ and $F^{\mathbf{T}} \to H$. Hence,

$$\begin{aligned}
d(z_1, z_2) &:= \lim b_{p1}(z_1)b_{p1}(z_2)\frac{1}{N}\mathrm{tr}(\mathbf{H}^{-1}(z_1)\mathbf{T}\mathbf{H}^{-1}(z_2)\mathbf{T}) \\
&= \int \frac{ct^2\underline{s}(z_1)\underline{s}(z_2)}{(1 + t\underline{s}(z_1))(1 + t\underline{s}(z_2))} dH(t) \\
&= 1 + \frac{\underline{s}(z_1)\underline{s}(z_2)(z_1 - z_2)}{\underline{s}(z_2) - \underline{s}(z_1)}. \qquad (10.7.26)
\end{aligned}$$

By the conditions of Theorem 10.21,

$$\begin{aligned}
h(z_1, z_2) &= \lim z_1 z_2 \underline{s}(z_1)\underline{s}(z_2)\mathbf{x}_p^*\mathbf{H}^{-1}(z_1)\mathbf{T}\mathbf{H}^{-1}(z_2)\mathbf{x}_p\mathbf{x}_p^*\mathbf{H}^{-1}(z_2)\mathbf{T}\mathbf{H}^{-1}(z_1)\mathbf{x}_p \\
&= \frac{\underline{s}(z_1)\underline{s}(z_2)}{z_1 z_2}\left(\int \frac{t^2\underline{s}(z_1)\underline{s}(z_2)}{(1 + t\underline{s}(z_1))(1 + t\underline{s}(z_2))} dH(t)\right)^2 \\
&= \frac{\underline{s}(z_1)\underline{s}(z_2)}{z_1 z_2}\left(\int \frac{t dH(t)}{(1 + t\underline{s}(z_1))(1 + t\underline{s}(z_2))}\right)^2 \\
&= \frac{\underline{s}(z_1)\underline{s}(z_2)}{z_1 z_2}\left(\frac{z_1 m(z_1) - z_2 m(z_2)}{(\underline{s}(z_2) - \underline{s}(z_1))}\right)^2. \qquad (10.7.27)
\end{aligned}$$

From (10.7.13) and (10.7.25)–(10.7.27), we get

$$(10.7.10) \xrightarrow{i.p.} h(z_1, z_2)\left(\int_0^1 \frac{1}{(1 - td(z_1, z_2))} dt + \int_0^1 \frac{td(z_1, z_2)}{(1 - td(z_1, z_2))^2} dt\right)$$

$$= \frac{h(z_1, z_2)}{1 - d(z_1, z_2)} = \frac{(z_2\underline{s}(z_2) - z_1\underline{s}(z_1))^2}{c^2 z_1 z_2 (z_2 - z_1)(\underline{s}(z_2) - \underline{s}(z_1))}.$$

10.7.3 Tightness of $M_n^1(z)$ and Convergence of $M_n^2(z)$

First, we proceed to the proof of the tightness of $M_n^1(z)$ by verifying that

$$\mathrm{E}\frac{|M_n^1(z_1) - M_n^1(z_2)|^2}{|z_1 - z_2|^2} \le K \quad \text{if} \quad z_1, z_2 \in \mathcal{C}. \tag{10.7.28}$$

By (10.7.4), we obtain

$$\mathrm{E}\left|\sum_{j=1}^n Y_j(z_i)\right|^2 = \sum_{j=1}^n \mathrm{E}|Y_j(z_i)|^2 \le K.$$

Therefore, we only need to consider z_1, z_2 when they are close to each other. Write

$$Q(z_1, z_2) = \sqrt{n}\mathbf{x}_p(\mathbf{D} - z_1\mathbf{I})^{-1}(\mathbf{D} - z_2\mathbf{I})^{-1}\mathbf{x}_p.$$

Recalling the definition of M_n^1, we have

$$\frac{|M_n^1(z_1) - M_n^1(z_2)|}{|z_1 - z_2|}$$
$$\le \begin{cases} |Q(z_1, z_2) - \mathrm{E}Q(z_1, z_2)|, & \text{if } \min(|\Im(z_1)|, |\Im(z_2)|) \ge \delta_p p^{-1}, \\ \sum_{*=+,-} |Q(z_1, z_*) - \mathrm{E}Q(z_1, z_*)|, & \text{if } |\Im(z_1)| \ge \delta_p p^{-1} \ \ \& \ \ |\Im(z_2)| \le \delta_p p^{-1}, \\ \sum_{*=+,-} |Q(z_2, z_*) - \mathrm{E}Q(z_2, z_*)|, & \text{if } |\Im(z_2)| \ge \delta_p p^{-1} \ \ \& \ \ |\Im(z_1)| \le \delta_p p^{-1}, \\ |Q(z_+, z_-) - \mathrm{E}Q(z_+, z_-)|, & \text{otherwise}, \end{cases}$$

where $\Re(z_\pm) = u_r$ or u_l, and $\Im(z_\pm) = \pm\delta_p p^{-1}$. We only give a proof of (10.7.28) for the case $\Im(z_1), \Im(z_2) \ge \delta_p p^{-1}$.

From the identity above (9.10.7), we get

$$\frac{M_n^1(z_1) - M_n^1(z_2)}{z_1 - z_2} = \sqrt{n}\sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1})\mathbf{x}_p^*\mathbf{D}^{-1}(z_1)\mathbf{D}^{-1}(z_2)\mathbf{x}_p \tag{10.7.29}$$

$$= V_1(z_1, z_2) + V_2(z_1, z_2) + V_3(z_1, z_2),$$

where

$$V_1(z_1, z_2) = \sqrt{n}\sum_{j=1}^n (\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_1)\beta_j(z_2)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\times$$

$$\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_2)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{r}_j,$$

$$V_2(z_1,z_2) = \sqrt{n}\sum_{j=1}^{n}(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_1)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{r}_j,$$

and

$$V_3(z_1,z_2) = \sqrt{n}\sum_{j=1}^{n}(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_2)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_2)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j.$$

Applying (10.7.1) and the bounds for $\beta_j(z)$ and $\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j$ argued below (9.10.2), we obtain

$$\begin{aligned} &\mathrm{E}|V_1(z_1,z_2)|^2 \qquad (10.7.30)\\ &= n\sum_{j=1}^{n}\mathrm{E}|(\mathrm{E}_j - \mathrm{E}_{j-1})\beta_j(z_1)\beta_j(z_2)\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_1)\mathbf{D}_j^{-1}(z_2)\mathbf{r}_j\\ &\qquad \mathbf{r}_j^*\mathbf{D}_j^{-1}(z_2)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{r}_j|^2\\ &\le Kn^2(\mathrm{E}|\mathbf{r}_j^*\mathbf{D}_j^{-1}(z_2)\mathbf{x}_p\mathbf{x}_p^*\mathbf{D}_j^{-1}(z_1)\mathbf{r}_j|^2 + v^{-12}n^2P(\|\mathbf{B}_p\| > u_r \ or \lambda_{\min}^{\mathbf{B}_p} < u_l))\\ &\le K. \end{aligned}$$

Here, to derive the inequality above, we have also used (9.7.8) and (9.7.9). It is easy to see that (9.7.8) and (9.7.9) hold for our truncated variables as well. Similarly, the argument above can also be used to handle $V_2(z_1,z_2)$ and $V_3(z_1,z_2)$. Therefore, we have completed the proof of (10.7.28).

Next we will consider the convergence of $M_n^2(z)$. Note that

$$s_{F^{y_p,H_p}}(z) = -\frac{1}{z}\int\frac{1}{1+t\underline{s}_{F^{y_p,H_p}}(z)}dH_p(t). \qquad (10.7.31)$$

Substituting (10.6.7) into (10.7.31), we obtain

$$\begin{aligned} &z\sqrt{n}((\mathbf{x}_p^*\mathrm{E}(\mathbf{D}^{-1})(z)\mathbf{x}_p - s_{F^{y_p,H_p}}(z))) \qquad (10.7.32)\\ &= \sqrt{p}(\mathbf{x}_p^*(-\mathrm{E}\underline{s}_p(z)\mathbf{T}-\mathbf{I})^{-1}\mathbf{x}_p + \int\frac{1}{1+t\underline{s}_{F^{y_p,H_p}}(z)}dH_p(t))\\ &\qquad +\sqrt{n}z(\delta_1+\delta_2+\delta_3+\delta_4). \end{aligned}$$

For δ_4, we have

$$|z\sqrt{n}\delta_4| \le K\sqrt{n}[\mathrm{E}|\xi_1^2|+v^{-5}n^{\frac{3}{2}}P(\|\mathbf{B}_p\| > u_r \ or \lambda_{\min}^{\mathbf{B}_p} < u_l)] = o(1). \quad (10.7.33)$$

Here and in the following, we use the notation $o(\cdot)$ for a limit uniform in $z \in \mathcal{C}$. Similarly, by (10.7.5) and the Hölder inequality, it is straightforward to verify that

$$|z\sqrt{n}\delta_3| \to 0.$$

Appealing to (9.8.6), whether for the complex case or the real case, we have

$$\mathrm{E}[\xi_1(z)(\mathbf{r}_1^*\mathbf{D}_1^{-1}(z)\mathbf{x}_p\mathbf{x}_p^*(\mathrm{E}\underline{s}_p(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{r}_1 \\ -n^{-1}\mathbf{x}_p^*(\mathrm{E}\underline{s}_p(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{T}\mathbf{D}_1^{-1}(z)\mathbf{x}_p)] = O(n^{-2}).$$

It follows that

$$|n^{\frac{3}{2}}\mathrm{E}[b_1(z)\beta_1(z)\xi_1(z)\alpha_1(z)]| \\ \le n^{\frac{3}{2}}|\mathrm{E}[b_1^2(z)\xi_1(z)\alpha_1(z)]| + n^{\frac{3}{2}}|\mathrm{E}[b_1^2(z)\beta_1(z)\xi_1^2(z)\alpha_1(z)]| = o(1), \quad (10.7.34)$$

where we have used the fact that

$$\mathrm{E}[b_1^2(z)\xi_1(z)\alpha_1(z)] = \mathrm{E}\{b_1^2(z)\mathrm{E}[\xi_1(z)\alpha_1(z)|\sigma(\mathbf{r}_j, j>1)]\}$$

and that

$$|\mathrm{E}[b_1^2(z)\beta_1(z)\xi_1^2(z)\alpha_1(z)]| \\ \le K[(\mathrm{E}|\xi_1(z)|^4\mathrm{E}|\alpha_1(z)|^2)^{\frac{1}{2}} + v^{-6}p^2P(\|\mathbf{B}\| > u_r \ or \lambda_{\min}^{\mathbf{B}} < u_l)] = O(n^{-2}).$$

Then (10.7.33) and (10.7.34) give

$$|z\sqrt{n}\delta_4| \to 0.$$

According to Section 9.11, we have

$$\sup_z \sqrt{n}(\underline{s}_{F^{y_p},H_p}(z) - \mathrm{E}\underline{s}_p(z)) \to 0. \quad (10.7.35)$$

Following lines similar to (9.11.6) and (9.11.5), we get

$$\sup_{p,z\in\mathcal{C}} \|(\underline{s}_{F^{y_p},H_p}(z)\mathbf{T}+\mathbf{I})^{-1}\| < \infty.$$

It follows that, via (9.11.6) and the assumption of Theorem 10.21,

$$\sup_{p,z\in\mathcal{C}_0} \int \frac{t}{(1+t\underline{s}_{F^{y_p},H_p}(z))(t\mathrm{E}\underline{s}_p(z)+1)} dH_p(t) < \infty. \quad (10.7.36)$$

Appealing to (9.11.4), (9.11.6), (10.7.35), and (10.7.36), we can conclude that

$$\sqrt{n}\left(\mathbf{x}_p^*(\mathrm{E}\underline{s}_p(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{x}_p - \int \frac{1}{\mathrm{E}\underline{s}_p(z)t+1} dH_p(t)\right)$$

$$= \sqrt{n}\left(\mathbf{x}_p^*(\underline{s}_{F^{y_p,H_p}}(z)\mathbf{T}+\mathbf{I})^{-1}\mathbf{x}_p - \int \frac{1}{\underline{s}_{F^{y_p,H_p}}(z)t+1}dH_p(t)\right) + o(1).$$

Using (10.7.35) and (10.7.36), we also have

$$\sqrt{n}\left(\int \frac{1}{\mathrm{E}\underline{s}_p(z)t+1}dH_p(t) - \int \frac{1}{1+t\underline{s}_{F^{y_p,H_p}}(z)}dH_p(t)\right)$$

$$= \sqrt{n}(\underline{s}_{F^{y_p,H_p}}(z) - \mathrm{E}\underline{s}_p(z))\int \frac{t}{(1+t\underline{s}_{F^{y_p,H_p}}(z))(t\mathrm{E}\underline{s}_p(z)+1)}dH_p(t) = o(1).$$

Combining the arguments above, we get that

$$(10.7.32) \to 0.$$

10.8 Proof of Theorem 10.23

In this part, when $\mathbf{T} = \mathbf{I}$, we will prove that (10.5.31) is a consequence of (10.5.29) under the conditions of Theorem 10.23.

Note that

$$z = -\frac{1}{\underline{s}(z)} + y\int \frac{t}{1+t\underline{s}(z)}dH(t). \tag{10.8.1}$$

It then follows that

$$z_1 - z_2 = \frac{(\underline{s}(z_1) - \underline{s}(z_2))}{\underline{s}(z_1)\underline{s}(z_2)}\left(1 - y\int \frac{t^2\underline{s}(z_1)\underline{s}(z_2)dH(t)}{(1+\underline{s}(z_1))(1+\underline{s}(z_2))}\right) \tag{10.8.2}$$

and

$$z_2\underline{s}(z_2) - z_1\underline{s}(z_1) = y(\underline{s}(z_2) - \underline{s}(z_1))\int \frac{tdH(t)}{(1+t\underline{s}(z_1))(1+t\underline{s}(z_2))}. \tag{10.8.3}$$

Combining (10.8.1)–(10.8.3), (10.5.29) equals

$$= -\frac{1}{2\pi^2}\int_{\mathcal{C}_1}\int_{\mathcal{C}_2} \frac{g_1(z_1)g_2(z_2)\underline{s}^2(z_1)\underline{s}^2(z_2)}{\int \frac{(ct\underline{s}(z_1)-1-t\underline{s}(z_1))}{1+t\underline{s}(z_1)}dH(t)\int \frac{(ct\underline{s}(z_2)-1-t\underline{s}(z_2))}{1+t\underline{s}(z_2)}dH(t)}$$

$$\times \frac{\left(\int \frac{tdH(t)}{(1+t\underline{s}(z_1))(1+t\underline{s}(z_2))}\right)^2 dz_1dz_2}{\left(1 - c\int \frac{\underline{s}(z_1)\underline{s}(z_2)t^2}{(1+t\underline{s}(z_1))(1+t\underline{s}(z_2))}dH(t)\right)}, \tag{10.8.4}$$

where the contours are defined as before.

On the other hand, by Cauchy's theorem, we get that

$$(10.5.31) = -\frac{1}{2c\pi^2}\int_{\mathcal{C}_1}\int_{\mathcal{C}_2} g_1(z_1)g_2(z_2)\Big(\frac{s(z_1)-s(z_2)}{z_1-z_2} - s(z_1)s(z_2)\Big)dz_1dz_2. \tag{10.8.5}$$

Applying (10.8.1)–(10.8.3) and the relation between $\underline{s}(z)$ and $s(z)$ given above (9.7.1), we get that

$$\Big(\frac{s(z_1)-s(z_2)}{z_1-z_2} - s(z_1)s(z_2)\Big) = \frac{M_1 + M_2\underline{s}(z_1)\underline{s}(z_2)}{1-y\int\frac{\underline{s}(z_1)\underline{s}(z_2)t^2}{(1+t\underline{s}(z_1))(1+t\underline{s}(z_2))}dH(t)},$$

where

$$M_1 = \frac{y\underline{s}^2(z_1)\underline{s}^2(z_2)\int\frac{dH(t)}{1+t\underline{s}(z_1)}\int\frac{dH(t)}{1+t\underline{s}(z_1)}\int\frac{t^2dH(t)}{(1+t\underline{s}(z_1))(1+t\underline{s}(z_2))}}{\int\frac{(yt\underline{s}(z_1)-1-t\underline{s}(z_1))}{1+t\underline{s}(z_1)}dH(t)\int\frac{(yt\underline{s}(z_2)-1-t\underline{s}(z_2))}{1+t\underline{s}(z_2)}dH(t)}$$

and

$$M_2 = \frac{\int\frac{(yt\underline{s}(z_1)-1-t\underline{s}(z_1))}{1+t\underline{s}(z_1)}dH(t)\int\frac{(yt\underline{s}(z_2)-1-t\underline{s}(z_2))}{1+t\underline{s}(z_2)}dH(t) - y\int\frac{dH(t)}{1+t\underline{s}(z_1)}\int\frac{dH(t)}{1+t\underline{s}(z_2)}}{y\int\frac{(yt\underline{s}(z_1)-1-t\underline{s}(z_1))}{1+t\underline{s}(z_1)}dH(t)\int\frac{(yt\underline{s}(z_2)-1-t\underline{s}(z_2))}{1+t\underline{s}(z_2)}dH(t)}.$$

Observe that

$$\begin{aligned}&\int\frac{(yt\underline{s}(z_1)-1-t\underline{s}(z_1))}{1+t\underline{s}(z_1)}dH(t)\int\frac{(yt\underline{s}(z_2)-1-t\underline{s}(z_2))}{1+t\underline{s}(z_2)}dH(t)\\ &\qquad -y\int\frac{dH(t)}{1+t\underline{s}(z_1)}\int\frac{dH(t)}{1+t\underline{s}(z_2)}\\ &= (1-y)\Big(1-y\int\frac{\underline{s}(z_1)\underline{s}(z_2)t^2}{(1+t\underline{s}(z_1))(1+t\underline{s}(z_2))}dH(t)\Big),\end{aligned}$$

which implies that

$$\begin{aligned}&\frac{1}{2y\pi^2}\int_{\mathcal{C}_1}\int_{\mathcal{C}_2} g_1(z_1)g_2(z_2)\frac{M_2\underline{s}(z_1)\underline{s}(z_2)}{1-y\int\frac{\underline{s}(z_1)\underline{s}(z_2)t^2}{(1+t\underline{s}(z_1))(1+t\underline{s}(z_2))}dH(t)}dz_1dz_2\\ &= \frac{1}{2\pi^2}\int_{\mathcal{C}_1}\int_{\mathcal{C}_2} g_1(z_1)g_2(z_2)\frac{(1-y)}{y^2z_1z_2}dz_1dz_2 = 0\end{aligned} \tag{10.8.6}$$

since the z_1 and z_2 contours do not enclose the origin. Note that

$$\begin{aligned}&\int\frac{dH(t)}{1+t\underline{s}(z_1)}\int\frac{dH(t)}{1+t\underline{s}(z_2)}\int\frac{\underline{s}(z_1)\underline{s}(z_2)t^2dH(t)}{(1+t\underline{s}(z_1))(1+t\underline{s}(z_2))}\\ &-\int\frac{\underline{s}(z_2)tdH(t)}{(1+t\underline{s}(z_1))(1+t\underline{s}(z_2))}\int\frac{\underline{s}(z_1)tdH(t)}{(1+t\underline{s}(z_1))(1+t\underline{s}(z_2))}\end{aligned} \tag{10.8.7}$$

$$= \left(\int \frac{dH(t)}{(1 + t\underline{s}(z_1))(1 + t\underline{s}(z_2))} - \int \frac{dH(t)}{(1 + t\underline{s}(z_1))} - \int \frac{dH(t)}{(1 + t\underline{s}(z_2))} \right)$$

$$\times \left(\int \frac{dH(t)}{(1 + t\underline{s}(z_1))} \int \frac{dH(t)}{(1 + t\underline{s}(z_2))} - \int \frac{dH(t)}{(1 + t\underline{s}(z_1))(1 + t\underline{s}(z_2))} \right). \quad (10.8.8)$$

Here one should note that the first item of the difference (10.8.7) is a factor of M_1 and that the second item is a factor of (10.8.4). Consequently, combining (10.8.4)–(10.8.8) and condition (10.5.30) (for identity matrix, the condition (10.5.30) holds automatically), one can conclude that

$$(10.5.29) = (10.5.31).$$

Thus we are done.

Chapter 11
Circular Law

11.1 The Problem and Difficulty

This is a famous conjecture that has been open for more than half a century. At present, only some partial answers are known. The conjecture is stated as follows. Suppose that $\mathbf{X}_n$ is an $n \times n$ matrix with entries x_{kj}, where $\{x_{kj},\ k,j = 1,2,\cdots\}$ forms an infinite double array of iid complex random variables of mean zero and variance one. Using the complex eigenvalues $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n$ of $\frac{1}{\sqrt{n}}\mathbf{X}_n$, we can construct a two-dimensional empirical distribution by

$$\mu_n(x,y) = \frac{1}{n}\#\left\{i \le n : \Re(\lambda_k) \le x,\ \Im(\lambda_k) \le y\right\},$$

which is called the empirical spectral distribution of the matrix $\frac{1}{\sqrt{n}}\mathbf{X}_n$.

Since the early 1950s, it has been conjectured that the distribution $\mu_n(x,y)$ converges to the so-called circular law; i.e., the uniform distribution over the unit disk in the complex plane. The first answer was given by Mehta [212] for the case where x_{ij} are iid complex normal variables. He used the joint density function of the eigenvalues of the matrix $\frac{1}{\sqrt{n}}\mathbf{X}_n$, which was derived by Ginibre [120]. The joint density is

$$p(\lambda_1,\cdots,\lambda_n) = c_n \prod_{i<j}|\lambda_i - \lambda_j|^2 \prod_{i=1}^{n} e^{-n|\lambda_i|^2},$$

where λ_i, $i \le n$, are the complex eigenvalues of the matrix $\frac{1}{\sqrt{n}}\mathbf{X}_n$ and c_n is a normalizing constant.

Partial answers under more general assumptions are made in Girko [123, 124] and Bai [14]. The problem under the only condition of finite second moment is still open. For details of the history of this problem, the reader is referred to Bai [14]. Some recent developments are given in Section 11.10.

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_11,

11.1.1 Failure of Techniques Dealing with Hermitian Matrices

In this section, we show that the methodologies to deal with Hermitian matrices do not apply to non-Hermitian matrices.

1. Failure of truncation method

It was seen in previous chapters that a small change to all entries or a large change to a small number of entries of Hermitian matrices will cause a small change in their empirical spectral distributions, and thus the truncation technique has played an important role in the spectral theory of large dimensional Hermitian matrices. However, it is not the case for non-Hermitian matrices. See the following example.

Example 11.1. Consider the following two $n \times n$ matrices:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ \frac{1}{n^3} & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

It is easy to see that all the n eigenvalues of $\mathbf{A}$ are 0, while those of $\mathbf{B}$ are

$$\lambda_k = n^{-3/n} e^{2k\pi/n}, \quad k = 0, 1, \cdots, n-1.$$

When n is large, $|\lambda_k| = n^{-3/n} \sim 1$. This example shows that the empirical spectral distributions of $\mathbf{A}$ and $\mathbf{B}$ are very different, although they only have one different entry, which is as small as n^{-3}. Therefore, the truncation technique does not apply.

2. Failure of moment method

Although the moment method has successively been used as a powerful tool in establishing the spectral theory of Hermitian (symmetric) large matrices, it fails to apply to nonsymmetric matrices. The reason can be seen from the following fact. For any complex random variable Z, its distribution can be uniquely determined by all mixed moments $\mathrm{E}Z^k\overline{Z}^\ell$, $k, \ell \geq 0$, with some additional Carleman type conditions. However, for any square matrix $\mathbf{X}$ with order larger than 1, there is no way to find a simple functional of $\mathbf{X}$ that gives

$$\frac{1}{n}\sum_{j=1}^{n} \lambda_j^k(\mathbf{X})\bar{\lambda}_j^\ell(\mathbf{X}) \tag{11.1.1}$$

unless k or ℓ is 0. Knowledge of just the latter is not sufficient. Indeed, if we only know

$$\mathrm{E}Z^k = 0, \tag{11.1.2}$$

then this relation is satisfied by any complex random variable Z that is uniformly distributed over any disk with center 0. Thus, even if we had proved

$$\frac{1}{n}\mathrm{tr}\left(\frac{1}{\sqrt{n}}\mathbf{X}_n\right)^k \to 0, \quad \text{a.s.}, \tag{11.1.3}$$

we could not claim that the spectral distribution of $\frac{1}{\sqrt{n}}\mathbf{X}_n$ tends to the circular law because (11.1.2) does not uniquely determine the distribution of Z.

3. Difficulty of the method of Stieltjes transform
When $|z| > 1$, by Theorem 5.18, we may write

$$s_n(z) =: \frac{1}{n}\mathrm{tr}\left(\frac{1}{\sqrt{n}}\mathbf{X}_n - z\right)^{-1} = -\frac{1}{z}\left(1 + \sum_{k=1}^{\infty}\frac{1}{z^k}\left(\frac{1}{n}\mathrm{tr}\left(\frac{1}{\sqrt{n}}\mathbf{X}_n^k\right)\right)\right). \tag{11.1.4}$$

By (11.1.3), we should have

$$s_n(z) \to -\frac{1}{z}, \quad \text{a.s.} \tag{11.1.5}$$

The limit is the same as the Stieltjes transform of any uniform distribution over a disk with center 0 and radius $\rho \le 1$. Although the Stieltjes transform of $\frac{1}{\sqrt{n}}\mathbf{X}_n$ can uniquely determine all eigenvalues of $\frac{1}{\sqrt{n}}\mathbf{X}_n$, even only using values at z with $|z| > 1$, limit (11.1.5) cannot determine convergence to the circular law.

Therefore, to establish the circular law by Stieltjes transforms, one has to consider the convergence of $s_n(z)$ for z with $|z| \le 1$. Unlike the case for Hermitian matrices, the Stieltjes transform $s_n(z)$ is not bounded for $|z| \le 1$, which is even impossible for any bound depending on n. This leads to serious difficulties in the mathematical analysis of $s_n(z)$.

11.1.2 Revisiting Stieltjes Transformation

Despite the difficulty shown in the last subsection, one usable piece of information is that the function $s_n(z)$ uniquely determines all eigenvalues of $\frac{1}{\sqrt{n}}\mathbf{X}_n$. We make some modification to it so that the resulting version is easier to deal with.

Denote the eigenvalues of $\frac{1}{\sqrt{n}}\mathbf{X}_n$ by

$$\lambda_k = \lambda_{kr} + i\lambda_{ki},\ k = 1, 2, \cdots, n,$$

and $z = s + it$. Then,

$$s_n(z) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{\lambda_k - z}. \tag{11.1.6}$$

Because $s_n(z)$ is analytic at all z except the n eigenvalues, the real (or imaginary) part can also determine the eigenvalues of $\frac{1}{\sqrt{n}}\mathbf{X}_n$. Write $s_n(z) = s_{nr}(z) + is_{ni}(z)$. Then we have

$$\begin{aligned} s_{nr}(z) &= \frac{1}{n}\sum_{k=1}^{n}\frac{\lambda_{kr} - s}{|\lambda_k - z|^2} \\ &= -\frac{1}{2n}\sum_{k=1}^{n}\frac{\partial}{\partial s}\log(|\lambda_k - z|^2) \\ &= -\frac{1}{2}\frac{\partial}{\partial s}\int_0^\infty \log x \nu_n(dx, z), \end{aligned} \tag{11.1.7}$$

where $\nu_n(\cdot, z)$ is the ESD of the Hermitian matrix $\mathbf{H}_n = (\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I})(\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I})^*$.

Now, let us consider the Fourier transform of the function $s_{nr}(z)$. We have

$$\begin{aligned} &-2\iint s_{nr}(z)e^{i(us+vt)}dtds \\ &= \iint e^{i(us+vt)}\frac{\partial}{\partial s}\int_0^\infty \log x\nu_n(dx, z)dtds \\ &= \frac{2}{n}\sum_{k=1}^{n}\iint \frac{s - \lambda_{kr}}{(\lambda_{kr} - s)^2 + (\lambda_{ki} - t)^2}e^{i(us+vt)}dtds \\ &= \frac{2}{n}\sum_{k=1}^{n}\iint \frac{s}{s^2 + t^2}e^{i(us+vt)+i(u\lambda_{kr}+v\lambda_{ki})}dtds. \end{aligned} \tag{11.1.8}$$

We note here that in (11.1.8) and throughout the following, when integration with respect to s and t is performed on unbounded domains, it is iterated, first with respect to t and then with respect to s. Fubini's theorem cannot be applied since $s/(s^2 + t^2)$ is not integrable on $\mathbb{R}^2$ (although it is integrable on bounded subsets of the plane).

Recalling the characteristic function of the Cauchy distribution, we have

$$\frac{1}{\pi}\int \frac{|s|}{s^2 + t^2}e^{itv}dt = e^{-|sv|}.$$

Therefore,

$$\int \Big[\int \frac{s}{s^2 + t^2}e^{i(us+vt)}dt\Big]ds$$

$$= \pi \int \operatorname{sgn}(s) e^{ius-|sv|} ds$$
$$= 2i\pi \int_0^\infty \sin(su) e^{-|v|s} ds$$
$$= \frac{2i\pi u}{u^2+v^2}.$$

Substituting the above into (11.1.8), we have

$$\iint e^{i(us+vt)} \frac{\partial}{\partial s} \int_0^\infty \log x \nu_n(dx,z) dt ds$$

$$= \frac{4i\pi u}{u^2+v^2} \iint \frac{1}{n} \sum_{k=1}^n e^{i(u\lambda_{kr}+t\lambda_{ki})}.$$

Therefore, we have established the following lemma.

Lemma 11.2. *For any $uv \neq 0$, we have*

$$c_n(u,v) =: \iint e^{iux+ivy} \mu_n(dx,dy)$$
$$= \frac{u^2+v^2}{4iu\pi} \iint \frac{\partial}{\partial s} \left[\int_0^\infty \ln x \nu_n(dx,z) \right] e^{ius+ivt} dt ds, \tag{11.1.9}$$

where $z = s+it$, $i = \sqrt{-1}$, and

$$\mu_n(x,y) = \frac{1}{n} \#\{k \le n : \lambda_{kr} \le x, \lambda_{ki} \le y\},$$

the empirical spectral distribution of $\frac{1}{\sqrt{n}}\mathbf{X}_n$.

Remark 11.3. The identity in the lemma was first given by Girko [123], which reveals a way toward proving the circular law conjecture.

If we assume the fourth moment of the underlying distribution is finite, then by Theorem 5.18, with probability 1, the family of distributions $\mu_n(x,y)$ is tight. In fact, by Lemma 11.6 given later, together with Theorem 3.7, one sees that the family of distributions $\mu_n(x,y)$ is also tight under only the finiteness of the second moment. Therefore, to prove the circular law, applying Lemma 11.2, one need only show that the right-hand side of (11.1.9) converges to its counterpart generated by the circular law.

Note that the function $\ln x$ is not bounded at both infinity and zero. Therefore, the convergence of the right-hand side of (11.1.9) cannot simply reduce to the convergence of ν_n. In view of Theorem 5.8, the upper limit of the inner integral does not pose a serious problem since the support of ν_n is bounded from the right by $(2+\varepsilon+|z|)^2$ under the assumption of a finite fourth moment. The most difficult part is in dealing with the lower limit of the integral.

11.2 A Theorem Establishing a Partial Answer to the Circular Law

We shall prove the following theorem by using Lemma 11.2.

Theorem 11.4. *Suppose that the underlying distribution of elements of $\mathbf{X}_n$ has finite $(2+\eta)$-th moment and that the joint distribution of the real and imaginary parts of the entries has a bounded density. Then, with probability 1, the ESD $\mu_n(x,y)$ of $\frac{1}{\sqrt{n}}\mathbf{X}_n$ tends to the uniform distribution over the unit disk in the complex plane.*

The proof of the theorem will be presented by showing that with probability one, $c_n(u,v) \to c(u,v)$ for every pair (u,v) such that $uv \neq 0$. The proof will be rather tedious. For ease of understanding, an outline of the proof is sketched as follows.

1. Reducing the range of integration. We need to reduce the range of integration to a finite rectangle, so that the dominated convergence theorem is applicable. As will be seen, the proof of the circular law reduces to showing that, for every large $A > 0$ and small $\varepsilon > 0$,

$$\begin{aligned} &\iint_T \left[\frac{\partial}{\partial s}\int_0^\infty \ln x \nu_n(dx,z)\right] e^{ius+ivt} ds dt \\ &\to \iint_T \left[\frac{\partial}{\partial s}\int_0^\infty \ln x \nu(dx,z)\right] e^{ius+ivt} ds dt, \end{aligned} \tag{11.2.1}$$

where $T = \{(s,t);\ |s| \le A, |t| \le A^3, |\sqrt{s^2+t^2}-1| \ge \varepsilon\}$ and $\nu(x,z)$ is the LSD of the sequence of matrices $\mathbf{H}_n = (\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I})(\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I})^*$ that determines the circular law. The rest of the proof will be divided into the following steps.

2. Identifying the limiting spectrum $\nu(\cdot,z)$ of $\nu_n(\cdot,z)$ and showing that it determines the circular law.

3. Establishing a convergence rate of $\nu_n(x,z)$ to $\nu(x,z)$ uniformly in every bounded region of z.

Then, we will be able to apply the convergence rate to establish (11.2.1). As argued earlier, it is sufficient to show the following.

4. For a suitably defined sequence ε_n with probability 1,

$$\limsup_{n\to\infty} \iint_T \left|\int_{\varepsilon_n}^\infty \ln x(\nu_n(dx,z) - \nu(dx,z))\right| = 0, \tag{11.2.2}$$

$$\limsup_{n\to\infty} \left|\iint_T \int_0^{\varepsilon_n} \ln x \nu_n(dx,z) ds dt\right| = 0, \tag{11.2.3}$$

and, for any fixed s,

$$\limsup_{n\to\infty} \left|\int_{(s,t)\in T} \int_0^{\varepsilon_n} \ln x \nu_n(dx,z) dt\right| = 0. \tag{11.2.4}$$

11.3 Lemmas on Integral Range Reduction

Let $\nu_n(x,z)$ denote the empirical distribution of the Hermitian matrix

$$\mathbf{H} = \mathbf{H}_n = \left(\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I}\right)\left(\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I}\right)^*$$

for each fixed $z = s + it \in \mathbb{C}$.

To establish Theorem 11.4, we need to find the limiting counterpart to

$$g_n(s,t) = \frac{\partial}{\partial s}\int_0^\infty \log x \nu_n(dx,z).$$

To this end, we have the following lemma.

Lemma 11.5. *For all $uv \neq 0$, we have*

$$\begin{aligned} c(u,v) &= \frac{1}{\pi}\iint_{x^2+y^2\le 1} e^{iux+ivy}dxdy \\ &= \frac{u^2+v^2}{4iu\pi}\int\left[\int g(s,t)e^{ius+ivt}dt\right]ds, \end{aligned} \tag{11.3.1}$$

where

$$g(s,t) = \begin{cases} \frac{2s}{s^2+t^2}, & \text{if } s^2+t^2>1, \\ 2s, & \text{otherwise.} \end{cases}$$

Proof. As in the proof of Lemma 11.2, we have seen that, for all $uv \neq 0$,

$$\int\left[\int \frac{s}{s^2+t^2}e^{i(us+vt)}dt\right]ds = \frac{2i\pi u}{u^2+v^2}.$$

Therefore,

$$\begin{aligned} c(u,v) &= \frac{1}{\pi}\iint_{x^2+y^2\le 1} e^{iux+ivy}dxdy \\ &= \frac{u^2+v^2}{2\pi^2 iu}\iint_{x^2+y^2\le 1}\int\left[\int \frac{s}{s^2+t^2}e^{i(us+vt)}dt\right]dse^{iux+ivy}dxdy \\ &= \frac{u^2+v^2}{4iu\pi}\iint\left[\frac{1}{\pi}\iint_{x^2+y^2\le 1}\frac{2(s-x)}{(s-x)^2+(t-y)^2}dxdy\right]e^{ius+ivt}dtds \\ &= \frac{u^2+v^2}{4iu\pi}\int\left[\int g(s,t)e^{ius+ivt}dt\right]ds. \end{aligned} \tag{11.3.2}$$

Note that the changes in the limits of integration are justified because

$$\int \frac{s}{s^2+t^2}e^{ivt}dt = \operatorname{sgn}(s)e^{-|svv|}$$

is an integrable function of s on $\mathbb{R}$ and, for each s, $s/(s^2+t^2)$ is an integrable function of t.

To complete the proof of the lemma, one need only note that

$$\begin{aligned}
g(s,t) &= \frac{1}{\pi}\iint_{x^2+y^2\le 1}\frac{2(s-x)}{(s-x)^2+(t-y)^2}dxdy \\
&= \Re\left(\frac{2}{\pi}\iint_{x^2+y^2\le 1}\frac{1}{(s-x)+i(t-y)}dxdy\right) \\
&= \Re\left(\frac{2}{\pi}\int_0^1\left[\int_0^{2\pi}\frac{\rho d\theta}{z-\rho e^{i\theta}}\right]d\rho\right) \quad \text{(by polar transformation)} \\
&= \Re\left(\frac{2}{i\pi}\int_0^1\left[\int_{|\zeta|=\rho}\frac{d\zeta}{\zeta(z-\zeta)}\right]\rho d\rho\right) \quad \text{(by setting } \zeta=\rho e^{i\theta}\text{)} \\
&= \Re\left(\frac{2}{i\pi}\int_0^{1\wedge|z|}\left[\frac{-2\pi i}{z}\right]\rho d\rho\right) \quad \text{(by residue theorem)} \\
&= 2\Re((|z|\wedge 1)^2 z^{-1}).
\end{aligned}$$

This completes the proof of the lemma.

Lemma 11.6. *Let λ_j and η_j denote the eigenvalues and singular values of an $n\times n$ matrix $\mathbf{A}$, respectively. Then, for any $k\le n$,*

$$\sum_{j=1}^k|\lambda_j|^2\le\sum_{j=1}^k\eta_j^2$$

if η_j is arranged in descending order.

Proof. For any order arrangement of the eigenvalues of $\mathbf{A}$, by the Schur decomposition we can find a unitary matrix $\mathbf{U}$ such that

$$\mathbf{A}=\mathbf{U}\mathbf{D}\mathbf{U}^*,$$

where $\mathbf{D}=(d_{ij})$ is an upper-triangular matrix with $d_{ij}=0$ for $i>j$ and $d_{ii}=\lambda_i$. By Lemma A.11, we have

$$\begin{aligned}
\sum_{j=1}^k\eta_j^2 &= \sup_{\mathbf{E}^*\mathbf{E}=\mathbf{F}^*\mathbf{F}=\mathbf{I}_k}|\mathrm{tr}(\mathbf{E}^*\mathbf{A}\mathbf{A}^*\mathbf{F})| \\
&\ge \mathrm{tr}\begin{pmatrix}\mathbf{I}_k & 0\end{pmatrix}\mathbf{D}\mathbf{D}^*\begin{pmatrix}\mathbf{I}_k\\ 0\end{pmatrix} \\
&= \sum_{j\le i\le k}|d_{ij}|^2 \\
&\ge \sum_{j=1}^k|\lambda_j|^2.
\end{aligned}$$

The proof is complete.

Lemma 11.7. *For any $uv \neq 0$ and $A > 2$, we have*

$$\left| \int_{|s| \geq A} \int_{-\infty}^{\infty} g_n(s,t) e^{ius+ivt} dtds \right| \leq \frac{4\pi}{|v|} e^{-\frac{1}{2}|v|A} + \frac{2\pi}{n|v|} \sum_{k=1}^{n} I\left(|\lambda_k| \geq \frac{1}{2} A \right) \tag{11.3.3}$$

and

$$\left| \int_{|s| \leq A} \int_{|t| \geq A^3} g_n(s,t) e^{ius+ivt} dsdt \right| \leq \frac{8A}{A^2 - 1} + \frac{4\pi A}{n} \sum_{k=1}^{n} I\left(|\lambda_k| \geq A \right). \tag{11.3.4}$$

Furthermore, if the function $g_n(s,t)$ is replaced by $g(s,t)$, the two inequalities above hold without the second terms.

Proof. We have

$$\begin{aligned}
&\left| \int_{|s| \geq A} \int_{-\infty}^{\infty} g_n(s,t) e^{ius+ivt} dsdt \right| \\
&= \left| \int_{|s| \geq A} \int_{-\infty}^{\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{2(s - \Re(\lambda_k))}{(s - \Re(\lambda_k))^2 + (t - \Im(\lambda_k))^2} e^{ius+ivt} dsdt \right| \\
&= \left| \frac{\pi}{n} \sum_{k=1}^{n} \int_{|s| \geq A} \operatorname{sign}(s - \Re(\lambda_k)) e^{ius - |v(s - \Re(\lambda_k))|} ds \right| \\
&\leq \left| \frac{\pi}{n} \sum_{k=1}^{n} \int_{|s| \geq A} e^{-\frac{1}{2}|vs|} ds + \int e^{-|v(s - \Re(\lambda_k))|} ds I\left(|\lambda_k| \geq \frac{1}{2} A \right) \right| \\
&\leq \frac{4\pi}{|v|} e^{-\frac{1}{2}|v|A} + \frac{2\pi}{n|v|} \sum_{k=1}^{n} I\left(|\lambda_k| \geq \frac{1}{2} A \right),
\end{aligned} \tag{11.3.5}$$

and

$$\begin{aligned}
&\left| \int_{|s| \leq A} \int_{|t| \geq A^3} g_n(s,t) e^{ius+ivt} dsdt \right| \\
&\leq \frac{1}{n} \sum_{k=1}^{n} \int_{|s| \leq A} \int_{|t| \geq A^3} \frac{2|s - \Re(\lambda_k)|}{(s - \Re(\lambda_k))^2 + (t - \Im(\lambda_k))^2} dtds \\
&\leq \frac{1}{n} \sum_{k=1}^{n} \left(\int_{|t| \geq A^3} \frac{8A^2}{(|t| - A)^2} dt + 4A\pi I(|\lambda_k| \geq A) \right) \\
&\leq \frac{8A}{(A^2 - 1)} + \frac{4\pi A}{n} \sum_{k=1}^{n} I(|\lambda_k| \geq A).
\end{aligned} \tag{11.3.6}$$

Similarly, one can prove the two inequalities above for $g(s,t)$. The proof of Lemma 11.7 is complete.

Under the conditions of Theorem 11.4, we have, by Lemma 11.6 and Kolmogorov's law of large numbers,

$$\frac{1}{n}\sum_{k=1}^{n} I(|\lambda_k| \geq A) \leq \frac{1}{n^2A^2}\mathrm{tr}(\mathbf{X}_n\mathbf{X}_n^*) \to \frac{1}{A^2}, \text{ a.s.}$$

From Lemma 11.7, one sees that the right-hand sides of (11.3.3) and (11.3.4) can be made arbitrarily small by making A large enough. The same is true when $g_n(s,t)$ is replaced by $g(s,t)$. Therefore, the proof of the circular law is reduced to showing

$$\int_{|s|\leq A}\int_{|t|\leq A^3}[g_n(s,t) - g(s,t)]e^{ius+ivt}dsdt \to 0. \tag{11.3.7}$$

Next, we define sets

$$T = \{(s,t) : \ |s| \leq A, |t| \leq A^3 \ \text{ and } \ ||z|-1| \geq \varepsilon\}$$

and

$$T_1 = \{(s,t) : \ ||z|-1| < \varepsilon\},$$

where $z = s + it$.

Lemma 11.8. *For all fixed A and $0 < \varepsilon < 1$,*

$$\iint_{T_1} |g_n(s,t)|dsdt \leq 32\sqrt{\varepsilon}. \tag{11.3.8}$$

Furthermore, when $g_n(s,t)$ in (11.3.8) is replaced by $g(s,t)$, estimation (11.3.8) still holds.

Proof. For any u and v, by a polar transformation, we obtain

$$\iint_{T_1} \frac{2|s-u|dtds}{(s-u)^2+(t-v)^2} = \left|\int_0^{2\pi} 2D(\theta)|\cos\theta|d\theta\right|,$$

where $D(\theta)$ is the total length of the intersection of the ring T_1 and the straight line $(s-u)\cos\theta + (t-v)\sin\theta = 0$, which consists of at most two pieces of segments. In fact, one can prove that the maximum value of $D(\theta)$ can be reached when the straight line is a tangent of the inner circle of the ring T_1 and hence $\max_{\theta,u.v} D(\theta) = 2\sqrt{(1+\varepsilon)^2 - (1-\varepsilon)^2} = 4\sqrt{\varepsilon}$.

The proof of (11.3.8) for $g_n(s,t)$ is then complete by noting that $\int_0^{2\pi}|\cos\theta|d\theta = 4$. The proof of (11.3.8) for $g(s,t)$ is straightforward and is thus omitted. The proof of Lemma 11.8 is complete.

Note that the right-hand side of (11.3.8) can be made arbitrarily small by choosing ε small. Thus, by Lemmas 11.7 and 11.8, to prove the circular law, one need only show that, for each fixed $A > 0$ and $\varepsilon \in (0, 1)$,

$$\iint_T [g_n(s,t) - g(s,t)]e^{ius+ivt}dsdt \to 0, \text{a.s.} \tag{11.3.9}$$

Before proving the main theorem, we first characterize the circular law.

11.4 Characterization of the Circular Law

In this section, we consider the cubic equation

$$\Delta^3 + 2\Delta^2 + \frac{\alpha + 1 - |z|^2}{\alpha}\Delta + \frac{1}{\alpha} = 0, \tag{11.4.1}$$

where $\alpha = x + iy$. The solution of the equation has three analytic branches when $\alpha \neq 0$ and when there is no multiple root. Lemma 11.10 below proves multiple roots only occur on the real line. In Lemma 11.14, it is shown that, with probability 1, the Stieltjes transform of $\nu_n(\cdot, z)$ converges to a root of (11.4.1) for an infinite number of $\alpha \in \mathbb{C}^+$ possessing a finite limit point. We claim there is only one of the three analytic branches, which we will henceforth denote by $\Delta(\alpha)$, to which the Stieltjes transforms are converging to. Let $m_2(\alpha)$ and $m_3(\alpha)$ denote the other two branches. If, say, $m_2(\alpha)$ is another Stieltjes transform, for α real converging to $-\infty$, from the fact that $\Delta(\alpha) + m_2(\alpha) + m_3(\alpha) = -2$, we would have $m_3(\alpha) \to -2$. But from $\alpha\Delta(\alpha)m_2(\alpha)m_3(\alpha) = -1$ and the fact that $\alpha\Delta(\alpha)$ converges as $\alpha \to -\infty$, we would have $m_3(\alpha)$ unbounded, a contradiction. The reader is reminded that Δ, m_2, and m_3 are also functions of $|z|$.

By Theorem B.9, there exists a distribution function $\nu(\cdot, z)$ such that

$$\Delta(\alpha) = \int \frac{1}{u - \alpha}\nu(du, z).$$

Furthermore, by Theorem B.10, $\nu(\cdot, z)$ has a density that is denoted by $p(\cdot, z)$.

We now discuss the properties of ν.

Lemma 11.9. *The limiting distribution function $\nu(x, z)$ satisfies*

$$|\nu(w + u, z) - \nu(w, z)| \leq 2\pi^{-1} \max\{2\sqrt{3|u|}, |u|\} \text{ for all } z. \tag{11.4.2}$$

Also, the limiting distribution function $\nu(u, z)$ has support in the interval $[x_1, x_2]$ when $|z| > 1$ and $[0, x_2]$ when $|z| \leq 1$, where

$$x_1 = \frac{1}{8|z|^2}[-1 + 20|z|^2 + 8|z|^4 - (\sqrt{1 + 8|z|^2})^3],$$

$$x_2 = \frac{1}{8|z|^2}[(\sqrt{1+8|z|^2})^3 - 1 + 20|z|^2 + 8|z|^4].$$

Proof. Since $p(u,z)$ is the density of the limiting spectral distribution $\nu(\cdot,z)$, $p(u,z) = 0$ for all $u < 0$. By (11.4.1), $p(u,z)$ is continuous for $u > 0$. Let $u > 0$ be a point in the support of $\nu(\cdot,z)$. Write $\Delta(u) = g(u) + ih(u)$. Then, to prove (11.4.2), it suffices to show

$$h(u) \le \max\{\sqrt{3/u}, 1\}.$$

Change (11.4.1) (for $\alpha = x > 0$) as

$$\Delta^2 + 2\Delta + 1 + \frac{1-|z|^2}{x} + \frac{1}{x\Delta} = 0.$$

Comparing the imaginary and real parts of both sides of the equation above, we obtain

$$2(g(x)+1) = \frac{1}{x(g^2(x)+h^2(x))}$$

and

$$\begin{aligned} h^2(x) &= \frac{1-|z|^2}{x} + (g(x)+1)^2 + \frac{g(x)}{x(g^2(x)+h^2(x))} \\ &= \frac{1}{x} + \frac{g(x)+1}{2x(g^2(x)+h^2(x))} + \frac{g(x)}{x(g^2(x)+h^2(x))} \\ &= \frac{1}{x} + \frac{3(g(x)+1)}{2x(g^2(x)+h^2(x))} \\ &\le \begin{cases} 1, & \text{if } h(x) \le 1, \\ \frac{3}{x}, & \text{otherwise.} \end{cases} \end{aligned} \tag{11.4.3}$$

Here, the last inequality follows from the fact that $(g+1)/(g^2+1)$ reaches maximum at $g = \sqrt{2} - 1$. Thus, (11.4.2) is established.

Now, we proceed to find the boundaries of the support of $\nu(.,z)$. Since $\nu(.,z)$ has no mass on the negative half line, we need only consider $x > 0$. Suppose $h(x) > 0$. Comparing the real and imaginary parts for both sides of (11.4.1) and then making x approach the boundary (namely, $h(x) \to 0$), we obtain

$$x(g^3 + 2g^2 + g) + (1-|z|^2)g + 1 = 0$$

and

$$x(3g^2 + 4g + 1) + 1 - |z|^2 = 0. \tag{11.4.4}$$

Thus, substituting $x(g+1) = -(1-|z|^2)/(3g+1)$ from the second identity into the first, we obtain

$$[(1-|z|^2)g + 1](3g+1) = (1-|z|^2)g(g+1).$$

For $|z| \neq 1$, the solution to this quadratic equation in g is

$$g = \frac{-3 \pm \sqrt{1+8|z|^2}}{4-4|z|^2} \quad \left(g = -\frac{1}{3}, \quad \text{if} \quad |z| = 1\right), \tag{11.4.5}$$

which, together with (11.4.4), implies that, for $|z| \neq 1$,

$$\begin{aligned} x_{1,2} &= -\frac{1-|z|^2}{(g+1)(3g+1)} \\ &= -\frac{1}{8|z|^2}\left\{1 - 20|z|^2 - 8|z|^4 \pm (\sqrt{1+8|z|^2})^3\right\}. \end{aligned} \tag{11.4.6}$$

Note that $0 < x_1 < x_2$ when $|z| > 1$. Hence, the interval $[x_1, x_2]$ is the support of $\nu(., z)$ since $p(x, z) = 0$ when x is very large. When $|z| < 1$, $x_1 < 0 < x_2$. Note that for the case $|z| < 1$, $g(x_1) < 0$, which contradicts the fact that $\Delta(x) > 0$ for all $x < 0$ and hence x_1 is not a solution of the boundary. Thus, the support of $\nu(., z)$ is the interval $(0, x_2)$. For $|z| = 1$, there is only one solution, $x_2 = -1/[g(g+1)^2] = 27/4$, which can also be expressed by (11.4.6). In this case, the support of $\nu(\cdot, z)$ is $(0, x_2)$. The proof of Lemma 11.9 is complete.

Next, we consider the separation between Δ and the other two solutions of equation (11.4.1).

Lemma 11.10. *For any given constants $N > 0$, $A > 0$, and $\varepsilon \in (0,1)$ (recall that A and ε are used to define the region T), there exist positive constants ε_1 and ε_0 (ε_0 may depend on ε_1) such that for all large n:*

(i) *for $|\alpha| \leq N$, $y \geq 0$, and $z \in T$,*

$$\max_{j=2,3} |\Delta(\alpha) - m_j(\alpha)| \geq \varepsilon_0, \tag{11.4.7}$$

(ii) *for $|\alpha| \leq N$, $y \geq 0$, $|\alpha - x_2| \geq \varepsilon_1$ (and $|\alpha - x_1| \geq \varepsilon_1$ if $|z| \geq 1 + \varepsilon$), and $z \in T$,*

$$\min_{j=2,3} |\Delta(\alpha) - m_j(\alpha)| \geq \varepsilon_0, \tag{11.4.8}$$

(iii) *for $z \in T$ and $|\alpha - x_2| < \varepsilon_1$,*

$$\min_{j=2,3} |\Delta(\alpha) - m_j(\alpha)| \geq \varepsilon_0 \sqrt{|\alpha - x_2|}, \tag{11.4.9}$$

(iv) *for $|z| > 1 + \varepsilon$, $z \in T$, and $|\alpha - x_1| < \varepsilon_1$,*

$$\min_{j=2,3} |\Delta(\alpha) - m_j(\alpha)| \geq \varepsilon_0 \sqrt{|\alpha - x_1|}. \tag{11.4.10}$$

Proof. We first prove (11.4.8). Because ε_1 is actually selected in the proofs of conclusions (iii) and (iv), here ε_1 is assumed to have already been chosen. Suppose that for some $z \in T$ and some α with $y \geq 0$, $\Delta(\alpha)$ is a multiple root of equation (11.4.1). Then, $\Delta(\alpha)$ should also be a root of the derivative of equation (11.4.1),

$$3\Delta^2 + 4\Delta + \frac{\alpha + 1 - |z|^2}{\alpha} = 0. \tag{11.4.11}$$

Since $\Delta(\alpha)$ satisfies both the above and (11.4.1), Euclidean division then yields

$$\Delta(\alpha) = \frac{1 + 8|z|^2}{2\alpha - 6 + 6|z|^2} - 1. \tag{11.4.12}$$

Since $\Im \Delta(\alpha) > 0$ for $\alpha \in \mathbb{C}^+$, we conclude that (11.4.1) can have multiple roots only when $\alpha \in \mathbb{R}$. Note that equation (11.4.11) is the same as (11.4.4). Thus, the only possible values for α are x_1 and x_2.

We claim that (11.4.8) is true. If not, then we have, for each positive integer k, that there exist α_k and z_k with $z_k \in T$ and $|\alpha - x_2| \geq \varepsilon_1$ and $|\alpha - x_1| \geq \varepsilon_1$ if $|z_k| \geq 1 + \varepsilon$ such that

$$\min_{j=2,3} |\Delta(\alpha_k) - m_j(\alpha_k)| < \frac{1}{k}.$$

Then, we may select a subsequence $\{k'\}$ such that the following are true: $\alpha_{k'} \to \alpha_0$ and $z_{k'} \to z_0$; $z_0 \in T$ and $|\alpha_0 - x_2| \geq \varepsilon_1$. If $|z_0| \geq 1 + \varepsilon$, then $|\alpha_0 - x_1| \geq \varepsilon_1$. For at least one of $j = 2$ or 3, say $j = 2$,

$$|\Delta(\alpha_{k'}) - m_2(\alpha_{k'})| < \frac{1}{k'}. \tag{11.4.13}$$

If $\alpha_0 \neq 0$, by continuity of $\Delta(\alpha)$ and $m_2(\alpha)$, we shall have $\Delta(\alpha_0) = m_2(\alpha_0)$, which contradicts the fact that $\Delta(\alpha)$ does not coincide with $m_2(\alpha)$ except for $\alpha = x_2$ or $\alpha = x_1$ when $|z| > 1$. If $\alpha_0 = 0$, then for $z_0 \neq 1$ it is straightforward to argue that one root must be bounded and hence converges to $1/(|z_0|^2 - 1)$, while the other two become unbounded as $\alpha_{k'} \to 0$. Thus, limit (11.4.13) requires that both $\Delta(\alpha_{k'})$ and $m_2(\alpha_{k'})$ are unbounded. On the other hand, since $\Delta(\alpha_{k'}) + m_2(\alpha'_k) + m_3(\alpha'_k) = -2$, we should have

$$|\Delta(\alpha_{k'}) - m_2(\alpha_{k'})| = |-2 - 2m_2(\alpha_{k'}) - m_3(\alpha_{k'})| \to \infty.$$

This contradiction shows that $\alpha_0 = 0$ is impossible. This concludes the proof of (11.4.8).

For (11.4.7) we argue as above and assume there are $\alpha_{k'} \to \alpha_0$, $z_{k'} \to z_0 \in T$ such that all three roots of (11.4.1) converge to each other. From the above we see immediately that α_0 cannot equal zero. For $\alpha_0 \neq 0$ we would have from the second derivative of (11.4.1) $\Delta(\alpha_0) = -2/3$. Again from the above, α has to be x_1 or x_2 and, from (11.4.6),

$$\Delta(\alpha_0) = -\frac{2}{3 \pm \sqrt{1 + 8|z_0|^2}},$$

which clearly cannot equal $-2/3$ since $|z| \neq 1$.

We now prove (11.4.9). Let $\Delta + \rho$ be either m_2 or m_3. Since both Δ and $\Delta + \rho$ satisfy (11.4.1), we obtain

$$\rho(\alpha) = -\frac{3\Delta^2(\alpha) + 4\Delta(\alpha) + 1 + (1 - |z|^2)/\alpha}{3\Delta(\alpha) + 2 + \rho(\alpha)}. \tag{11.4.14}$$

Write $\hat{\rho} = \Delta(\alpha) - \Delta(x_2)$. By (11.4.4), we have

$$\begin{aligned} &3\Delta^2(\alpha) + 4\Delta(\alpha) + 1 + (1 - |z|^2)/\alpha \\ &= 3\Delta^2(\alpha) + 4\Delta(\alpha) + 1 + (1 - |z|^2)/\alpha \\ &\quad -3\Delta^2(x_2) + 4\Delta(x_2) + 1 + (1 - |z|^2)/x_2 \\ &= \hat{\rho}[6\Delta(x_2) + 4 + 3\hat{\rho}] + \frac{(1 - |z|^2)(x_2 - \alpha)}{x_2\alpha}. \end{aligned} \tag{11.4.15}$$

From (11.4.1) and (11.4.4), it follows that

$$\begin{aligned} 0 &= [3\Delta^2(x_2) + 4\Delta(x_2) + 1 + (1 - |z|^2)/\alpha]\hat{\rho} + [3\Delta(x_2) + 2]\hat{\rho}^2 + \hat{\rho}^3 \\ &\quad + \frac{(x_2 - \alpha)(\Delta(x_2)(1 - |z|^2) + 1)}{x_2\alpha} \\ &= [3\Delta(x_2) + 2 + \hat{\rho}]\hat{\rho}^2 + \frac{((\Delta(x_2) + \hat{\rho})(1 - |z|^2) + 1)(x_2 - \alpha)}{x_2\alpha}. \end{aligned} \tag{11.4.16}$$

Let $M > 1$ be a bound on all $|z|$, $|\alpha|$, and x_2. Then $\Delta(x_2) = -\frac{2}{3+\sqrt{1+8|z|^2}} \in \left(-\frac{1}{2}, -\frac{2}{3+\sqrt{1+8M^2}}\right)$. When $|\hat{\rho}| < \frac{2}{3+\sqrt{1+8M^2}}$, we have

$$|3\Delta(x_2) + 2 + \hat{\rho}| = \left| \frac{\sqrt{1 + 8|z|^2}}{3 + \sqrt{1 + 8|z|^2}} + \hat{\rho} \right| < 2.$$

Also, since $\Delta(x_2) + \Re\hat{\rho} < 0$, we have $|(\Delta(x_2) + \hat{\rho})(1 - |z|^2) + 1| > 1$ when $|z| > 1$ and $> \frac{-4}{3+\sqrt{1+8M^2}} + 1 > 1/3$ when $|z| < 1$. Therefore, regardless of the size of $\hat{\rho}$, equation (11.4.16) implies

$$|\hat{\rho}| \geq \min\left(\frac{2}{3 + \sqrt{1 + 8M^2}}, \frac{1}{2M^2}\sqrt{|x_2 - \alpha|}\right) \geq c_1\sqrt{|x_2 - \alpha|} \tag{11.4.17}$$

for some positive constant c_1. Note that Δ is continuous in an open set containing the rectangle $\{(\alpha, z);\ z \in T,\ x_{2,\min} \leq x \leq x_{2,\max},\ 0 \leq y \leq N\}$, where $x_{2,\min} = 4$ (corresponding to $z = 0$) and $x_{2,\max} = \frac{1}{8M^2}[(1 + 8M^2)^{3/2} - 1 + 20M^2 + 8M^4]$ (corresponding to $|z| = M$). Therefore, we may select a positive constant $\varepsilon_1 \leq \min(1, c_1^2/M^4)$ such that, for all $|z| \leq M$ and $|\alpha - x_2| \leq \varepsilon_1$, we have $|\hat{\rho}| \leq \min(\frac{1}{8}, c_1^2/M^4)$. Then, from (11.4.14) and (11.4.15) and the fact that when $|\rho(\alpha)| \le \frac{1}{8}$

$$|3\Delta(\alpha) + 2 + \rho(\alpha)| = |3\Delta(x_2) + 3\hat{\rho} + 2 + \rho(\alpha)| \le 4,$$

we conclude that

$$\begin{aligned} |\rho(\alpha)| &\ge \frac{1}{4}\min\left(\frac{1}{2}, \left|\hat{\rho}[6\Delta(x_2) + 4 + 3\hat{\rho}] + \frac{(1-|z|^2)(x_2-\alpha)}{x_2\alpha}\right|\right) \\ &\ge \frac{1}{4}\min\left(\frac{1}{2}, \frac{1}{2}c_1\sqrt{|x_2-\alpha|} - \frac{1}{9}M^2|x_2-\alpha|\right) \\ &\ge c_2\sqrt{|x_2-\alpha|}. \end{aligned} \tag{11.4.18}$$

This concludes the proof of (11.4.9).

The proof of (11.4.10) is similar to that of (11.4.9). Checking the proof of (11.4.9), one finds that equations (11.4.14)–(11.4.16) are still true if x_2 is replaced by x_1. The rest of the proof depends on the fact that, for all $z \in T$, $|z| \ge 1+\varepsilon$, and $|\alpha - x_1| \le \varepsilon_1$, $|3\Delta(\alpha)+2+\rho(\alpha)|$ has a uniform upper bound and $\hat{\rho}$ can be made arbitrarily small provided ε_1 is small enough. Indeed, these can be done because x_1 has a minimum $x_{1,\min}$ at $|z| = 1+\varepsilon$ that is strictly greater than 0 and hence $\Delta(\alpha)$ is uniformly continuous in an open set containing the rectangle $\{(\alpha, z);\ z \in T,\ x_{1,\min} - \varepsilon_1 \le x \le x_{1,\max},\ 0 \le y \le N\}$, provided ε_1 is chosen so that $x_{1,\min} - \varepsilon_1 > 0$.

The proof of Lemma 11.10 is then complete.

The purpose of the following lemma is to show that $\nu(\cdot, z)$ defines the circular law.

Lemma 11.11. *We have*

$$\frac{\partial}{\partial s}\int_0^\infty \ln x \nu(dx, z) = g(s,t). \tag{11.4.19}$$

Proof. From (11.4.1) and the fact that Δ is the Stieltjes transform of $\nu(\cdot, z)$, we have, for $x < 0$,

$$\Delta(x)\begin{cases} > 0, & \text{if } x < 0, \\ \to 0, & \text{as } x \to -\infty, \\ \le \sqrt{\frac{2(1-|z|^2)}{|x|}}, & \text{as } x \uparrow 0 \text{ if } |z| < 1, \\ \le |x|^{-1/3}, & \text{as } x \uparrow 0 \text{ if } |z| = 1, \\ \uparrow \frac{1}{|z|^2-1}, & \text{as } x \uparrow 0 \text{ if } |z| > 1. \end{cases}$$

Thus, for any $C > 0$, the integral $\int_{-C}^0 \Delta(x)dx$ exists. We have, using integration by parts,

$$\int_{-C}^0 \Delta(x)dx = \int_0^C \Delta(-x)dx = \int_0^C \int_0^\infty \frac{1}{u+x}\nu(du, z)dx$$

$$= \int_0^\infty [\log(C+u) - \log u]\nu(du, z)$$
$$= \log C + \int_0^\infty \log(1+u/C)\nu(du,z) - \int_0^\infty \log u\nu(du, z).$$

Differentiating both sides with respect to s, we get

$$\frac{\partial}{\partial s}\int_0^\infty \log u\nu(du, z)$$
$$= \frac{\partial}{\partial s}\int_0^\infty \log(1+u/C)\nu(du,z) - \frac{\partial}{\partial s}\int_{-C}^0 \Delta(x)dx. \tag{11.4.20}$$

Differentiating both sides of (11.4.1) with respect to s and x, respectively, we obtain

$$\frac{\partial}{\partial s}\Delta(x)\left[3\Delta^2(x) + 4\Delta(x) + \frac{x+1-|z|^2}{x}\right] = \frac{2s\Delta(x)}{x} \tag{11.4.21}$$

and

$$\frac{\partial}{\partial x}\Delta(x)\left[3\Delta^2(x) + 4\Delta(x) + \frac{x+1-|z|^2}{x}\right] = \frac{\Delta(x)(1-|z|^2)+1}{x^2}.$$

Comparing the two equations, we get

$$\frac{\partial}{\partial s}\Delta(x) = \frac{2sx\Delta(x)}{1+\Delta(x)(1-|z|^2)}\frac{\partial}{\partial x}\Delta(x) = -\frac{2s}{(1+\Delta(x))^2}\frac{\partial}{\partial x}\Delta(x), \tag{11.4.22}$$

where the last equality follows from the fact that

$$x = \frac{1+\Delta(x)(1-|z|^2)}{\Delta(x)(1+\Delta(x))^2}.$$

We now determine the behavior of $\frac{\partial}{\partial s}\Delta(x)$ near the boundary of the support of $\nu(\cdot, z)$. For the following, x_1 will denote the left endpoint of the support of $\nu(\cdot, z)$ regardless of the value of z. Let $\hat{x}$ denote x_2 or, when $|z| > 1$, $x_1 > 0$. Using (11.4.21), it is a simple matter to show

$$\lim_{x\to\hat{x}}\frac{\partial}{\partial x}\left[3\Delta^2(x) + 4\Delta(x) + \frac{x+1-|z|^2}{x}\right]^2$$
$$= \lim_{x\to\hat{x}}\frac{\partial}{\partial x}\Delta(x)\left[3\Delta^2(x) + 4\Delta(x) + \frac{x+1-|z|^2}{x}\right]\left[12\Delta(x)+8\right]$$
$$-\lim_{x\to\hat{x}}\left[3\Delta^2(x) + 4\Delta(x) + \frac{x+1-|z|^2}{x}\right]\frac{(1-|z|^2)}{x^2}$$
$$= \frac{[\Delta(\hat{x})(1-|z|^2)+1](12\Delta(\hat{x})+8)}{\hat{x}^2},$$

which is continuous in z. This implies that, for x near $\hat{x}$,

$$\left|3\Delta^2(x) + 4\Delta(x) + \frac{x+1-|z|^2}{x}\right| \geq K_1\sqrt{|x-\hat{x}|},$$

so that from (11.4.21) we get that, for x near $\hat{x}$,

$$\left|\frac{\partial}{\partial s}\Delta(x)\right| \leq \frac{K_2}{\sqrt{|x-\hat{x}|}}.$$

When $|z| \leq 1$ (so that $x_1 = 0$), it follows from (11.4.1) that, as $x \to 0$, $x\Delta^2(x) \to -1+|z|^2$ when $|z|<1$ and $x\Delta^3(x) \to -1$ when $|z|=1$. Therefore, from (11.4.21) we have, for x small,

$$\left|\frac{\partial}{\partial s}\Delta(x)\right| \leq K_3|x|^{-a},$$

where $a = 1/2$ when $|z|<1$ and is $2/3$ when $|z|=1$.

The bounds on $\frac{\partial}{\partial s}\Delta(x)$ above, together with the dominated convergence theorem, justify the interchange of differentiation and integration about to be performed below.

Substituting (11.4.22) into the second term on the right-hand side of (11.4.20), it follows that

$$\begin{aligned}\frac{\partial}{\partial s}\int_{-C}^{0}\Delta(x)dx &= \int_{-C}^{0}\frac{\partial}{\partial s}\Delta(x)dx = -\int_{-C}^{0}\frac{\partial}{\partial x}\Delta(x)\frac{2s}{(1+\Delta(x))^2}dx\\ &= -2s\int_{\Delta(-C)}^{\Delta(0_-)}\frac{1}{(1+\Delta)^2}d\Delta\\ &= \frac{2s}{1+\Delta(0_-)} - \frac{2s}{1+\Delta(-C)}.\end{aligned} \tag{11.4.23}$$

Taking $x \uparrow 0$ in (11.4.1), we get

$$\Delta(0_-) = \begin{cases}\infty, & \text{if } |z|^2 \leq 1,\\ \frac{1}{|z|^2-1}, & \text{if } |z|^2 > 1.\end{cases}$$

We also have $\Delta(-C) \to 0$ as $C \to \infty$. Thus, we get

$$\int_{-C}^{0}\frac{\partial}{\partial s}\Delta(x)dx \to -g(s,t). \tag{11.4.24}$$

By noting that $\nu(dx,z)/dx = \pi^{-1}\Im(\Delta(x))$, we have

$$\left|\frac{\partial}{\partial s}\int_{0}^{\infty}\log(1+u/C)\nu(du,z)\right| = \left|\frac{\partial}{\partial s}\int_{x_1}^{x_2}\log(1+u/C)\nu(du,z)\right|$$

$$
\begin{aligned}
&= \left|\frac{1}{\pi}\Im\left(\int_{x_1}^{x_2}\log(1+u/C)\frac{\partial}{\partial s}(\Delta(u))du\right)\right| \\
&= \left|\frac{1}{\pi}\Im\left(\int_{x_1}^{x_2}\log(1+u/C)\frac{2s}{(1+\Delta(u))^2}\frac{\partial}{\partial u}\Delta(u)du\right)\right| \\
&= \left|\frac{2s}{\pi}\Im\left(\frac{\log(1+x_1/C)}{1+\Delta(x_1+0)}-\frac{\log(1+x_2/C)}{1+\Delta(x_2)}\right.\right. \\
&\qquad \left.\left.+\int_{x_1}^{x_2}\frac{1}{(1+\Delta(u))(C+u)}du\right)\right| \\
&\to 0 \quad \text{as } C\to\infty. \qquad (11.4.25)
\end{aligned}
$$

Assertion (11.4.19) then follows from (11.4.20), (11.4.24), and (11.4.25), and the proof of Lemma 11.11 is complete.

11.5 A Rough Rate on the Convergence of $\nu_n(x,z)$

In this section, we shall establish a convergence rate of $\nu_n(x,z)$ to a limiting distribution $\nu(x,z)$ and discuss properties of the limiting distribution $\nu(x,z)$. In the remainder of this chapter, if the quantities represented by the symbols $o(1)$ or $O(1)$ are involved with indices j, ℓ, or k, or variables α or z, then their order is uniform with respect to these indices and variables.

11.5.1 Truncation and Centralization

Let $\widehat{\mathbf{X}}_n$ and $\widetilde{\mathbf{X}}_n$ be the $n\times n$ matrices with entries

$$\widehat{x}_{ij} = x_{ij}I(|x_{ij}|<n^{\delta}) - \mathrm{E}x_{ij}I(|x_{ij}|<n^{\delta})$$

and $\tilde{x}_{ij} = \widehat{x}_{ij}/\sqrt{\mathrm{E}|\widehat{x}_{ij}|^2}$, respectively, where $\delta\in(0,\frac{1}{4})$. Further, denote by $\hat{\nu}_n(\cdot,z)$ and $\tilde{\nu}_n(\cdot,z)$ the empirical spectral distributions of $\widehat{\mathbf{H}} = (\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_n - z\mathbf{I})(\frac{1}{\sqrt{n}}\widehat{\mathbf{X}}_n - z\mathbf{I})^*$ and $\widetilde{\mathbf{H}} = (\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n - z\mathbf{I})(\frac{1}{\sqrt{n}}\widetilde{\mathbf{X}}_n - z\mathbf{I})^*$. Then, by Corollary A.42, we have

$$
\begin{aligned}
&L^4(\nu_n(\cdot,z),\hat{\nu}_n(\cdot,z)) \\
&\le \frac{2}{n^3}\left[\mathrm{tr}(\mathbf{H}+\widehat{\mathbf{H}})\mathrm{tr}((\mathbf{X}_n-\widehat{\mathbf{X}}_n)(\mathbf{X}_n-\widehat{\mathbf{X}}_n)^*)\right] \qquad (11.5.1)
\end{aligned}
$$

and

$$L^4(\hat{\nu}_n(\cdot,z),\tilde{\nu}_n(\cdot,z))$$

$$\leq \frac{2}{n^3}\left[\mathrm{tr}(\widehat{\mathbf{H}}+\widetilde{\mathbf{H}})\mathrm{tr}(\widehat{\mathbf{X}}_n\widehat{\mathbf{X}}_n^*)\Big(1-1/\sqrt{\mathrm{E}|\widehat{x}_{11}^2|}\Big)^2\right]. \qquad (11.5.2)$$

By Kolmogorov's law of large numbers, it is easy to see that

$$\frac{1}{n}\mathrm{tr}H = \frac{1}{n^2}\sum_{ij}|x_{ij}|^2 - 2\Re\left(\frac{\bar{z}}{n^{3/2}}\sum_{k=1}^{n}x_{kk}\right) + |z|^2 \to (1+|z|), \text{ a.s.}$$

Similarly, it can be proved that $\frac{1}{n}\mathrm{tr}\widehat{\mathbf{H}} \to 1+|z|^2$, a.s.

Furthermore, by the assumption that $\mathrm{E}|x_{11}|^{2+\eta} < \infty$, for any $L > 0$,

$$\begin{aligned}
&\limsup_{n\to\infty} n^{\delta\eta}\frac{1}{n^2}\mathrm{tr}(\mathbf{X}_n-\widehat{\mathbf{X}}_n)(\mathbf{X}_n-\widehat{\mathbf{X}}_n)^* \\
&= \limsup_{n\to\infty} n^{\delta\eta}\frac{1}{n^2}\sum_{ij}|x_{ij}I(|x_{ij}|>n^\delta) - \mathrm{E}x_{11}I(|x_{11}|>n^\delta)|^2 \\
&\leq 2\limsup_{n\to\infty} n^{\delta\eta}\left(\mathrm{E}|x_{11}|^2 I(|x_{11}|>n^\delta) + \frac{1}{n^2}\sum_{ij}|x_{ij}|^2 I(|x_{ij}|>n^\delta)\right) \\
&\leq 2\lim_{n\to\infty}\left(\mathrm{E}|x_{11}|^{2+\eta}I(|x_{11}|>L) + \frac{1}{n^2}\sum_{ij}|x_{ij}|^{2+\eta}I(|x_{ij}|>L)\right) \\
&= 4\mathrm{E}|x_{11}|^{2+\eta}I(|x_{11}|>L), \text{ a.s.,}
\end{aligned}$$

which can be made arbitrarily small be making L suitably large. This, together with (11.5.1), implies that

$$L^2(\nu_n(\cdot,z),\hat{\nu}_n(\cdot,z)) = o_{\mathrm{a.s.}}(n^{-\delta\eta}).$$

Here the convergence rate is uniform for any $|z| \leq M$.

Again, by the assumption that $\mathrm{E}|x_{11}|^{2+\eta} < \infty$, it is easy to show that

$$1-\sqrt{\mathrm{E}|\widehat{x}_{11}|^2} \leq 1-\mathrm{E}|\widehat{x}_{11}|^2 = o(n^{-\sigma\eta}).$$

Thus, by (11.5.2),

$$L^4(\hat{\nu}_n(\cdot,z),\tilde{\nu}_n(\cdot,z)) = o_{\mathrm{a.s.}}(n^{-2\delta\eta}).$$

Combining the above, we have proved that

$$L(\nu_n(\cdot,z),\tilde{\nu}_n(\cdot,z)) = o_{\mathrm{a.s.}}(n^{-\delta\eta/4}), \qquad (11.5.3)$$

where the convergence rate $o_{\mathrm{a.s.}}(n^{-\delta\eta/4})$ is uniform for $|z| \leq M$.

11.5.2 A Convergence Rate of the Stieltjes Transform of $\nu_n(\cdot, z)$

In this subsection, we assume the conditions of Theorem 11.4 hold. Also, according to what was proved in the last subsection, we assume that the additional condition $|x_{ij}| \le n^\delta$ holds.

Lemma 11.12. *If δ is chosen to be $< \frac{1}{4}$, then, for any fixed k, we have*

$$\mathrm{E}\mathrm{tr}\mathbf{H}^k = O(n),$$

where the order $O(n)$ is uniform for $|z| \le M$.

Proof. Write $\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I}_n = \mathbf{W} = (w_{ij})_{i,j=1}^n$. Then, we have

$$\begin{aligned} &|w_{ii}| \le |z| + n^{\delta - 1/2} \le M + 1, \\ &\mathrm{E}w_{ij} = 0, \quad \text{if } i \ne j, \\ &\mathrm{E}|w_{ij}|^\mu \le n^{-\mu/2 + \delta(\mu - 2)}, \quad \text{if} \quad i \ne j \quad \text{and} \quad \mu \ge 2. \end{aligned}$$

We have

$$\mathrm{E}\mathrm{tr}(\mathbf{H}^k) = \sum\nolimits^* \mathrm{E} w_{i_1,j_1} \bar{w}_{i_2,j_1} w_{i_2,j_2} \bar{w}_{i_3,j_2} \cdots w_{i_k,j_{2k}} \bar{w}_{i_1,j_{2k}},$$

where the summation $\sum^*$ is taken for $i_1, j_1, \cdots, i_k, j_k$ run over $1, \cdots, n$. Similar to the proof of Theorem 3.6, we may construct a $G(\mathbf{i}, \mathbf{j})$-graph. We further distinguish a vertical edge as a perpendicular or skew edge if its I-vertex and J-vertex have equal value or not, respectively.

If there is a single skew edge in $G(\mathbf{i}, \mathbf{j})$, the value of the corresponding term is zero. For other graphs $G(\mathbf{i}, \mathbf{j})$, let r denote the number of distinguished values of its vertices and s denote the number of noncoincident skew edges with multiplicities $\nu_1, \cdots, \nu_s$. Note that a new vertex value must be led by a skew edge which implies that $r \le s + 1$. For such a graph, the contribution of the term is dominated by

$$(M+1)^{2k - \nu_1 - \cdots - \nu_2} n^{-s} n^{-(\frac{1}{2} - \delta)(\nu_1 + \cdots + \nu_2 - 2s)}.$$

For an isomorphism class, there are $n(n-1)\cdots(n-r+1) \sim n^r$ isomorphic graphs and there are a finite number of isomorphism classes. The lemma then follows.

Remark 11.13. With more accurate counting, one can prove that

$$\frac{1}{n}\mathrm{tr}(\mathbf{H}^k) \to \mu_k(|z|^2) \text{ a.s.} \quad \text{as } n \to \infty.$$

From this, one may conclude that $\nu_n(\cdot, z)$ tends to a limiting spectral distribution. However, we need a convergence rate, which has to be obtained by an alternate way. Thus, we shall omit the details here. Interested readers may try to derive it themselves as an exercise.

Denote by

$$\Delta_n(\alpha, z) = \int \frac{\nu_n(dx, z)}{x - \alpha},$$

where $\alpha = x + iy$ with $y > 0$, the Stieltjes transforms of $\nu_n(\cdot, z)$. For brevity, the variable z will be suppressed when there is no confusion.

In this subsection, we shall prove that ν_n tends to the limiting spectral distribution ν with a certain rate by the following lemmas.

Lemma 11.14. *Suppose the conditions of Theorem 11.4 hold, and each* $|x_{ij}| \leq n^{\delta}$. *Write*

$$\Delta_n(\alpha)^3 + 2\Delta_n(\alpha)^2 + \frac{\alpha + 1 - |z|^2}{\alpha}\Delta_n(\alpha) + \frac{1}{\alpha} = r_n. \tag{11.5.4}$$

If δ *is chosen such that* $\delta\eta < 1/14$ *and* $\delta < 1/14$, *then the remainder term* r_n *satisfies*

$$\begin{aligned} &\sup\{|r_n| : \alpha = x + iy \quad \text{with} - \infty < x < \infty, y \geq y_n, |z| \leq M\} \\ &= o_{\text{a.s.}}(\delta_n), \end{aligned} \tag{11.5.5}$$

$y_n = n^{-\delta\eta}$ *and* $\delta_n = n^{-\delta\eta}$.

Remark 11.15. As shown earlier, only one branch of (11.4.1) can be a Stieltjes transform. Using Theorem B.9, this, together with the a.s. convergence of $(1/n)\text{tr}H$ (which shows $\{\nu_n\}$ to be almost surely tight), proves that, with probability one, ν_n converges in distribution to ν.

Proof. Consider the case where $|\alpha| > n^{2\delta\eta}$, $y \geq y_n$, and $|z| \leq M$, where M is a given constant. Then, by Lemma 11.12,

$$\begin{aligned} \sup_{|\alpha|>n^{2\delta\eta}, y \geq y_n,\ |z|\leq M} |\Delta_n(\alpha)| &\leq 2n^{-2\delta\eta} + y^{-1} I\left(\Lambda_{\max}(\mathbf{H}_n) > \frac{1}{2}n^{2\delta\eta}\right) \\ &\leq 2n^{-2\delta\eta} + 2^k y^{-1} n^{-2k\delta\eta}\text{tr}(H_n^k)) \\ &= o_{\text{a.s.}}(n^{-\delta\eta}), \end{aligned}$$

provided k is chosen such that $(k-1)\delta\eta > 1$.

Consequently,

$$\begin{aligned} &\sup_{|\alpha|>n^{2\delta\eta}, y\geq y_n,\ |z|\leq M} |r_n| \\ &= \sup_{|\alpha|>n^{\delta\eta},\ |z|\leq M} \left| \Delta_n^3 + 2\Delta_n^2 + \frac{\alpha + 1 - |z|^2}{\alpha}\Delta_n + \frac{1}{\alpha} \right| \\ &\leq o_{\text{a.s.}}(\delta_n). \end{aligned} \tag{11.5.6}$$

Let $\alpha' = x' + iy'$. If $|\alpha - \alpha'| \le n^{-6\delta\eta}$ and $y,\ y' \ge y_n$, then, for $k = 1, 2, 3$,

$$|\Delta_n^k(\alpha) - \Delta_n^k(\alpha')| \le k y_n^{-k-1} |\alpha - \alpha'| = o(\delta_n^{4-k}),$$

which implies

$$\sup_{\substack{|\alpha-\alpha'|\le n^{-6\delta\eta} \\ y,\ y'\ge y_n,\quad |z|\le M}} |r_n(\alpha) - r_n(\alpha')| = o(\delta_n). \tag{11.5.7}$$

Let $\Lambda_k(z)$ (arranging in increasing order) denote the eigenvalues of the matrix $\mathbf{H}_n(z) = (\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I})(\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I})^*$. In the proof of Corollary A.42, we have seen that

$$\left(\sum_{k=1}^{n} |\Lambda_k(z) - \Lambda_k(z')|\right)^2 \le 2(\mathrm{tr}\mathbf{H}_n(z) + \mathrm{tr}\mathbf{H}_n(z'))\mathrm{tr}(|z - z'|^2 \mathbf{I}_n).$$

Thus we have, for $k = 1, 2, 3$,

$$\begin{aligned}
&\sup_{\substack{|z|,|z'|\le M, |z-z'|\le n^{-6\delta\eta} \\ y\ge y_n}} |\Delta_n^k(\alpha, z) - \Delta_n^k(\alpha, z')| \\
\le &\sup_{\substack{|z|,|z'|\le M, |z-z'|\le n^{-6\delta\eta} \\ y\ge y_n}} k y^{-k+1} |\Delta_n(\alpha, z) - \Delta_n(\alpha, z')| \\
\le &\sup_{\substack{|z|,|z'|\le M, |z-z'|\le n^{-6\delta\eta} \\ y\ge y_n}} k y^{-k+1} \frac{1}{n} \sum_{k=1}^{n} \frac{|\Lambda_k(z) - \Lambda_k(z')|}{|\Lambda_k(z) - \alpha||\Lambda_k(z') - \alpha|} \\
\le &\sup_{\substack{|z|,|z'|\le M, |z-z'|\le n^{-6\delta\eta} \\ y\ge y_n}} y^{-k-1} |z - z'| \left(\frac{2}{n}\mathrm{tr}(\mathbf{H}_n(z) + \mathbf{H}_n(z'))\right)^{1/2} \\
= &\, o_{\mathrm{a.s.}}(\delta_n^{4-k}).
\end{aligned} \tag{11.5.8}$$

Here, we have used the fact that $\frac{1}{n}\mathrm{tr}(H_n(z)) \to 1 + |z|^2$ a.s. uniformly for $|z| \le M$.

This, together with (11.5.6) and (11.5.7), shows that in order to finish the proof of (11.5.5), it is sufficient to show that

$$\max_{\ell\le\mu_n,\ j\le\mu_n'} \{|r_n(\alpha_\ell, z_j)|\} = o_{\mathrm{a.s.}}(\delta_n), \tag{11.5.9}$$

where $\alpha_\ell = x(\ell) + iy(\ell)$, $\ell = 1, 2, \cdots, \mu_n$, and z_j, $j = 1, 2, \cdots, \mu_n'$, are selected so that $|\alpha_\ell| \le n^{2\delta\eta}$, $y_n \le y(\ell)$, and for each $|\alpha| \le n^{2\delta\eta}$ with $y \ge y_n$, there is an ℓ such that $|\alpha - \alpha_\ell| < n^{-6\delta\eta}$, and for each $|z| \le M$, there is a j such that $|z - z_j| \le n^{-6\delta\eta}$.

By the conditions on the α_ℓ and z_j's, we may assume that

$$\mu_n \le 2n^{18\delta\eta} \quad \text{and } \mu_n' \le 3M^2 n^{12\delta\eta}.$$

In the rest of the proof of this lemma, we shall suppress the indices ℓ and j from the variables α_ℓ and z_j. The reader should remember that we shall only consider those α_ℓ and z_j that were selected as above.

Now the proof of (11.5.5) reduces to showing

$$\sup_{\alpha_\ell, z_j} \left| (\mathrm{E}\Delta_n(\alpha_\ell))^3 + 2(\mathrm{E}\Delta_n(\alpha_\ell))^2 + \frac{\alpha_\ell + 1 - |z|^2}{\alpha_\ell} \mathrm{E}\Delta_n(\alpha_\ell) + \frac{1}{\alpha_\ell} \right| \\ = o(\delta_n) \tag{11.5.10}$$

and

$$\sup_{\alpha_\ell, z_j} |\Delta_n(\alpha_\ell) - \mathrm{E}\Delta_n(\alpha_\ell)| = o_{\mathrm{a.s.}}(\delta_n^3). \tag{11.5.11}$$

Use the notation $\mathbf{W} = \mathbf{W}_n(z) = \frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I}_n = (w_{ij})$, where $w_{ij} = \frac{1}{\sqrt{n}} x_{ij}$ for $i \neq j$ and $w_{ii} = \frac{1}{\sqrt{n}} x_{ii} - z$. Then, $\mathbf{H} = \mathbf{W}\mathbf{W}^*$. We first prove (11.5.11). As before, we use the method of martingale decomposition. Let E_k denote the conditional expectation given $\{x_{ij}, i \leq k, j \leq n\}$. Then,

$$\Delta_n(\alpha) - \mathrm{E}\Delta_n(\alpha) = \frac{1}{n}\sum_{k=1}^n \gamma_k,$$

where $\mathrm{E}_0 = \mathrm{E}$,

$$\begin{aligned}\gamma_k &= \mathrm{E}_k \mathrm{tr}(\mathbf{H} - \alpha\mathbf{I})^{-1} - \mathrm{E}_{k-1}\mathrm{tr}(\mathbf{H} - \alpha\mathbf{I})^{-1} \\ &= (\mathrm{E}_k - \mathrm{E}_{k-1}) \frac{1 + \mathbf{w}_k'\mathbf{W}_k^*(\mathbf{H}_k - \alpha\mathbf{I}_{n-1})^{-2}\mathbf{W}_k\overline{\mathbf{w}}_k}{|\mathbf{w}_k|^2 - \alpha - \mathbf{w}_k'\mathbf{W}_k^*(\mathbf{H}_k - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_k\overline{\mathbf{w}}_k},\end{aligned}$$

$\mathbf{w}_k'$ denotes the k-th row vector of $\mathbf{W}$, $\mathbf{W}_k$ consists of the rest of the $n-1$ rows of $\mathbf{W}$ when $\mathbf{w}_k'$ is removed, and $\mathbf{H}_k = \mathbf{W}_k\mathbf{W}_k^*$. By the fact that

$$\begin{aligned}&|1 + \mathbf{w}_k'\mathbf{W}_k^*(\mathbf{H}_k - \alpha\mathbf{I}_{n-1})^{-2}\mathbf{W}_k\overline{\mathbf{w}}_k| \\ &\leq 1 + \mathbf{w}_k'\mathbf{W}_k^*[(\mathbf{H}_k - x\mathbf{I}_{n-1})^2 + y^2\mathbf{I}_{n-1}]^{-1}\mathbf{W}_k\overline{\mathbf{w}}_k \\ &= -y^{-1}\Im(|\mathbf{w}_k|^2 - \alpha - \mathbf{w}_k'\mathbf{W}_k^*(\mathbf{H}_k - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_k\overline{\mathbf{w}}_k),\end{aligned}$$

it follows that

$$|\gamma_k| \leq 2y^{-1}.$$

Then, by Lemma 2.12, we obtain

$$\delta_n^{-6m}\mathrm{E}|\Delta_n(\alpha) - \mathrm{E}\Delta_n(\alpha)|^{2m} \leq C_m n^{-2m+6m\delta\eta}\mathrm{E}\left(\sum_{k=1}^n |\gamma_k|^2\right)^m \\ \leq C_m 2^m n^{-m+6m\delta\eta} y^{-2m} \leq C_m n^{-m\delta\eta}. \tag{11.5.12}$$

Then, conclusion (11.5.11) follows by choosing m such that $(m-30)\delta\eta > 1$.

Next, we proceed to the proof of (11.5.10). We have

$$\Delta_n(\alpha) = \frac{1}{n}\mathrm{tr}\,(\mathbf{H} - \alpha\mathbf{I})^{-1} = \frac{1}{n}\sum_{k=1}^{n}\beta_k, \tag{11.5.13}$$

where

$$\beta_k = \frac{1}{|\mathbf{w}_k|^2 - \alpha - \mathbf{w}_k'\mathbf{W}_k^*(\mathbf{H}_k - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_k\overline{\mathbf{w}}_k}.$$

Since distributions of the β_k's are the same, we have from (11.5.13)

$$\mathrm{E}\Delta_n(\alpha) = b_n + \mathrm{E}\varepsilon_1^2 b_n^2\beta_1, \tag{11.5.14}$$

where

$$\begin{aligned}
b_n &= \frac{1}{\mathrm{E}(|\mathbf{w}_1|^2 - \alpha - \mathbf{w}_1'\mathbf{W}_1^*(\mathbf{H}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_1\overline{\mathbf{w}}_1)},\\
\varepsilon_1 &= \beta_1^{-1} - b_n^{-1} = \beta_1^{-1} - \mathrm{E}\beta_1^{-1}\\
&= |\mathbf{w}_1|^2 - (1 + |z|^2) - \mathbf{w}_1'\mathbf{W}_1^*(\mathbf{H}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_1\overline{\mathbf{w}}_1\\
&\quad + \mathrm{E}\mathbf{w}_1'\mathbf{W}_1^*(\mathbf{H}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_1\overline{\mathbf{w}}_1.
\end{aligned}$$

Since the imaginary parts of the denominators of both β_1 and b_n are at least y, we have $|\beta_1| \le y^{-1}$ and $|b_n| < y^{-1}$, so that

$$\begin{aligned}
&|\mathrm{E}\Delta_n(\alpha)\mathrm{E}(|\mathbf{w}_1|^2 - \alpha - \mathbf{w}_1'\mathbf{W}_1^*(\mathbf{H}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_1\overline{\mathbf{w}}_1) - 1|\\
&\le \frac{1}{y^2}\mathrm{E}|\varepsilon_1|^2.
\end{aligned} \tag{11.5.15}$$

Now, let us split the estimation of $\mathrm{E}|\varepsilon_1|^2$ into several parts. Since $\mathrm{E}|x_{11}|^4 \le bn^{\delta(2-\eta)}$, where b is a bound for $\mathrm{E}|x_{11}|^{2+\eta}$, we have

$$\begin{aligned}
&\mathrm{E}\left||\mathbf{w}_1|^2 - (1 + |z|^2)\right|^2\\
&= \mathrm{E}\left|\frac{1}{n}\sum_{j=1}^{n}(|x_{1j}|^2 - 1) - \frac{2}{\sqrt{n}}\Re(\bar{z}x_{11})\right|^2\\
&\le \frac{n-1}{n^2}\mathrm{E}|x_{11}|^4 + \mathrm{E}\left|\frac{1}{n}(|x_{11}|^2 - 1) - \frac{2}{\sqrt{n}}\Re(\bar{z}x_{11})\right|^2\\
&\le \frac{n+1}{n^2}\mathrm{E}|x_{11}|^4 + \frac{4|z|^2}{n}\\
&= O(n^{-1+\delta(2-\eta)}).
\end{aligned}$$

Write $\mathbf{W}_1^*(\mathbf{H}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_1 = (a_{ij})_{i,j=1}^{n-1} := \mathbf{A}$. By Lemma B.26, we have the following estimates

$$\mathrm{E}\left|\frac{1}{n}(\mathbf{x}_1'\mathbf{A}\mathbf{x}_1 - \mathrm{tr}\mathbf{A})\right|^2$$

$$\begin{aligned}
&\le \frac{C}{n^2}\mathrm{E}|x_{11}|^4\mathrm{E}\mathrm{tr}(\mathbf{A}\mathbf{A}^*)\\
&= \frac{C}{n^2}\mathrm{E}|x_{11}|^4\mathrm{E}\mathrm{tr}((\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{H}_1(\mathbf{H}_1-\bar{\alpha}\mathbf{I}_{n-1})^{-1}\mathbf{H}_1)\\
&\le \frac{C}{n^2}\mathrm{E}|x_{11}|^4\mathrm{E}\left[|\alpha|^2\mathrm{tr}(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}(\mathbf{H}_1-\bar{\alpha}\mathbf{I}_{n-1})^{-1}+n-1\right]\\
&\le \frac{C}{n^2}n^{\delta(2-\eta)}[ny^{-2}|\alpha|^2+n]\\
&\le Cn^{-1+2\delta+5\delta\eta},
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left|\tfrac{1}{\sqrt{n}}\bar{z}\mathbf{e}_1'\mathbf{A}\mathbf{x}_1\right|^2 &= \frac{|z|^2}{n}\mathrm{E}\mathbf{e}_1'\mathbf{A}\mathbf{A}^*\mathbf{e}_1\\
&\le \frac{2|z|^2}{n}(1+|z|^2y^{-2})\\
&\le Cn^{-1}y^{-2}\\
&\le Cn^{-1+2\delta\eta},
\end{aligned}$$

$$\begin{aligned}
&\mathrm{E}\left|\frac{1}{n}\mathrm{tr}\mathbf{A}-\mathrm{E}\frac{1}{n}\mathrm{tr}\mathbf{A}\right|^2\\
&= n^{-2}\mathrm{E}|\mathrm{tr}(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{H}_1-\mathrm{E}\mathrm{tr}(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{H}_1|^2\\
&= n^{-2}|\alpha|^2\mathrm{E}|\mathrm{tr}(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}-\mathrm{E}\mathrm{tr}(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}|^2\\
&\le n^{-2}|\alpha|^2\left[\mathrm{E}\left||\mathrm{tr}(\mathbf{H}-\alpha\mathbf{I}_n)^{-1}-\mathrm{E}\mathrm{tr}(\mathbf{H}-\alpha\mathbf{I}_n)^{-1}|+2y^{-1}\right|^2\right]\\
&\le Cn^{-2}|\alpha|^2\left[ny^{-2}+y^{-2}\right] \qquad \text{(by (11.5.12))}\\
&\le Cn^{-1+6\delta\eta}.
\end{aligned}$$

Finally, we need to estimate

$$\begin{aligned}
&\mathrm{E}|\mathbf{e}_1'\mathbf{A}\mathbf{e}_1-\mathrm{E}\mathbf{e}_1'\mathbf{A}\mathbf{e}_1|^2\\
&= \frac{1}{n^2}[\mathrm{E}|\mathbf{v}_1^*(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1-\mathrm{E}\mathbf{v}_1^*(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1|^2],
\end{aligned}$$

where $\frac{1}{\sqrt{n}}\mathbf{v}_1$ is the first column vector (of dimension $n-1$) of $\mathbf{W}_1$, which is also the first column of $\frac{1}{\sqrt{n}}\mathbf{X}_n$ with the first element removed.

Let $\mathbf{X}_{n1}$ be the matrix obtained from $\mathbf{X}_n$ by removing its first row and column. Note that $\mathbf{H}_1=\widehat{\mathbf{H}}_1+\frac{1}{n}\mathbf{v}_1\mathbf{v}_1^*$, where $\widehat{\mathbf{H}}_1=(\frac{1}{\sqrt{n}}\mathbf{X}_{n1}-z\mathbf{I}_{n-1})(\frac{1}{\sqrt{n}}\mathbf{X}_{n1}-z\mathbf{I}_{n-1})^*$. Note that

$$\frac{1}{n}\mathbf{v}_1^*(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1$$

$$= 1 - \frac{1}{1 + \frac{1}{n}\mathbf{v}_1^*(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1}$$
$$= 1 - \frac{1}{1 + \frac{1}{n}\mathrm{Etr}(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1}}$$
$$- \frac{\frac{1}{n}(\mathbf{v}_1^*(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1 - \mathrm{Etr}(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1})}{(1 + \frac{1}{n}\mathbf{v}_1^*(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1)(1 + \frac{1}{n}\mathrm{Etr}(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1})}.$$

Also,

$$\Im\left(\alpha\left(1 + \frac{1}{n}\mathbf{v}_1'(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1}\bar{\mathbf{v}}_1\right)\right) \geq y. \tag{11.5.16}$$

The two observations above yield

$$\frac{1}{n^2}\mathrm{E}|\mathbf{v}_1^*(\mathbf{H}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1 - \mathrm{E}\mathbf{v}_1 * (\mathbf{H}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1|^2$$
$$\leq \mathrm{E}\left|\frac{\frac{1}{n}(\mathbf{v}_1^*(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1 - \mathrm{Etr}(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1})}{(1 + \frac{1}{n}\mathbf{v}_1^*(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1)(1 + \frac{1}{n}\mathrm{Etr}(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1})}\right|^2$$
$$\leq |\alpha|^2 y^{-2} n^{-2}\mathrm{E}|\mathbf{v}_1^*(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1 - \mathrm{Etr}(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1}|^2$$
$$\leq |\alpha|^2 y^{-2} n^{-2}\left[\mathrm{E}\left|\mathbf{v}_1^*(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1 - \mathrm{tr}(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1}\right|^2\right.$$
$$\left. + \mathrm{E}\left|\mathrm{tr}(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1} - \mathrm{Etr}(\widehat{\mathbf{H}}_1 - \alpha\mathbf{I}_{n-1})^{-1}\right|^2\right]$$
$$\leq Cn^{-1}|\alpha|^2 y^{-4} \leq Cn^{-1+8\delta\eta}. \tag{11.5.17}$$

Here, in the last inequality, the first term is less than Cny^{-2} by Lemma B.26 and the second term is less than Cny^{-2}, which can be proved by the same method as for (11.5.12). Combining the above, we obtain

$$\frac{1}{ny^2}\mathrm{E}|\varepsilon_1|^2 = o(\delta_n^3).$$

Substituting this into (11.5.15), we get

$$|\mathrm{E}\Delta_n(\alpha)\mathrm{E}(|\mathbf{w}_1|^2 - \alpha - \mathbf{w}_1'\mathbf{W}_1(\mathbf{H}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_1\bar{\mathbf{w}}_1) - 1|$$
$$\leq o(\delta_n^3). \tag{11.5.18}$$

To conclude the proof of the lemma, we need to estimate

$$\mathrm{E}(|\mathbf{w}_1|^2 - \alpha - \mathbf{w}_1'\mathbf{W}_1^*(\mathbf{H}_1 - \alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_1\overline{\mathbf{w}}_1). \tag{11.5.19}$$

First, we have

$$\mathrm{E}|\mathbf{w}_1|^2 = 1 + |z|^2.$$

Second, we have

$$\begin{aligned}
&\mathrm{E}\mathbf{w}_1'\mathbf{W}_1^*(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_1\overline{\mathbf{w}}_1\\
&=\frac{1}{n}\mathrm{Etr}(\mathbf{W}_1^*(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_1)\\
&\qquad+|z|^2\mathrm{E}\mathbf{e}_1'\mathbf{W}_1^*(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_1\mathbf{e}_1\\
&=\frac{n-1}{n}+\frac{\alpha}{n}\mathrm{Etr}(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}\\
&\qquad+|z|^2\left(1-\mathrm{E}\frac{1}{1+\frac{1}{n}\mathbf{v}_1^*(\widehat{\mathbf{H}}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1}\right).
\end{aligned}$$

By Lemma A.1.12, we have

$$\left|\frac{1}{n}\mathrm{Etr}(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}-\mathrm{E}\Delta_n(\alpha)\right|\le\frac{1}{ny}.$$

By (6.9), we have

$$|\mathrm{tr}(\widehat{\mathbf{H}}_1-\alpha\mathbf{I}_{n-1})^{-1}-\mathrm{tr}(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}|\le\frac{1}{y}.$$

Furthermore, by the above, Lemma B.26, (11.5.16), and (11.5.12),

$$\begin{aligned}
&\left|\mathrm{E}\frac{1}{1+\frac{1}{n}\mathbf{v}_1^*(\widehat{\mathbf{H}}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1}-\frac{1}{1+\mathrm{E}\Delta_n(\alpha)}\right|\\
&\le\left|\frac{\mathrm{E}\frac{1}{n}\mathbf{v}_1^*(\widehat{\mathbf{H}}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1-\mathrm{E}\Delta_n(\alpha)}{(1+\mathrm{E}\Delta_n(\alpha))^2}\right|\\
&\quad+\mathrm{E}\frac{|\frac{1}{n}\mathbf{v}_1^*(\widehat{\mathbf{H}}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1-\mathrm{E}\Delta_n(\alpha)|^2}{|(1+\mathrm{E}\Delta_n(\alpha))^2(1+\frac{1}{n}\mathbf{v}_1^*(\widehat{\mathbf{H}}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1)|}\\
&\le 2|\alpha|^2n^{-1}y^{-3}+|\alpha|^3y^{-3}\mathrm{E}\left|\frac{1}{n}\mathbf{v}_1^*(\widehat{\mathbf{H}}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1-\mathrm{E}\Delta_n(\alpha)\right|^2\\
&\le 2|\alpha|^2n^{-1}y^{-3}+C|\alpha|^2y^{-3}\left[\mathrm{E}\left|\frac{1}{n}\mathbf{v}_1^*(\widehat{\mathbf{H}}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{v}_1\right.\right.\\
&\quad\left.\left.-\frac{1}{n}\mathrm{tr}(\widehat{\mathbf{H}}_1-\alpha\mathbf{I}_{n-1})^{-1}\right|^2+n^{-2}y^{-2}+\mathrm{E}|\Delta_n(\alpha)-\mathrm{E}\Delta_n(\alpha)|^2\right]\\
&\le Cn^{-1}(|\alpha|^2y^{-3}+|\alpha|^3y^{-5})\\
&\le Cn^{-1+8\delta\eta}.
\end{aligned}\tag{11.5.20}$$

Combining the above, we obtain an approximation for quantity (11.5.19) as

$$\begin{aligned}
&\mathrm{E}(|\mathbf{w}_1|^2-\alpha-\mathbf{w}_1'\mathbf{W}_1^*(\mathbf{H}_1-\alpha\mathbf{I}_{n-1})^{-1}\mathbf{W}_1\overline{\mathbf{w}}_1)\\
&=-\alpha-\alpha\mathrm{E}\Delta_n(\alpha)+\frac{|z|^2}{1+\mathrm{E}\Delta_n(\alpha)}+O(n^{-1+8\delta\eta}).
\end{aligned}$$

Substituting this into (11.5.18), the proof of the lemma is complete.

Lemma 11.16. *Under the conditions of Theorem 11.4, for any $M_2 > M_1 > 0$, and δ defined in Lemma 11.14,*

$$\sup_{M_1 \le |z| \le M_2} ||\nu_n(\cdot, z) - \nu(\cdot, z)|| := \sup_{x, M_1 \le |z| \le M_2} |\nu_n(x,z) - \nu(x,z)|$$
$$= O_{\text{a.s.}}(n^{-\delta\eta/2}). \tag{11.5.21}$$

Remark 11.17. Lemma 11.16 is used only in proving (11.2.2) for a suitably chosen ε_n. From the proof of the lemma and comparson with Theorem 8.10, one can see that a better rate of this convergence can be obtained by more detailed calculus. As the rate given in (11.5.21) is enough for our purpose, we restrict ourselves to the weaker result (11.5.21) by using a simpler proof rather than trying to get a better rate by long and tedious arguments.

Proof. We shall prove (11.5.21) by employing Corollary B.15. The supports of all $\nu(\cdot, z)$ are bounded for all $z \in T$. Therefore, we may select a constant N such that, for some absolute constant C,

$$\begin{aligned} &\|\nu_n(\cdot, z) - \nu\cdot, z)\| \\ &\le C\Bigg(\int_{|x|\le N} |\Delta_n(\alpha) - \Delta(\alpha)| dx \\ &\quad + y_n^{-1} \sup_x \int_{|y| \le 2y_n} |\nu_n(x+y, z) - \nu(x,z)| dy\Bigg) \\ &\le C\left(\int_{|x|\le N} |\Delta_n(\alpha) - \Delta(\alpha)| dx + \sqrt{y_n}\right), \end{aligned} \tag{11.5.22}$$

where $\alpha = x + iy_n$ and the last step of the inequality above follows from (11.4.2).

To complete the proof of the lemma, we need only estimate the integral of (11.5.22). To this end, consider a realization in which (11.5.5) holds. We first prove that for $\alpha = x + iy$, $|x| \le N$, $|x - x_2| \ge \varepsilon_1$ (also $|x - x_1| \ge \varepsilon_1$ if $|z| < 1$), $y \ge y_n$, $M_1 \le |z| \le M_2$, and all large n,

$$|\Delta_n(\alpha) - \Delta(\alpha)| < \frac{1}{3}\varepsilon_0\delta_n, \tag{11.5.23}$$

where ε_0 (and ε_1 in what follows) is defined in Lemma 11.10.

Equation (11.4.1) has three solutions denoted by $m_1(\alpha) = \Delta(\alpha)$, $m_2(\alpha)$, and $m_3(\alpha)$. As mentioned earlier, all three solutions are analytic in α on the upper half complex plane.

By Lemma 11.10, we assume that (11.4.7)–(11.4.10). By Lemma 11.14, there is an integer n_0 such that, for all $n \ge n_0$,

$$|(\Delta_n - m_1)(\Delta_n - m_2)(\Delta_n - m_3)| = o(\delta_n) \le \frac{4}{27}\varepsilon_0^3\delta_n. \tag{11.5.24}$$

Now, choose an $\alpha_0 = x_0 + iy_0$ with $|x_0| \le N$, $y_0 > 0$, and $\min_{k=1,2}(|x_0 - x_k|) \ge \varepsilon_1$. Fix $z \in T$. We first show that (11.5.23) holds for $\alpha = \alpha_0$. As argued earlier, $\Delta_n(\alpha_0)$ converges to $\Delta(\alpha_0)$. Therefore, we can find an $n_1 > n_0$ such that $|\Delta_n(\alpha_0) - \Delta(\alpha_0)| < \varepsilon_0/3$ for all $n > n_1$. Therefore, for these n,

$$\begin{aligned}
&\min_{k=2,3}(|\Delta_n(\alpha_0) - m_k(\alpha_0)|) \\
&\ge \min_{k=2,3}(|\Delta(\alpha_0) - m_k(\alpha_0)| - |\Delta_n(\alpha_0) - \Delta(\alpha_0)|) > \frac{2}{3}\varepsilon_0.
\end{aligned}$$

This and (11.5.24) imply

$$|\Delta_n(\alpha_0) - \Delta(\alpha_0)| = o(\delta_n) \le \frac{1}{3}\varepsilon_0\delta_n. \tag{11.5.25}$$

Next, we claim that (11.5.23) is true for all $n > n_0$, $y \ge y_n$, and $|x| \le N$, $\max_{k=1,2}(|x - x_k|) \ge \varepsilon_1$.

By (11.4.8) and (11.5.24), one sees that (11.5.23) is implied by

$$\min_{k=2,3}(|\Delta_n(\alpha) - m_k(\alpha)|) > \frac{2}{3}\varepsilon_0. \tag{11.5.26}$$

Note that both Δ_n and $m_j(\alpha)$, $j = 2, 3$, are continuous functions in both α and z. Therefore, if (11.5.23) is ever false, an α and z would exist for which $|\Delta_n(\alpha) - \Delta(\alpha)| = \frac{1}{3}\varepsilon_0\delta_n$. As with $\alpha = \alpha_0$, this equality, together with (11.4.8), implies (11.5.26), a contradiction.

Finally, we consider the case where $|\alpha - x_k| \le \varepsilon_1$, with $k = 1$ or 2. We see that (11.4.7), (11.4.9), and (11.4.10) imply that

$$|\Delta_n(\alpha) - \Delta(\alpha)| \le o(\delta_n/\sqrt{|\alpha - x_k|}). \tag{11.5.27}$$

This, together with (11.5.22) and (11.5.23), implies (11.5.21). The proof of Lemma 11.16 is complete.

11.6 Proofs of (11.2.3) and (11.2.4)

For this section, we return to the original assumptions on the variables. Note that truncation techniques cannot be applied from this point on. The proofs of (11.2.3) and (11.2.4) are almost the same. Thus, we shall only show (11.2.3), that is, with probability 1,

$$\int_{z\in T}\left|\int_0^{\varepsilon_n}\log x\nu_n(dx,z)\right|dtds \to 0, \tag{11.6.1}$$

where $\varepsilon_n = e^{-n^{\delta\eta}}$.

Recall that $\mathbf{W}(z) = \frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I}$ is nonsingular when $z \in T$ and z does not coincide with any eigenvalues of $\frac{1}{\sqrt{n}}\mathbf{X}_n$, an almost sure event. For the following, we may assume $\mathbf{W}(z)$ is nonsingular. Now, denote respectively by $\mathbf{Z}_1$ and $\mathbf{Z}$ the matrix of the first two columns of $\mathbf{W}(z)$ and that formed by the last $n-2$ columns. Let $\Lambda_1 \le \cdots \le \Lambda_n$ denote the eigenvalues of the matrix $\mathbf{W}^*(z)\mathbf{W}(z)$ and $\eta_1 \le \cdots \le \eta_{n-2}$ denote the eigenvalues of $\mathbf{Z}^*\mathbf{Z}$. Then, for any $k \le n-2$, by the interlacing theorem (see Theorem A.43), we have $\Lambda_k \le \eta_k \le \Lambda_{k+2}$. We also have $\det(\mathbf{W}^*(z)\mathbf{W}(z)) = \det(\mathbf{Z}^*\mathbf{Z})\det(\mathbf{Z}_1^*\mathbf{Q}\mathbf{Z}_1)$, where $\mathbf{Q} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^*\mathbf{Z})^{-1}\mathbf{Z}^*$. This identity can be written as

$$\sum_{k=1}^{n} \log(\Lambda_k) = \log(\det(\mathbf{Z}_1^*\mathbf{Q}\mathbf{Z}_1)) + \sum_{k=1}^{n-2} \log(\eta_k).$$

If ℓ is the smallest integer such that $\eta_\ell \ge \varepsilon_n$, then $\Lambda_{\ell-1} < \varepsilon_n$ and $\Lambda_{\ell+2} > \varepsilon_n$. Therefore, we have

$$\begin{aligned} 0 > \int_0^{\varepsilon_n} \log x \nu(dx, z) &= \frac{1}{n} \sum_{\Lambda_k < \varepsilon_n} \log \Lambda_k \\ &\ge \frac{1}{n} \min\{\log(\det(\mathbf{Z}_1^*\mathbf{Q}\mathbf{Z}_1)), 0\} + \frac{1}{n} \sum_{\eta_k < \varepsilon_n} \log(\eta_k) \\ &\quad - \frac{2}{n} \log(\max(\Lambda_n, 1)). \end{aligned} \tag{11.6.2}$$

To prove (11.6.1), we first estimate the integral of $\frac{1}{n}\log(\det(\mathbf{Z}_1^*\mathbf{Q}\mathbf{Z}_1))$ with respect to s and t. Note that $\mathbf{Q}$ is a projection matrix of rank 2. Hence, there are two orthogonal complex unit vectors $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ such that $\mathbf{Q} = \boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^* + \boldsymbol{\gamma}_2\boldsymbol{\gamma}_2^*$. Denote the i-th column vector of $\mathbf{W}$ by $\mathbf{w}_i$. Then we have

$$\frac{1}{n}\log(\det(\mathbf{Z}_1^*\mathbf{Q}\mathbf{Z}_1)) = \frac{1}{n}\log(|\boldsymbol{\gamma}_1^*\mathbf{w}_1\boldsymbol{\gamma}_2^*\mathbf{w}_2 - \boldsymbol{\gamma}_2^*\mathbf{w}_1\boldsymbol{\gamma}_1^*\mathbf{w}_2|^2).$$

Define the random sets

$$\mathcal{E} = \Big\{(s,t) : |\boldsymbol{\gamma}_1^*\mathbf{w}_1\boldsymbol{\gamma}_2^*\mathbf{w}_2 - \boldsymbol{\gamma}_2^*\mathbf{w}_1\boldsymbol{\gamma}_1^*\mathbf{w}_2| \ge n^{-14}, \\ \Big|\frac{1}{\sqrt{n}}\mathbf{X}_1\Big| \le n, \Big|\frac{1}{\sqrt{n}}\mathbf{X}_2\Big| \le n\Big\}$$

and

$$\mathcal{F} = \Big\{(s,t) : \big|\boldsymbol{\gamma}_1^*\mathbf{w}_1\boldsymbol{\gamma}_2^*\mathbf{w}_2 - \boldsymbol{\gamma}_2^*\mathbf{w}_1\boldsymbol{\gamma}_1^*\mathbf{w}_2\big| < n^{-14}, \\ \Big|\frac{1}{\sqrt{n}}\mathbf{x}_1\Big| \le n, \Big|\frac{1}{\sqrt{n}}\mathbf{x}_2\Big| \le n\Big\},$$

where $\mathbf{x}_i$ is the i-th column of $\mathbf{X}_n$. It is trivial to see that

$$\mathrm{P}\Big(\Big|\frac{1}{\sqrt{n}}\mathbf{x}_1\Big| > n \ \text{ or } \ \Big|\frac{1}{\sqrt{n}}\mathbf{x}_2\Big| > n\Big) < 2n^{-2}. \tag{11.6.3}$$

When $|\frac{1}{\sqrt{n}}\mathbf{x}_1| \le n$ and $|\frac{1}{\sqrt{n}}\mathbf{x}_2| \le n$, we have $\det(\mathbf{Z}_1^*\mathbf{Q}\mathbf{Z}_1) = |\boldsymbol{\gamma}_1^*\mathbf{w}_1\boldsymbol{\gamma}_2^*\mathbf{w}_2 - \boldsymbol{\gamma}_2^*\mathbf{w}_1\boldsymbol{\gamma}_1^*\mathbf{w}_2|^2 \le Cn^4$. Thus,

$$\frac{1}{n}\int_{z\in T} |I_{\mathcal{E}} \log(\det(\mathbf{Z}_1^*\mathbf{Q}\mathbf{Z}_1))|\, dtds \le Cn^{-1}\log n \to 0. \tag{11.6.4}$$

On the other hand, for any $\varepsilon > 0$, we have

$$\begin{aligned} &\mathrm{P}\left(\frac{1}{n}\int_{z\in T} |I_{\mathcal{F}} \log(\det(\mathbf{Z}_1^*\mathbf{Q}\mathbf{Z}_1))|\, dtds \ge \varepsilon\right) \\ &\le \frac{1}{\varepsilon n}\int_{z\in T} \mathrm{E}\left|I_{\mathcal{F}} \log(|\boldsymbol{\gamma}_1^*\mathbf{w}_1\boldsymbol{\gamma}_2^*\mathbf{w}_2 - \boldsymbol{\gamma}_2^*\mathbf{w}_1\boldsymbol{\gamma}_1^*\mathbf{w}_2|^2)\right| dtds. \end{aligned} \tag{11.6.5}$$

Note that the elements of $\sqrt{n}\mathbf{w}_1 = \mathbf{x}_1 - z\sqrt{n}\mathbf{e}_1$ and $\sqrt{n}\mathbf{w}_2 = \mathbf{x}_2 - z\sqrt{n}\mathbf{e}_2$ are independent of each other and the joint densities of their real and imaginary parts have a common upper bound K_d. Also, they are independent of $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$. Therefore, by Corollary 11.21, the conditional joint density of the real and imaginary parts of $\sqrt{n}\boldsymbol{\gamma}_1^*\mathbf{w}_1$, $\sqrt{n}\boldsymbol{\gamma}_2^*\mathbf{w}_2$, $\sqrt{n}\boldsymbol{\gamma}_2^*\mathbf{w}_1$, and $\sqrt{n}\boldsymbol{\gamma}_1^*\mathbf{w}_2$, when $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are given, is bounded by $(2K_dn)^4$. Hence, the conditional joint density of the real and imaginary parts of $\boldsymbol{\gamma}_1^*\mathbf{w}_1$, $\boldsymbol{\gamma}_2^*\mathbf{w}_2$, $\boldsymbol{\gamma}_2^*\mathbf{w}_1$, and $\boldsymbol{\gamma}_1^*\mathbf{w}_2$, when $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are given, is bounded by $K_d^4 2^4 n^8$. Set $\mathbf{x} = (\boldsymbol{\gamma}_1^*\mathbf{w}_1, \boldsymbol{\gamma}_2^*\mathbf{w}_1)'$ and $\mathbf{y} = (\mathbf{w}_2^*\boldsymbol{\gamma}_2, -\mathbf{w}_2^*\boldsymbol{\gamma}_1)'$. Note that, by Corollary 11.21, the joint density of $\mathbf{x}$ and $\mathbf{y}$ is bounded by $K_d^4 2^4 n^8$.

If $|\frac{1}{\sqrt{n}}\mathbf{x}_1| \le n$, $|\frac{1}{\sqrt{n}}\mathbf{x}_2| \le n$, then $\max(|\mathbf{x}|, |\mathbf{y}|) \le 2n + 2|z| \le 2n + 2M$. Applying Lemma 11.22 with $f(t) = \log t$, $M = \mu = 1$, we obtain

$$\begin{aligned} &\frac{2}{n}\int_{z\in T} \left|\mathrm{E}\left(I_{(|\mathbf{y}^*\mathbf{x}|<n^{-14},\, |\frac{1}{\sqrt{n}}\mathbf{x}_1|\le n,\, |\frac{1}{\sqrt{n}}\mathbf{x}_2|\le n)} \log(|\mathbf{y}^*\mathbf{x}|)\right| \gamma_1, \gamma_2\right)\right| dtds \\ &\le Cn^{12}n^{-14} \le Cn^{-2} \end{aligned} \tag{11.6.6}$$

for some positive constant C.

From (11.6.3), (11.6.5), and (11.6.6), it follows that

$$\frac{1}{n}\int_{z\in T} |I_{\mathcal{F}} \log(\det(\mathbf{Z}_1^*\mathbf{Q}\mathbf{Z}_1))|\, dtds \to 0, \text{ a.s.} \tag{11.6.7}$$

Next, we estimate the second term in (11.6.2). Using the fact that $x\ln x$ is decreasing on $(0, e^{-1})$, we have

$$\frac{1}{n}\left|\sum_{\eta_k<\varepsilon_n} \log(\eta_k)\right| \le n^{\delta\eta-1}\varepsilon_n \sum_{k=1}^{n-2} \frac{1}{\eta_k}$$

$$= n^{\delta\eta-1}\varepsilon_n \mathrm{tr}(\mathbf{Z}^*\mathbf{Z})^{-1} = n^{\delta\eta-1}\varepsilon_n \sum_{k=3}^{n} \frac{1}{\mathbf{w}_k^*\mathbf{Q}_k\mathbf{w}_k}$$

$$= n^{\delta\eta-1}\varepsilon_n \sum_{k=3}^{n} \frac{1}{|\mathbf{w}_k^*\boldsymbol{\gamma}_{k1}|^2 + |\mathbf{w}_k^*\boldsymbol{\gamma}_{k2}|^2 + |\mathbf{w}_k^*\boldsymbol{\gamma}_{k3}|^2}, \tag{11.6.8}$$

where, for each k, $\boldsymbol{\gamma}_{kj}$, $j = 1, 2, 3$, are orthonormal complex vectors such that $\mathbf{Q}_k = \boldsymbol{\gamma}_{k1}\boldsymbol{\gamma}_{k1}^* + \boldsymbol{\gamma}_{k2}\boldsymbol{\gamma}_{k2}^* + \boldsymbol{\gamma}_{k3}\boldsymbol{\gamma}_{k3}^*$, which is the projection matrix onto the orthogonal complement of the space spanned by the third,$\cdots$, $k-1$st,$k+1$st,$\cdots$, n-th columns of $\mathbf{W}(z)$.

As in the proof of (11.6.7), one can show that the conditional joint density of the real and imaginary parts of $\mathbf{w}_k^*\boldsymbol{\gamma}_{k1}$, $\mathbf{w}_k^*\boldsymbol{\gamma}_{k2}$, and $\mathbf{w}_k^*\boldsymbol{\gamma}_{k3}$ when $\boldsymbol{\gamma}_{kj}$, $j = 1, 2, 3$, are given, is bounded by $CK_d^3 n^6$. Therefore, we have

$$\begin{aligned}
& n^{\delta\eta-1}\varepsilon_n \sum_{k=3}^{n} \int_{z\in T} \mathrm{E}\left(\frac{dtds}{|\mathbf{w}_k^*\boldsymbol{\gamma}_{k1}|^2 + |\mathbf{w}_k^*\boldsymbol{\gamma}_{k2}|^2 + |\mathbf{w}_k^*\boldsymbol{\gamma}_{k3}|^2} \right) \\
& \le Cn^{\delta\eta-1}\varepsilon_n \left(K_d^3 n^7 \int\cdots\int_{u_1^2+\cdots+u_6^2<1} \frac{du_1\cdots du_6}{u_1^2+\cdots+u_6^2} + 1 \right) \\
& \le Cn^{6+\delta\eta}\varepsilon_n \quad \text{by a polar transformation} \\
& \le Cn^{-2}.
\end{aligned} \tag{11.6.9}$$

Therefore, by the Borel-Cantelli lemma,

$$n^{\delta\eta-1}\varepsilon_n \sum_{k=3}^{n} \int_{z\in T} \left(\frac{dtds}{|\mathbf{w}_k^*\boldsymbol{\gamma}_{k1}|^2 + |\mathbf{w}_k^*\boldsymbol{\gamma}_{k2}|^2 + |\mathbf{w}_k^*\boldsymbol{\gamma}_{k3}|^2} \right) \to 0, \text{ a.s.},$$

and hence, with probability 1,

$$\frac{1}{n}\int_{z\in T} \left| \sum_{\eta_k<\varepsilon_n} \log(\eta_k) \right| dtds \le n^{\delta\eta-1}\varepsilon_n \int_{z\in T} \left| \sum_{k=1}^{n-2} \frac{1}{\eta_k} \right| dtds \to 0. \tag{11.6.10}$$

Finally, we estimate the integral of the third term in (11.6.2). By Theorem 3.7, we have

$$\begin{aligned}
\frac{1}{n}\max(\Lambda_n, 1) & \le \frac{2}{n}\left(\max\left(\|\frac{1}{\sqrt{n}}\mathbf{X}_n\|^2, 1 \right) + |z|^2 \right) \\
& \le 2\int_1^\infty x dF^{\mathbf{S}_n}(x) + \frac{|z|^2}{n} \\
& \to 2\int_1^4 x dF_1(x), \text{ a.s.}
\end{aligned} \tag{11.6.11}$$

We conclude that

$$\frac{2}{n}\int_{z\in T}\log(\max(\Lambda_n,1))dtds \to 0, \text{ a.s.} \tag{11.6.12}$$

Hence, (11.6.1) follows from (11.6.7), (11.6.10), and (11.6.12).

11.7 Proof of Theorem 11.4

In Section 11.3, the problem is reduced to showing (11.3.9). Recalling the definitions of $g_n(s,t)$ and $g(s,t)$, we have, by integration by parts,

$$\begin{aligned}
&\left|\int_{z\in T}(g_n(s,t)-g(s,t))e^{ius+itv}dtds\right| \\
&=\left|-\int_{z\in T}iu\tau(s,t)dtds+\int_{|t|\le A^3}[\tau(A,t)dt-\tau(-A,t)]dt\right. \\
&\quad -\int_{|t|\le 1+\varepsilon}[\tau(\sqrt{(1+\varepsilon)^2-t^2},t)+\tau(-\sqrt{(1+\varepsilon)^2-t^2},t)]dt \\
&\quad \left.+\int_{|t|\le 1-\varepsilon}[\tau(\sqrt{(1-\varepsilon)^2-t^2},t)-\tau(-\sqrt{(1-\varepsilon)^2-t^2},t)]dt\right|,
\end{aligned} \tag{11.7.1}$$

where

$$\tau(s,t)=e^{ius+itv}\int_0^\infty \log x d(\nu_n(x,z)-\nu(x,z)).$$

Let $\varepsilon_n = e^{-n^{\delta\eta}}$. In the last section, we proved that

$$\int_{z\in T}\left|\int_0^{\varepsilon_n}\log x\nu_n(dx,z)\right|dtds \to 0, \quad \text{a.s.}$$

By (11.4.2), we have

$$\int_{z\in T}\left|\int_0^{\varepsilon_n}\log x\nu(dx,z)\right|dtds \to 0.$$

By (11.6.11), we have the existence of a_n, for which the support of $\nu_n(\cdot,z)$ lies in $[0,a_n n]$ for all $z\in T$ and converges to a nonrandom value. Therefore, from Lemma 11.16, with probability 1,

$$\begin{aligned}
&\int_{z\in T}\left|\int_{\varepsilon_n}^\infty \log x d(\nu_n(x,z)-\nu(x,z))\right|dtds \\
&=\int_{z\in T}\left|\int_{\varepsilon_n}^{an}\log x d(\nu_n(x,z)-\nu(x,z))\right|dtds
\end{aligned}$$

$$\leq [|\log(\varepsilon_n)| + \log(an)] \max_{z \in T} \|\nu_n(\cdot, z) - \nu(\cdot, z)\| \to 0.$$

This proves that

$$iu \int_{z \in T} \tau(s,t) dt ds \to 0.$$

By applying (11.2.4), we can similarly prove that

$$\int_{|t| \leq A^3} \tau(\pm A, t) dt \to 0, \text{ a.s.},$$

$$\int_{|t| \leq 1+\varepsilon} \tau(\pm\sqrt{(1+\varepsilon)^2 - t^2}, t) dt \to 0, \text{ a.s.},$$

and

$$\int_{|t| \leq 1-\varepsilon} \tau(\pm\sqrt{(1-\varepsilon)^2 - t^2}, t) dt \to 0, \text{ a.s.}$$

The proof of Theorem 11.4 is complete.

11.8 Comments and Extensions

11.8.1 Relaxation of Conditions Assumed in Theorem 11.4

1. On the smoothness of the underlying distribution
The assumption that the real and imaginary parts of the entries of the matrix $\mathbf{X}_n$ have a bounded joint density is too restrictive because the circular law for a real Gaussian matrix does not follow from Theorem 11.4. In what follows, we shall extend Theorem 11.4 to a more general case to cover the real Gaussian case and in general to random variables with bounded densities.

Theorem 11.18. *Assume that there are two directions such that the conditional density of the projection of the underlying random variable onto one direction given the projection onto the other direction is uniformly bounded, and assume that the underlying distribution has zero mean and finite* $2 + \eta$ *moment. Then the circular law holds.*

Sketch of the proof. Suppose that the two directions are $(\cos(\theta_j), \sin(\theta_j))$, $j = 1, 2$. Then, the density condition is equivalent to:

> *The conditional density of the linear combination* $\Re(x_{11})\cos(\theta_1) + \Im(x_{11})\sin(\theta_1) = \Re(e^{-i\theta_1} x_{11})$ *given* $\Re(x_{11})\cos(\theta_2) + \Im(x_{11})\sin(\theta_2) = \Re(e^{-i\theta_2} x_{11}) = \Im(ie^{-i\theta_2} x_{11})$ *is bounded.*

Consider the matrix $\mathbf{Y}_n = (y_{jk}) = e^{-i\theta_2+i\pi/2}\mathbf{X}_n$. The circular law $\frac{1}{\sqrt{n}}\mathbf{X}_n$ is obviously equivalent to the circular law for $\frac{1}{\sqrt{n}}\mathbf{Y}_n$. Then, the density condition for $\mathbf{Y}_n$ then becomes:

The conditional density of $\Re(y_{11})\sin(\theta_2-\theta_1)+\Im(y_{11})\cos(\theta_2-\theta_1)$ *given* $\Im(y_{11})$ *is bounded.*

This condition simply implies that $\sin(\theta_2-\theta_1)\neq 0$. Thus, the density condition is further equivalent to:

The conditional density of $\Re(y_{11})$ *given* $\Im(y_{11})$ *is bounded.*

Therefore, we shall prove Theorem 11.18 under this latter condition.

Examining the proof of Theorem 11.4, one finds that it is sufficient to prove inequalities (11.6.7) and (11.6.10) under the new density condition. We start with the proof of (11.6.7) from (11.6.5). Rewrite

$$\log(|\mathbf{y}^*\mathbf{x}|^2) = \log(|\mathbf{y}|^2) + \log(|\tilde{\mathbf{y}}^*\mathbf{x}|^2),$$

where $\tilde{\mathbf{y}} = \mathbf{y}/|\mathbf{y}|$.

Denote by $\mathbf{x}_{jr}$ and $\mathbf{x}_{ji}$ the real and imaginary parts of the vector $\mathbf{x}_j$. Since $i\boldsymbol{\gamma}_1$ also yields $\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^*$, we may, without loss of generality, assume that $|\boldsymbol{\gamma}_{1r}|\geq 1/\sqrt{2}$. Then, we have

$$\begin{aligned}|\mathbf{y}|^2 &= \mathbf{w}_2^*(\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^*+\boldsymbol{\gamma}_2\boldsymbol{\gamma}_2^*)\mathbf{w}_2\\ &\geq (\boldsymbol{\gamma}_{1r}'\mathbf{w}_{2r}+\boldsymbol{\gamma}_{1i}'\mathbf{w}_{2i})^2.\end{aligned}$$

Applying Lemma 11.20, we find that the conditional density of $\boldsymbol{\gamma}_{1r}'\mathbf{w}_{2r}+\boldsymbol{\gamma}_{1i}'\mathbf{w}_{2i}$ when $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, and $\mathbf{w}_{2i}$ are given is bounded by $2K_d n$. Therefore,

$$\begin{aligned}&\frac{1}{n}\int_{z\in T}\left|\mathrm{E}\left(I_{(|\mathbf{y}|^2<n^{-14},\,|\frac{1}{\sqrt{n}}\mathbf{x}_2|\leq n)}\log(|\mathbf{y}|^2)\right)\right|dtds\\ &\leq\frac{1}{n}\int_{z\in T}\mathrm{E}\Bigg(I\left(|\boldsymbol{\gamma}_{1r}'\mathbf{w}_{2r}+\boldsymbol{\gamma}_{1i}'\mathbf{w}_{2i}|^2<n^{-14},\ \left|\frac{1}{\sqrt{n}}\mathbf{x}_2\right|\leq n\right)\\ &\qquad\times|\log(|\boldsymbol{\gamma}_{1r}'\mathbf{w}_{2r}+\boldsymbol{\gamma}_{1i}'\mathbf{w}_{2i}|^2)|\,\Bigg|\,\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\mathbf{w}_{2i}\Bigg)dtds\\ &\leq CK_d\int_0^{n^{-7}}|\log x|dx\leq Cn^{-7}\log n\end{aligned}\tag{11.8.1}$$

for some positive constant C.

Rewrite

$$|\tilde{\mathbf{y}}^*\mathbf{x}|^2 = (\boldsymbol{\beta}_1\mathbf{w}_{1r}+\boldsymbol{\zeta}_1\mathbf{w}_{1i})^2+(\boldsymbol{\beta}_2\mathbf{w}_{1r}+\boldsymbol{\zeta}_2\mathbf{w}_{1i})^2,$$

where

$$\begin{aligned}
\boldsymbol{\beta}_1 &= (\tilde{\mathbf{y}}_r', \tilde{\mathbf{y}}_i')(\gamma_{1r}, \gamma_{2r}, -\gamma_{1i}, -\gamma_{2i})', \\
\boldsymbol{\beta}_2 &= (\tilde{\mathbf{y}}_r', \tilde{\mathbf{y}}_i')(-\gamma_{1i}, -\gamma_{2i}, -\gamma_{1r}, -\gamma_{2r})', \\
\boldsymbol{\zeta}_1 &= (\tilde{\mathbf{y}}_r', \tilde{\mathbf{y}}_i')(\gamma_{1i}, \gamma_{2i}, \gamma_{1r}, \gamma_{2r})', \\
\boldsymbol{\zeta}_2 &= (\tilde{\mathbf{y}}_r', \tilde{\mathbf{y}}_i')(\gamma_{1r}, \gamma_{2r}, -\gamma_{1i}, -\gamma_{2i})'.
\end{aligned}$$

It is easy to verify that $|\boldsymbol{\beta}_1|^2 + |\boldsymbol{\beta}_2|^2 = 1$. Thus, we may assume that $|\boldsymbol{\beta}_1| \geq 1/\sqrt{2}$. By Lemma 11.20, the conditional density of $\boldsymbol{\beta}_1'\mathbf{w}_{1r} + \boldsymbol{\zeta}_1'\mathbf{w}_{1i}$ when $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, $\mathbf{y}$, and $\mathbf{w}_{1i}$ are given is bounded by $2K_d n$. Consequently, we can prove that

$$\begin{aligned}
&\frac{1}{n}\int_{z\in T} \left|\mathrm{E}\left(I(|\tilde{\mathbf{y}}^*\mathbf{x}| < n^{-7})\log(|\tilde{\mathbf{y}}^*\mathbf{x}|^2)\right| \gamma_1, \gamma_2, \mathbf{y}, \mathbf{w}_{1i}\right)\Big| dtds \\
&\leq \frac{1}{n}\int_{z\in T}\Bigg|\mathrm{E}\Bigg(I(|\boldsymbol{\beta}_1'\mathbf{w}_{1r} + \boldsymbol{\zeta}_1'\mathbf{w}_{1i}|^2| < n^{-7}) \\
&\qquad \times \log(|\boldsymbol{\beta}_1'\mathbf{w}_{1r} + \boldsymbol{\zeta}_1'\mathbf{w}_{1i}|^2)\Bigg|\gamma_1, \gamma_2, \mathbf{y}, \mathbf{w}_{1i}\Bigg)\Bigg| dtds \\
&\leq CK_d\int_0^{n^{-7}} \log x dx \leq Cn^{-7}\log n.
\end{aligned}$$

This, together with (11.8.1), completes the proof of (11.6.7).

Now, we prove (11.6.10) under the new condition. For each k, consider the $2n\times 6$ matrix $\mathbf{A}$ whose first three columns are $(\gamma_{kjr}', -\gamma_{kji}')'$, $j = 1, 2, 3$, and other three columns are $(\gamma_{kji}', \gamma_{kjr}')'$. Since $\boldsymbol{\gamma}_{kj}$ are orthonormal, we have $\mathbf{A}'\mathbf{A} = \mathbf{I}_6$. Using the same approach as in the proof of Lemma 11.20, one may select a 6×6 submatrix $\mathbf{A}_1$ of $\mathbf{A}$ such that $|\det(\mathbf{A}_1)| \geq n^{-3}$. Within the six rows of $\mathbf{A}_1$, either its first three rows come from the first n rows of $\mathbf{A}$ or its last three come from the last n rows. Without loss of generality, assume that the first three rows of $\mathbf{A}_1$ come from the first n rows of $\mathbf{A}$. Then, consider the Laplace expansion of the determinant of $\mathbf{A}_1$ with respect to these three rows. Within the 20 terms, we may select one with an absolute value not less than $\frac{1}{20}n^{-3}$. This term is the product of a minor from those three rows of $\mathbf{A}_1$ and its cofactor. Note that the absolute value of the entries of $\mathbf{A}$ is not greater than 1. Thus, the absolute value of the cofactor is not greater than 6. Therefore, the absolute value of the minor is not less than $\frac{1}{120}n^{-3}$. Suppose the three columns of the minor come from the first, second, and fourth columns of $\mathbf{A}$; i.e., they come from γ_{k1r}, γ_{k2r}, and γ_{k1i} (the proof of the other 19 cases is similar). Then, as in the proof of Lemma 11.20, one can prove that the conditional joint density of $\gamma_{k1r}'\mathbf{w}_{kr}$, $\gamma_{k2r}'\mathbf{w}_{kr}$, and $\gamma_{k1i}'\mathbf{w}_{kr}$ when $\boldsymbol{\gamma}_{kj}$ and $\mathbf{w}_{ki}$ are given is uniformly bounded by $120K_d^3 n^{4.5}$. Finally, from (11.6.8), we have

$$n^{\delta\eta-1}\varepsilon_n\sum_{k=3}^{n}\left(|\mathbf{w}_k^*\boldsymbol{\gamma}_{k1}|^2 + |\mathbf{w}_k^*\boldsymbol{\gamma}_{k2}|^2 + |\mathbf{w}_k^*\boldsymbol{\gamma}_{k3}|^2\right)^{-1}$$

$$\leq n^{\delta\eta-1}\varepsilon_n \sum_{k=3}^{n}\Big((\mathbf{w}'_{kr}\gamma_{k1r}+\mathbf{w}'_{ki}\gamma_{k1i})^2+(\mathbf{w}'_{kr}\gamma_{k2r}+\mathbf{w}'_{ki}\gamma_{k2i})^2$$
$$+(\mathbf{w}'_{kr}\gamma_{k1i}-\mathbf{w}'_{ki}\gamma_{k1r})^2\Big)^{-1}.$$

Using this and the same approach as given in Section 11.7, one may prove that the right-hand side of the above tends to zero almost surely. Thus, (11.6.10) is proved and consequently Theorem 11.18 follows.

2. Extension to the nonidentical case

Reviewing the proofs of Theorems 11.4 and 11.18, one finds that the moment condition was used only in establishing the convergence rate of $\nu_n(\cdot, z)$. To this end, we only need, for any $\delta > 0$ and some constant $\eta > 0$,

$$\frac{1}{n^2}\sum_{ij} \mathrm{E}|x_{ij}|^{2+\eta} I(|x_{ij}| \geq n^{\delta}) \to 0. \tag{11.8.2}$$

After the convergence rate is established, the proof of the circular law then reduces to showing (11.2.3) and (11.2.4). To guarantee this, we need only the following:

> *There are two directions such that the conditional density of the projection of each random variable x_{ij} onto one direction given the projection onto the other direction is uniformly bounded.* (11.8.3)

Therefore, we have the following theorem.

Theorem 11.19. *Assume that the entries of $\mathbf{X}_n$ are independent and have mean zero and variance* 1. *Also, we assume that conditions (11.8.2) and (11.8.3) are true. Then the circular law holds.*

11.9 Some Elementary Mathematics

Lemma 11.20. *Let $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ be a $p\times n$ real random matrix of n independent column vectors whose probability densities $f_1, \cdots, f_n$, have a common bound K_d, and let $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_k$ $(k < n)$ be k orthogonal real unit n-vectors. Then, the joint density of the random p-vectors $\mathbf{y}_j = \mathbf{X}\boldsymbol{\alpha}_j$, $j = 1, \cdots, k$, is bounded by $K_d^k n^{kp/2}$.*

Proof. Write $\mathbf{C} = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_k)'$, and let $\mathbf{C}(j_1, \cdots, j_k)$ denote the $k \times k$ submatrix formed by the $j_1 \cdots j_k$-th columns of C. By Bennett's formula, we have

$$\sum_{1\leq j_1<\cdots<j_k\leq n} \det{}^2(\mathbf{C}(j_1, \cdots, j_k)) = \det(\mathbf{CC}') = 1.$$

Thus, we may select $1 \le j_1 < \cdots < j_k \le n$, say $1, 2, \cdots, k$ for simplicity, such that $|\det(\mathbf{C}(1, \cdots, k))| \ge n^{-k/2}$. Write $\mathbf{C} = (\mathbf{C}(1, 2, \cdots, k), \mathbf{C}_2)$ and $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, where $\mathbf{X}_1$ is $p \times k$. Let $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_k)$. It is straightforward to show that the transformation $\mathbf{Y} = \mathbf{X}_1\mathbf{C}'(1, 2, \cdots, k) + \mathbf{X}_2\mathbf{C}_2'$, $\mathbf{Z} = \mathbf{X}_2$, has Jacobian $\det^p(\mathbf{C}(1, 2, \cdots, k))$. Furthermore, denote by $\mathbf{c}_1, \cdots, \mathbf{c}_k$, the row vectors of the matrix $\mathbf{C}^{-1}(1, 2, \cdots, k)$. Then, the joint density of $\mathbf{y}_1, \cdots, \mathbf{y}_k$ is given by

$$p(\mathbf{y}_1, \cdots, \mathbf{y}_k) = |\det{}^{-p}(\mathbf{C}(1, 2, \cdots, k))|\mathrm{E}\left(\prod_{i=1}^{k} f_i((\mathbf{Y} - \mathbf{X}_2\mathbf{C}_2')\mathbf{c}_i')\right),$$
$$\le K_d^k n^{kp/2},$$

where $\mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_k)$. The proof of the lemma is complete.

For the complex case, we have the following corollary.

Corollary 11.21. *Assume the vectors and matrices in Lemma 11.20 are complex and the joint distribution density of the real and imaginary parts of* $\mathbf{x}_j$ *are uniformly bounded by* K_d*, and define* $\mathbf{y}_j = \mathbf{X}\overline{\boldsymbol{\alpha}}_j$*. Then, the joint density of the real and imaginary parts of* $\mathbf{y}_1, \cdots, \mathbf{y}_k$ *is bounded by* $K_d^{2k}(2n)^{kp}$*.*

Proof. The proof is similar to the lemma above. Form the $p \times 2n$ matrix $(\mathbf{X}_r, \mathbf{X}_i)$, where respective j-th columns in $\mathbf{X}_r$ and $\mathbf{X}_i$ are the real and imaginary parts of x_j. Each α_j yields two real unit $2n$-vectors, $(\alpha_{jr}', \alpha_{ji}')'$ and $(\alpha_{ji}', -\alpha_{jr}')'$, resulting in $2k$ orthonormal vectors. As above, form the $2k \times 2n$ matrix $\mathbf{C}$ so that $\mathbf{Y} = (\mathbf{X}_r, \mathbf{X}_i)\mathbf{C}'$ is the $p \times 2k$ matrix containing the real and imaginary parts of the y_j's. Let $\mathbf{C}_1$ be the $2k \times 2k$ submatrix for which $|\det \mathbf{C}_1| \ge (2n)^{-k}$. With a rearrangement of the columns of $(\mathbf{X}_r, \mathbf{X}_i)$, we can write the transformation $\mathbf{Y} = \mathbf{X}_1\mathbf{C}_1' + \mathbf{X}_2\mathbf{C}_2'$, $\mathbf{Z} = \mathbf{X}_2$, with $\mathbf{X}_1$ $p \times 2k$, $\mathbf{X}_2$ $p \times (2n - 2k)$, and $\mathbf{C}_2$ $2k \times (2n - 2k)$. Its Jacobian is $\det^p(\mathbf{C}_1)$. Notice there are at most $2k$ columns of $\mathbf{X}_2$, each of whose counterpart is a column of $\mathbf{X}_1$, or, in other words, there are at least $n - 2k$ densities whose pairs of real and imaginary variables are present in $\mathbf{X}_2$. Therefore, the joint density of $\mathbf{Y}$ is bounded by

$$|\det{}^{-p}(\mathbf{C}_1)|K_d^{2k} \le K_d^{2k}(2n)^{kp}.$$

Lemma 11.22. *Suppose that* $f(t)$ *is a function such that* $\int_0^\delta t|f(t)|dt \le M\delta^\mu$ *for some* $\mu > 0$ *and all small* δ*. Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be two complex random* k*-vectors* $(k > 1)$ *whose joint density of the real and imaginary parts of* $\mathbf{x}$ *and* $\mathbf{y}$ *is bounded by* K_d*. Then,*

$$\mathrm{E}(f(|\mathbf{x}^*\mathbf{y}|)I(|\mathbf{x}^*\mathbf{y}| < \delta, |\mathbf{x}| \le K_e, |\mathbf{y}| \le K_e)) \le C_k M\delta^\mu K_d K_e^{4k-4}, \quad (11.9.1)$$

where C_k *is a positive constant depending on* k *only.*

Proof. Note that the measure of the set on which $\mathbf{x} = 0$ is zero. For each $\mathbf{x} \ne 0$, define a unitary $k \times k$ matrix U with $\mathbf{x}/|\mathbf{x}|$ as its first column. Now, make a change of variables $\mathbf{u} = \mathbf{x}$ and $\mathbf{v} = \mathbf{U}^*\mathbf{y}$. It is known that the Jacobian of this variable transformation is 1. This leads to $|\mathbf{x}^*\mathbf{y}| = |\mathbf{u}||v_1|$. Thus,

$$\begin{aligned}
&\mathrm{E}\big(f(|\mathbf{x}^*\mathbf{y}|)I(|\mathbf{x}^*\mathbf{y}| < \delta, |\mathbf{x}| \le K_e, |\mathbf{y}| \le K_e)\big) \\
&= \int \cdots \int f(|\mathbf{u}||v_1|)I(|\mathbf{u}||v_1| < \delta,\ |\mathbf{u}| \le K_e, |\mathbf{v}| \le K_e)p(\mathbf{u}, U\mathbf{v})d\mathbf{u}d\mathbf{v} \\
&\le K_d s_{2k} 2\pi (2K_e)^{2k-2} \int_0^{K_e} \rho_1^{2k-1} d\rho_1 \int_0^{\delta/\rho_1} \rho_2 f(\rho_1\rho_2) d\rho_2 \qquad (11.9.2) \\
&\le K_d s_{2k} 2\pi (2K_e)^{2k-2} (2k-2)^{-1} (K_e)^{2k-2} \int_0^{\delta} t f(t) dt, \qquad (11.9.3)
\end{aligned}$$

where s_{2k} denotes the surface area of the $2k$-dimensional unit sphere. Here, inequality (11.9.2) follows from a polar transformation for the real and imaginary parts of $\mathbf{u}$ (dimension $= 2k$) and from a polar transformation for the real and imaginary parts (dimension$= 2$) of v_1. The lemma now follows from (11.9.3).

11.10 New Developments

From the truncation and the proof of Lemma 11.14, we can see, if we only require that the Stieltjes transform $\Sigma_n(\alpha)$ converge to a limit $\Delta(\alpha)$ for any $z \in \mathbb{C}$ and $\alpha \in \mathbb{C}^+$, that it is enough to assume $\mathrm{E}(x_{11}) = 0$ and $\mathrm{E}|x_{11}^2| = 1$. The assumption $\mathrm{E}|x_{11}|^{2+\eta} < \infty$ is merely to establish some rate for r_n so that the rate of $\varepsilon_n = e^{-n^{\delta\eta}} = e^{-1/r_n}$ in (11.2.4) can be very fast and hence helps the proof of (11.2.4).

On the other hand, the density condition is only needed for the proof of (11.2.4); that is, to handle the convergence of the smallest eigenvalues of $\mathbf{H}$. For removing the density assumption, we thank Rudelson and Vershynin [247], who proved the following theorem.

Theorem 11.23. *Let $\xi_1, \cdots, \xi_n$ be independent centered random variables with variances at least 1 and fourth moments at most B. Let $\mathbf{A}$ be an $n \times n$ matrix whose rows are independent copies of the random vector $(\xi_1, \cdots, \xi_n)$. Let $K \ge 1$. Then, for every $\varepsilon > 0$, one has*

$$\mathrm{P}(s_n(\mathbf{A}) \le \varepsilon n^{-1/2}) \le C\varepsilon + c^n + \mathrm{P}(\|\mathbf{A}\| > K n^{1/2}), \qquad (11.10.4)$$

where $s_n(\mathbf{A})$ denotes the smallest singular value and $C > 0$ and $c \in (0, 1)$ depend (polynomially) only on B and K.

To apply the estimate of the smallest singular values to the proof of the circular law, Pan and Zhou [227] extended Theorem 11.23 as follows.

Theorem 11.24. *Let* $\mathbf{W} = \mathbf{X} + \mathbf{A}_n$*, where* $\mathbf{A}_n$ *is an* $n \times n$ *non-random complex matrix and* $\mathbf{X} = (X_{jk})_{n\times n}$*, a random matrix.*

Assume $\{X_{jk}\}$ *are iid complex random variables with* $\mathrm{E}X_{11} = 0$*,* $\mathrm{E}|X_{11}^2| = 1$*, and* $\mathrm{E}|X_{11}^3| < B$*. Let* $K \geq 1$*. Then, for every* $\varepsilon > 0$*, that may depend on* n*,*

$$\mathrm{P}(s_n(\mathbf{W}) \leq \varepsilon n^{-1/2}) \leq C\varepsilon + c^n + \mathrm{P}(\|\mathbf{W}\| > Kn^{1/2}), \tag{11.10.5}$$

where $C > 0$ *and* $c \in (0, 1)$ *depend only on* K, B*,* $\mathrm{E}\Re(X_{11})^2$*,* $\mathrm{E}\Im(X_{11})^2$*, and* $\mathrm{E}\Re(X_{11})\Im(X_{11})$*.*

It should be noted that the extension to Rudelson and Vershynin's theorem is crucial because we can apply it to $\mathbf{X} - z\mathbf{I}$, whose entries are no longer iid or centered. Based on their extension, Pan and Zhou [227] proved the following circular law.

Theorem 11.25. *Suppose that* $\{X_{jk}\}$ *are iid complex random variables with* $\mathrm{E}X_{11} = 0$*,* $\mathrm{E}|X_{11}^2| = 1$*, and* $\mathrm{E}|X_{11}^4| < \infty$*. Then, with probability* 1*, the empirical spectral distribution* $\mu_n(x, y)$ *converges to the uniform distribution over the unit disk in two-dimensional space.*

Tao and Vu [273] further generalized Theorem 11.24 to reduce the moment requirement to the existence of the $2 + \eta$-th moment; i.e., $\mathrm{E}|X_{11}|^{2+\eta} < \infty$.

Because the proof involves the estimation of small ball probability, which may be beyond the knowledge of most graduates and junior researchers, we omit the proofs of these theorems. We refer readers who are interested in the detailed proofs to Tao and Vu [273].

Chapter 12
Some Applications of RMT

In recent decades, data sets have become large in both size and dimension, and thus statisticians are confronted with large dimensional data analysis in both theoretical investigation and real applications. Consequently, RMT has found applications to modern statistics and many applied disciplines. As an illustration, we briefly mention some basic concepts and applications in wireless communications and statistical finance.

12.1 Wireless Communications

In the past decade, RMT has found wide application in wireless communications. Random matrices are employed to describe the propagation of two important wireless communication systems: the multiple-input multiple-output (MIMO) antenna system and the direct-sequence code-division multiple-access (DS-CDMA) system. In an MIMO antenna system, multiple antennas are used at the transmitter side for simultaneous data transmission and at the receiver side for simultaneous reception. For a rich multipath environment, the channel responses between the transmit antennas and the receive antennas can be simply modeled as independent and identically distributed (iid) random variables. Thus the wireless channel for such a communication scenario can be described by a random matrix. DS-CDMA is a multiple- access scheme supporting multiple users communicating with a single base station using the same time and frequency resources but different spreading codes. CDMA is the key physical layer air interface in third-generation (3G) cellular mobile communications. In a frequency-flat, synchronous DS-CDMA uplink system with random spreading codes, the channel can also be described by a random matrix.

Foschini [113] and Telatar [274] may have been the first to introduce RMT into wireless communications. They have proven that, for a given power budget and a given bandwidth, the ergodic capacity of an MIMO Rayleigh fading

Z. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices, Second Edition*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0661-8_12,

channel increases with the minimum number of transmit antennas and receive antennas. Furthermore, it is this promising result that makes MIMO an attractive solution for achieving high-speed wireless connections over a limited bandwidth (for details, we refer the reader to the monograph [232]).

Further applications of RMT to wireless communications may be attributed to [135], [215], [281], [286], where it was derived that, by using spectral theory of large dimensional random matrices, for a frequency-flat synchronous DS-CDMA uplink with random spreading codes, the output signal-to-interference-plus-noise ratios (SINRs) using well-known linear receivers such as matched filter, decorrelator, and minimum mean-square-error (MMSE) receivers converge to deterministic values for large systems; i.e., when both spreading gain and number of users proportionally tend to infinity. These provide us a fundamental guideline in designing system parameters and predicting the system performance without requiring the exact knowledge of the specific spreading codes for each user.

Some important application results of RMT to wireless communications are listed below, among many others.

(i) limiting capacity and asymptotic capacity distribution for random MIMO channels [286], [284];
(ii) asymptotic SINR distribution analysis for random channels [282], [194];
(iii) limiting SINR analysis for linearly precoded systems, such as the multi-carrier CDMA using linear receivers [86];
(iv) limiting SINR analysis for random channels with interference cancellation receivers [151], [280], [195];
(v) asymptotic performance analysis for reduced-rank receivers [152], [226];
(vi) limiting SINR analysis for coded multiuser systems [70];
(vii) design of receivers, such as the reduced-rank minimum mean-square-error (MMSE) receiver [187]; and
(viii) the asymptotic normality study for multiple-access interference (MAI) [309] and linear receiver output [142].

For more details, the reader is referred to Tulino and Verdú in [284].

For recent applications of RMT to an emerging area in wireless communications, the “cognitive radio” networks, we refer the reader to [307], [308], [190], [214], and [148].

In the following subsections, our main objective is to show that there are indeed many wireless channels and problems that can be modeled using random matrices and to review some of the typical applications of RMT in wireless communications. Since there are new results published every year, we are not in a position to review all of the results. Instead, we concentrate on some of the representative examples and main results. Hopefully this part of the book can help readers from both mathematical and engineering backgrounds see the link between the two different societies, identify new problems to work on, and promote interdisciplinary collaborations.

12.1.1 Channel Models

1. Basics of wireless communication systems

The working principle of a wireless communication system is illustrated in Fig. 12.1, the block diagram of a wireless communication system, which consists of three basic parts: transmitter, channel, and receiver. The objective of the transmitter design is to transform the information bits into a signal format that is suitable for transmission over the wireless channels. The key components in the transmitter side include channel coding, modulation, and linear or nonlinear precoding. When the signal passes through the channel, the signal strength will be attenuated due to propagation loss, shadowing, and multipath fading, and the received signal waveform will be different from the transmitted signal waveform due to multipath delay, time/frequency selectivity of the channel, and the addition of noise and unwanted interference. Finally, at the receiver side, the transmitted information bits are to be recovered through the operations of equalization, demodulation, and channel decoding.

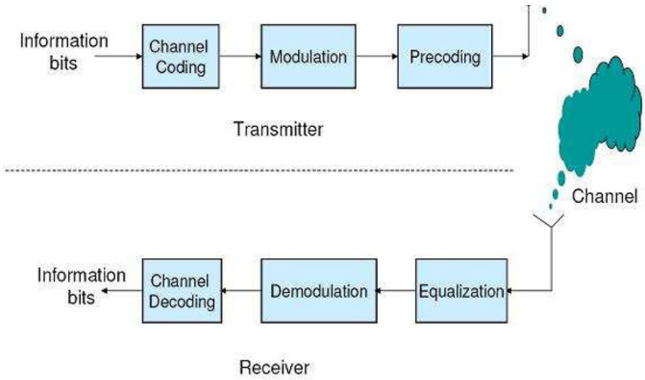

Fig. 12.1 Block diagram of wireless communication system.

2. Mathematical formulation of channels by matrices

In this and the next subsection, we formulate the input-output model arising from wireless communication systems,

$$\begin{aligned}\mathbf{x} &= \sum_{i=1}^{n} \mathbf{h}_i s_i + \mathbf{u} \\ &= \mathbf{Hs} + \mathbf{u},\end{aligned} \tag{12.1.1}$$

where

$$\mathbf{s} = [s_1, s_2, \cdots, s_n]'$$

represents the transmitted signal vector of dimension $n \times 1$; $\mathbf{h}_i$ represents the channel vector of dimension $p \times 1$, corresponding to symbol s_i;

$$\mathbf{x} = [x_1, x_2, \cdots, x_p]',$$
$$\mathbf{u} = [u_1, u_2, \cdots, u_p]',$$

denote the received signal vector and received noise vector, respectively, both with dimension $p \times 1$; and

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_n]$$

is the $p \times n$ channel matrix. In (12.1.1), n and p are referred to as the signal dimension and observation dimension, respectively. The matrix model in (12.1.1) can be derived either in the time, frequency, space, or code domain, or any combination of them. In the following subsections, we describe two popular matrix models in wireless communications: random matrix channels and linearly precoded channels.

12.1.2 random matrix channelRandom Matrix Channels

Random matrix channels include DS-CDMA uplink, MIMO antenna systems, and spatial division multiple access (SDMA) uplink.

1. DS-CDMA uplink

In a DS-CDMA system, all users within the same cell communicate with a single base station using the same time and frequency resources. The transmission from the users to the base station is called uplink, while the transmission from the base station to the users is referred to as downlink. The block diagram of the DS-CDMA uplink is illustrated in Fig. 12.2. In order to achieve user differentiation, each user is assigned a unique spreading sequence. The matrix model in (12.1.1) directly represents the frequency-flat synchronous DS-CDMA uplink, where s_i and $\mathbf{h}_i$ represent the transmitted symbol and spreading sequence of user i, respectively. In this case, n and p denote the number of active users and processing gain, respectively. In the third-generation (3G) wideband CDMA system, which is one of the 3G physical layer standards, the uplink spreading codes are designed as random codes, and thus the equivalent propagation channel from the users to the base station can be modeled as a random matrix channel.

2. MIMO antenna systems
Figure 12.3 shows the block diagram of a MIMO antenna system. The transmitter has n transmit antennas and the receiver has p receive antennas. The matrix model in (12.1.1) can also be used to represent such a system, where s_i and $\mathbf{h}_i$ denote respectively the transmitted symbol from the i-th transmit antenna and the channel responses from that transmit antenna to all receive antennas. While the MIMO channel modeling is related to the antenna configurations at both the transmitter and receiver sides, as well as to the multipath environment [232], when there are rich local scatters surrounding both sides, $\mathbf{h}_i$ can be simply modeled as an iid vector (i.e., a vector of iid entries), and thus the channel in (12.1.1) becomes a random matrix channel.

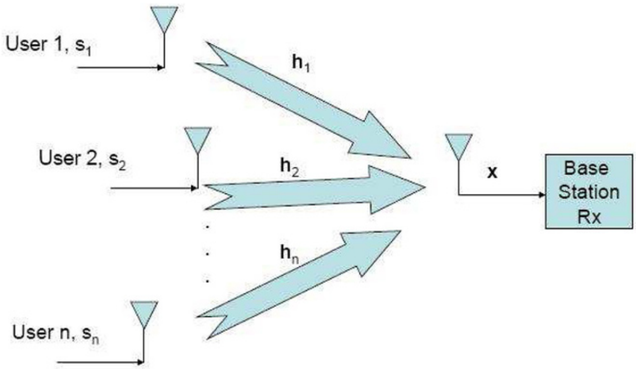

Fig. 12.2 Block diagram of DS-CDMA uplink.

3. SDMA uplink
In an SDMA system, the base station supports multiple users for simultaneous transmission using the same time and frequency resources [119, 238, 239]. This is achieved by equipping the base station with multiple antennas, and by doing so the spatial channels for different users are different, which allows the signals from different users to be distinguishable. Figure 12.4 shows the block diagram of the SDMA uplink, where n users communicate with the same base station equipped with p antennas. The matrix channel in (12.1.1) can be used to represent the uplink scenario, where s_i and $\mathbf{h}_i$ denote respectively the transmitted symbol from user i and the channel responses from this user to all receive antennas at the base station. When the base station antennas and the users are surrounded with rich scatters, the channel matrix can be modeled as an iid matrix (that is, a matrix of iid entries). Note that SDMA and CDMA can be further combined to generate SDMA-CDMA systems [192, 193].

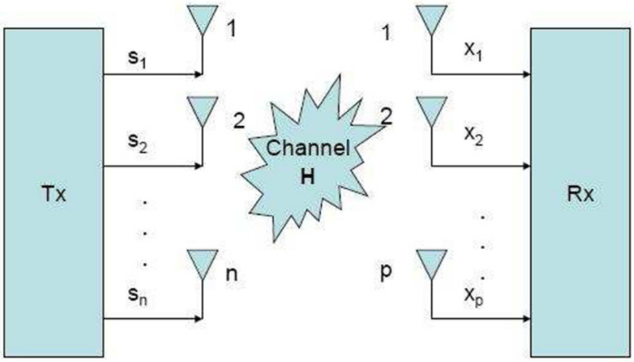

Fig. 12.3 Block diagram of MIMO antenna system.

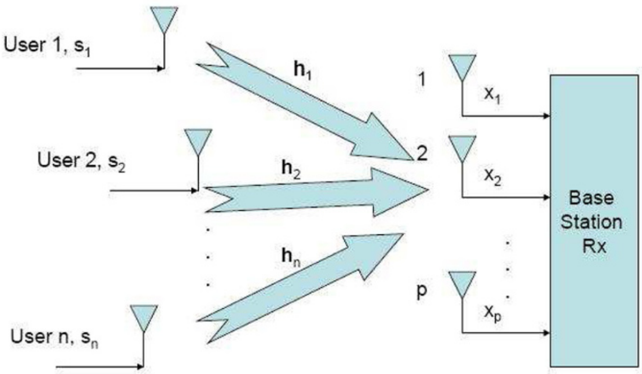

Fig. 12.4 Block diagram of SDMA uplink.

12.1.3 Linearly Precoded Systems

In broadband communications, the wireless channels usually have memories and thus are frequency selective. The frequency selectiveness of the channel introduces intersymbol interferences (ISI) at the receiver side. Linear or nonlinear equalizers are designed to suppress the ISI. To simplify the complexity of the equalizers, linear precoding at the transmitter side can be applied. The matrix channel in (12.1.1) can be used to represent wireless channels of cyclic-prefix (CP)-based block transmissions, which include, for example, orthogonal frequency division multiplexing (OFDM), single-carrier CP (SCCP) systems, multicarrier CDMA (MC-CDMA), and CP-CDMA.

1. CP-based block transmissions

The block diagram of a CP-based block transmission system is illustrated in Fig. 12.5. Consider the case where a CP portion of M symbols is inserted prior to the transmission of each data block of p symbols. Suppose the frequency-selective channel can be represented by $(L+1)$ equally spaced time domain taps, $h_0, h_1, \cdots, h_L$. Here, L is also referred to as the channel memory. The insertion of CP alleviates the interblock interference if the CP length M is larger than the channel memory and, more importantly, it transforms the linear convolution operation into a circular convolution operation.

Let $\mathbf{y}$ be the signal block before CP insertion at the transmitter and $\mathbf{z}$ the received signal block after CP removal at the receiver, with circular convolution. The relation between $\mathbf{y}$ and $\mathbf{z}$ is given by

$$\mathbf{z} = \mathbf{W}_p^* \mathbf{\Lambda}_p \mathbf{W}_p \mathbf{y} + \tilde{\mathbf{u}}, \tag{12.1.2}$$

where $\mathbf{W}_p \in \mathbb{C}^{p\times p}$ is the $p \times p$ discrete Fourier transform matrix,

$$\mathbf{W} = \frac{1}{\sqrt{p}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{\frac{-j2\pi}{p}} & \cdots & e^{\frac{-j2\pi\times(p-1)}{p}} \\ \cdots & & & \\ 1 & e^{\frac{-j2\pi\times(p-1)}{p}} & \cdots & e^{\frac{-j2\pi\times(p-1)(p-1)}{p}} \end{bmatrix}.$$

$\mathbf{\Lambda}_p = \text{diag}\{[f_0, \cdots, f_{p-1}]\}$ is the $p \times p$ diagonal matrix with $f_k = \sum_{l=0}^{L} h_l e^{-j\frac{2\pi kl}{p}}$, and $\tilde{\mathbf{u}}$ is the received noise vector.

The data block $\mathbf{y}$ is the linear transform of the modulated block $\mathbf{s}$ with size $n \times 1$ and is described as $\mathbf{y} = \mathbf{W}_p^* \mathbf{Q}_p \mathbf{s}$, where $\mathbf{Q}_p \in \mathbb{C}^{p\times n}$ is the first precoding matrix. Performing discrete Fourier transformation (DFT) on the CP-removed block $\mathbf{z}$, we then have the input-output relation

$$\mathbf{x} = \mathbf{\Lambda}_p \mathbf{Q}_p \mathbf{s} + \mathbf{u}, \tag{12.1.3}$$

where $\mathbf{x} = \mathbf{W}_p \mathbf{z}$ and $\mathbf{u} = \mathbf{W}_p \tilde{\mathbf{u}}$.

In (12.1.3), if we choose $\mathbf{Q}_p^* \mathbf{Q}_p = \mathbf{I}_n$, the system is referred to as an *isometrically precoded system*. If $\mathbf{Q}_p$ is chosen as an iid matrix, then the system is called a *randomly precoded system*. Finally, if $\mathbf{\Lambda}_p = \mathbf{I}_p$ and $\mathbf{Q}_p$ is an iid matrix, the system is equivalent to a *random MIMO system* [281, 286].

2. Orthogonal frequency division multiplexing (OFDM) systems

In OFDM systems, we choose $n = p$ and $\mathbf{Q}_p = \mathbf{I}_p$. Thus, we have

$$\mathbf{y} = \mathbf{W}_p^* \mathbf{s}, \tag{12.1.4}$$

$$\mathbf{x} = \mathbf{\Lambda}_p \mathbf{s} + \mathbf{u}. \tag{12.1.5}$$

In (12.1.5), since $\mathbf{\Lambda}_p$ is a diagonal matrix, the received signals have been completely decoupled; i.e.,

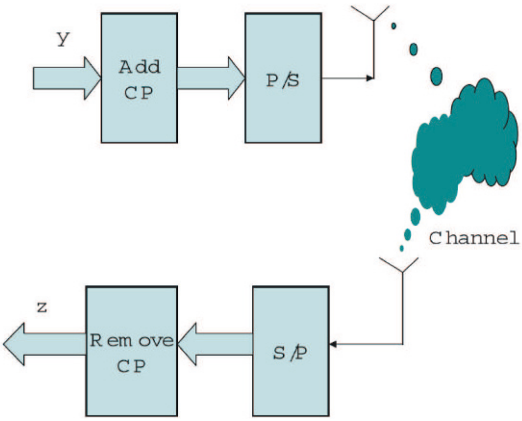

Fig. 12.5 Block diagram of CP-based block transmission system.

$$x_i = f_i s_i + u_i$$

for $i = 0, \cdots, p-1$, where x_i and u_i are the i-th elements of $\mathbf{x}$ and $\mathbf{u}$, respectively. Thus the signal detection problem for recovering the transmitted signals becomes very simple to implement. Therefore, OFDM has become the most popular scheme to handle the ISI issue and has been adopted in various wireless standards; e.g., IEEE802.11 wireless local area networks, IEEE802.16 wireless metropolitan area networks, and third-generation long-term evolution (3G-LTE).

3. Single-Carrier CP (SCCP) systems
In an OFDM system, from (12.1.4), $\mathbf{y}$ is the IDFT (inverse DFT) output of the modulated symbols $\mathbf{s}$, and thus the signals transmitted may suffer from a high peak-to-average power ratio (PAPR), which causes difficulty in practical implementation. In an SCCP system, we directly choose $n = p$ and $\mathbf{y} = \mathbf{s}$, and thus we have

$$\mathbf{x} = \mathbf{\Lambda}_p \mathbf{W}_p \mathbf{s} + \mathbf{u}.$$

Obviously, SCCP is an isometrically precoded system. From (12.1.6), it can be seen that the transmitted symbols are mixed up together and thus the equalization for SCCP is more complicated compared with that for an OFDM system. However, due to its simplicity on the transmitter side and lower PAPR, SCCP has been adopted in IEEE802.16, and its multiuser version, interleaved frequency division multiple access (IFDMA), has been adopted in 3G-LTE uplink.

4. MC-CDMA

A single-user scenario is considered in OFDM and SCCP systems. In order to support multiple users simultaneously, in the next two subsections we will introduce downlink models for MC-CDMA and CP-CDMA. To do so, we use the following common notations: G for processing gain common to all users; T for the number of active users; $\mathbf{D}(q)$ for the long scrambling codes used at the q-th block, where

$$\mathbf{D}(q) = \text{diag}\left\{[d(q;0), \cdots, d(q;p-1)]\right\},$$

with $|d(q;k)| = 1$; and $\mathbf{c}_i$ for the short codes of user i, where

$$\mathbf{c}_i = [c_i(0), \cdots, c_i(G-1)]'$$

with $\mathbf{c}_i'\mathbf{c}_j = 1$ for $i = j$ and $\mathbf{c}_i'\mathbf{c}_j = 0$ for $i \neq j$. From now on, we look at the channel model from the base station to one particular mobile user.

MC-CDMA performs frequency domain spreading by transmitting the chip signals associated with each modulated symbol over different subcarriers within the same time block [147]. Denote Q as the number of symbols transmitted in one block for each user, G as the processing gain, T as the number of users, and $p = QG$ as the total number of subcarriers. There are then $n = TQ$ multiuser symbols in each block.

At the receiver side, the q-th received block after CP removal and FFT operation can be represented as

$$\mathbf{x}(q) = \mathbf{\Lambda}_p \mathbf{D}(q)\mathbf{C}\mathbf{s}(q) + \mathbf{u}(q), \tag{12.1.6}$$

where

$$\begin{aligned} \mathbf{x}(q) &= [x(q;0), \cdots, x(q;p-1)]', \\ \mathbf{s}(q) &= \left[\bar{\mathbf{s}}_1'(q), \cdots, \bar{\mathbf{s}}_Q'(q)\right]', \\ \mathbf{u}(q) &= [u(q;0), \cdots, u(q;p-1)]', \\ \mathbf{C} &= \text{diag}\left\{\bar{\mathbf{C}}, \cdots, \bar{\mathbf{C}}\right\}, \end{aligned}$$

with $\bar{\mathbf{s}}_i(q) = [s_0(q;i), \cdots, s_{T-1}(q;i)]'$ and $\bar{\mathbf{C}} = [\mathbf{c}_0, \cdots, \mathbf{c}_{T-1}]$.

5. CP-CDMA

CP-CDMA is the single-carrier dual of MC-CDMA. The $Q = p/G$ symbols of each user are first spread out with user-specific spreading codes and then the chip sequence for all users is summed up; the total chip signal of size p is then passed to a CP inserter, which adds a CP. Using the duality between CP-CDMA and MC-CDMA, from (12.1.6), the q-th received block of CP-CDMA after FFT can be written as

$$\mathbf{x}(q) = \mathbf{\Lambda}_p \mathbf{W}_p \mathbf{D}(q)\mathbf{C}\mathbf{s}(q) + \mathbf{u}(q), \tag{12.1.7}$$

where $\mathbf{x}(q)$, $\mathbf{C}$, $\mathbf{s}(q)$, and $\mathbf{u}(q)$ are defined the same way as in the MC-CDMA case. Again, there are $P = TQ$ multiuser symbols in each block.

For MC-CDMA downlink, $\mathbf{Q}_p = \mathbf{D}(q)\mathbf{C}$, and for CP-CDMA downlink, $\mathbf{Q}_p = \mathbf{W}_p\mathbf{D}(q)\mathbf{C}$. Since $\mathbf{Q}_p^*\mathbf{Q}_p = \mathbf{I}_n$, both systems belong to the isometrically precoded category.

12.1.4 Channel Capacity for MIMO Antenna Systems

Channel capacity is a fundamental performance indicator used in communication theory study; it describes the maximum rate of data transmission that the channel can support with an arbitrarily small probability of error incurred due to the channel impairment. The channel capacity for additive white Gaussian noise channels was derived by Claude Shannon in 1948 [84]. For single-input single-output systems, the capacity limits for fading channels have been well documented for example in [130, 55, 69, 54]. In this section, we consider the channel capacity of MIMO antenna systems in the fading channel environment.

1. Single-input single-output channels

Let us first consider the AWGN channel

$$x(q) = s(q) + u(q) \tag{12.1.8}$$

and assume that (i) the transmitted signal $s(q)$ is zero-mean iid Gaussian with $\mathrm{E}[|s(q)|^2] = \sigma_s^2$ and (ii) the noise $u(q)$ is zero-mean iid Gaussian with $\mathrm{E}[|u(q)|^2] = \sigma_u^2$, and denote $\Gamma = \sigma_s^2/\sigma_u^2$ as the signal-to-noise ratio (SNR) of the channel.

The capacity of the channel is determined by the mutual information between the input and output, which is given by

$$C = \log_2(1 + \Gamma).$$

Here the unit of capacity is bits per second per Hertz (bits/sec/Hz). In the high-SNR regime, the channel capacity increases by 1 bit/sec/Hz for every 3 dB increase in SNR. Note that the channel capacity determines the maximum rate of codes that can be transmitted over the channel and recovered with arbitrarily small error.

Next, we consider the SISO block fading channel

$$x(q) = hs(q) + u(q) \tag{12.1.9}$$

and assume that (i) the transmitted signal $s(q)$ is zero-mean iid Gaussian with $\mathrm{E}[|s(q)|^2] = \sigma_s^2$; (ii) the noise $u(q)$ is zero-mean iid Gaussian with $\mathrm{E}[|u(q)|^2] =$

σ_u^2, and denote $\varGamma = E_s/\sigma_u^2$; and (iii) the fading state h is a random variable with $\mathrm{E}[|h|^2] = 1$.

Let us first introduce the concept of a *block fading channel.* A block fading channel refers to a slow fading channel whose coefficient is constant over an interval of large time T and changes to another independent value, which is again constant over an interval of time T, and so on. The *instantaneous* mutual information between $s(q)$ and $x(q)$ of channel (12.1.9) conditional on channel state h is given by

$$I(s;x|h) = \log_2(1 + |h|^2\varGamma).$$

Since h is a random variable, the instantaneous mutual information is also a random variable. Thus, if the distribution of $|h|^2$ is known, the distribution of $I(s;x|h)$ can be calculated accordingly.

The channel capacity of a fading channel can be quantified either in an ergodic sense or in an outage sense, yielding *ergodic capacity* and *outage capacity.*

The ergodic capacity of the SISO fading channel (12.1.9) is defined as

$$C = \mathrm{E}[\log_2(1 + |h|^2\varGamma)],$$

where the expectation is taken over the channel state variable h. Physically speaking, the ergodic capacity defines the maximum (constant) rate of codes that can be transmitted over the channel and recovered with arbitrarily small probability of error when the codes are long enough to cover all the possible channel states.

In Fig. 12.6, we compare the capacities of the AWGN channel and the SISO Rayleigh fading channel with respect to the received SNR. Here, for the fading channel case, we have used the average received SNR. It can be seen that, at high SNR, the capacity of the fading channel increases by 1 bit/sec/Hz for every 3 dB increase in SNR, which is the same as for the AWGN channel.

Since the instantaneous mutual information is a random variable, if a code with constant rate C_0 is transmitted over the fading channel, this code cannot be correctly recovered at the receiver at a fading block whose instantaneous mutual information is lower than the code rate C_0, thus causing an outage event. We define the *outage probability* as the probability that the instantaneous mutual information is less than the rate of C_0; i.e.,

$$P_{\mathrm{out}}(C_0) = \Pr(I(s;x|h) < C_0).$$

Based on this, the $\alpha\%$ *outage capacity* $C_{out,\alpha\%}$ is defined as the maximum information rate of codes transmitted over the fading channel for which the outage probability does not exceed $\alpha\%$.

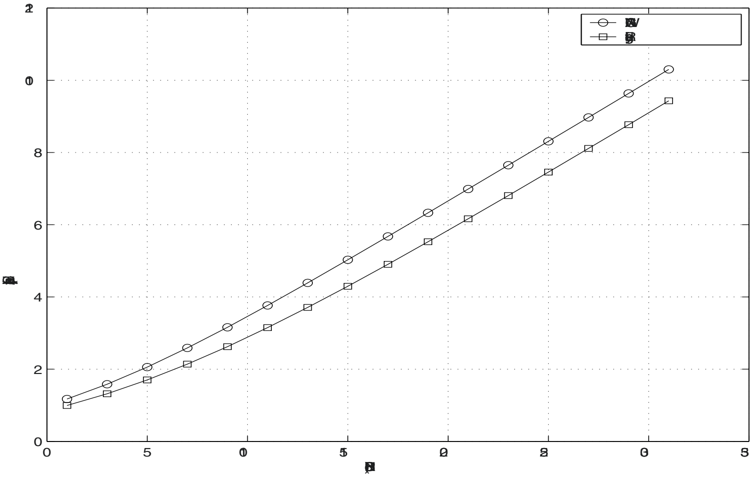

Fig. 12.6 The capacity comparison for AWGN channels and SISO Rayleigh fading channels.

2. MIMO fading channels

In order to analyze the capacity of MIMO fading channels, we make the following assumptions:

(A1) The channel vectors $\mathbf{h}_i$ can be represented as

$$\mathbf{h}_i = [X_{1i}, X_{2i}, \cdots, X_{pi}]' \tag{12.1.10}$$

for $i = 1, \cdots, n$, where X_{ki}'s are iid random variables with zero mean and unit variance; i.e., $\mathrm{E}[|X_{ki}|^2] = 1$ for all k's and i's.

(A2) The n symbols constituting the transmitted signal vector are drawn from a *Gaussian codebook* with $\mathbf{R}_x = \mathrm{E}[\mathbf{x}(q)\mathbf{x}^*(q)]$ and $\mathrm{tr}(\mathbf{R}_x) = \sigma_s^2$. This is the total power constraint.

(A3) The elements of $\mathbf{u}(q)$ are zero-mean, circularly symmetric complex Gaussian with $\mathbf{R}_u = \mathrm{E}[\mathbf{u}(q)\mathbf{u}^*(q)] = \sigma_u^2\mathbf{I}$.

The following two cases need to be considered separately when we study the MIMO channel capacity. In the first case, the channel state information (CSI) $\mathbf{H}$ is available at the transmitter side. This case is called the *CSI-known case*. In the second case, the CSI is unavailable at the transmitter side. This case is referred to as the *CSI-unknown case*. In both cases, we assume that the CSI is perfectly known at the receiver side.

Let us first look at the instantaneous mutual information under one CSI realization $\mathbf{H}$. Based on the distribution of $\mathbf{H}$, the distribution of the instantaneous mutual information can be derived. Similar to the SISO fading channel case, we study the ergodic capacity and outage capacity of MIMO fading channels.

Using singular-value decomposition (SVD), the $p \times n$ channel matrix $\mathbf{H}$ can be represented as

$$\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*,$$

where matrix $\mathbf{U}$ is a $p \times p$ unitary matrix ($\mathbf{U}^*\mathbf{U} = \mathbf{I}_p$) and is called the left singular vector matrix of $\mathbf{H}$; matrix $\mathbf{V}$ is an $n \times n$ unitary matrix ($\mathbf{V}^*\mathbf{V} = \mathbf{I}_n$), which is referred to as the right singular vector matrix of $\mathbf{H}$; and $\boldsymbol{\Sigma}$ is the $p \times n$ singular value matrix, the elements of which are zeros except that $(\boldsymbol{\Sigma})_{i,i} = \sigma_i \geq 0$, where $\sigma_1 \geq \ldots \geq \sigma_M \geq 0$ ($M = \min(n, p)$) are the singular values of $\mathbf{H}$. Note that $\mathbf{H}\mathbf{H}^* = \mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^*\mathbf{U}^*$, and thus $\lambda_i = \sigma_i^2$, $\forall\ i$ are the nonzero eigenvalues of $\mathbf{H}\mathbf{H}^*$.

(i). MIMO fading channels for CSI-known case

When CSI is available at the transmitter, joint transmit and receive beamforming can be used to decouple the MIMO fading channel into M SISO fading channels. Figure 12.7 shows the block diagram of the decoupling process.

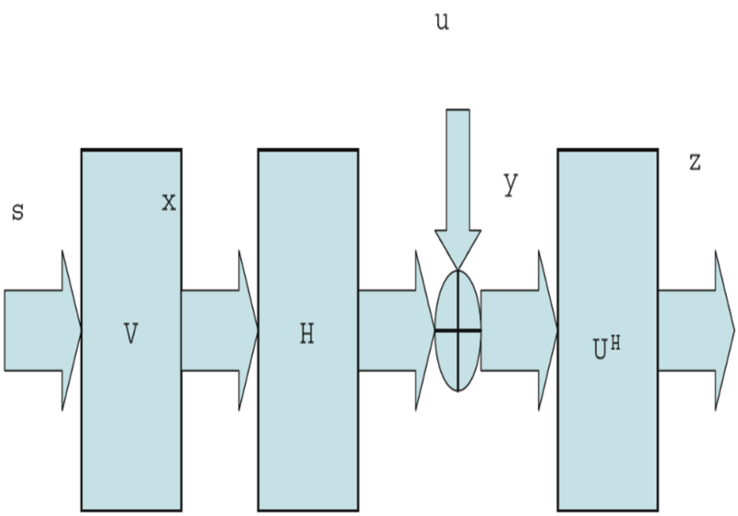

Fig. 12.7 Block diagram of joint transmit and receive eigen-beamforming.

At the transmitter side, we precode the transmitted signals using transmit beamforming:

$$\mathbf{x} = \mathbf{V}\mathbf{s}.$$

Then the received signal vector is given by

$$\begin{aligned}\mathbf{y} &= \mathbf{U\Sigma V^* V s} + \mathbf{u} \\ &= \mathbf{U\Sigma s} + \mathbf{u}.\end{aligned}$$

At the receiver side, if we premultiply $\mathbf{y}$ with $\mathbf{U}^*$ (this processing is called *receive eigenbeamforming*), we then have

$$\begin{aligned}\mathbf{z} &= \mathbf{\Sigma s} + \mathbf{U}^*\mathbf{u} \\ &= \mathbf{\Sigma s} + \tilde{\mathbf{u}}.\end{aligned}$$

Note that $\tilde{\mathbf{u}} = \mathbf{U}^*\mathbf{u}$, and thus $\mathbf{R}_{\tilde{u}} = \mathrm{E}[\tilde{\mathbf{u}}\tilde{\mathbf{u}}^*] = \mathrm{E}[\mathbf{U}^*\mathbf{u}\mathbf{u}^*\mathbf{U}] = \sigma_u^2\mathbf{I}$. Therefore, we have

$$z_i = \sigma_i s_i + \tilde{u}_i, \quad i = 1, \ldots, M.$$

From the above, it can be seen that with CSI known, the MIMO channel has been decoupled into M SISO channels through joint transmit and receive beamforming. That is to say, M data streams can be transmitted in a parallel manner.

Suppose the transmission power to the i-th data stream is $\gamma_i = \mathrm{E}[|s_i|^2]$. Then the SNR for this data stream is given by

$$SNR_i = \frac{\sigma_i^2 \mathrm{E}[|s_i|^2]}{\mathrm{E}[|\tilde{u}_i|^2]} = \frac{\sigma_i^2\gamma_i}{\sigma_u^2}.$$

Note that the total power over the M data streams has to be less than or equal to σ_s^2; i.e., $\sum_{i=1}^M \gamma_i \le \sigma_s^2$.

The capacity of the MIMO channel under channel state $\mathbf{H}$ is equal to the sum of the individual SISO channel's capacity

$$I_{\mathbf{H}} = \sum_{i=1}^M \log_2\left(1 + \frac{\sigma_i^2\gamma_i}{\sigma_u^2}\right) = \sum_{i=1}^M \log_2\left(1 + \frac{\lambda_i\gamma_i}{\sigma_u^2}\right)$$

under the power constraint $\sum_{i=1}^M \gamma_i \le \sigma_s^2$ (see Fig 12.8).

If equal power is allocated to each data stream (i.e., $\gamma_i = \frac{\sigma_s^2}{M}$), then

$$I_{\mathbf{H}}^{(\mathrm{EP})} = \sum_{i=1}^M \log_2\left(1 + \frac{\lambda_i\sigma_s^2}{M\sigma_u^2}\right). \tag{12.1.11}$$

In order to achieve maximum capacity, we can allocate different powers to the data streams using the Lagrangian method. Define the objective function

$$J(\gamma_1, \cdots, \gamma_M) = \sum_{i=1}^M \log_2\left(1 + \frac{\lambda_i\gamma_i}{\sigma_u^2}\right) - \mu_1\left(\sum_{i=1}^M \gamma_i - \sigma_s^2\right).$$

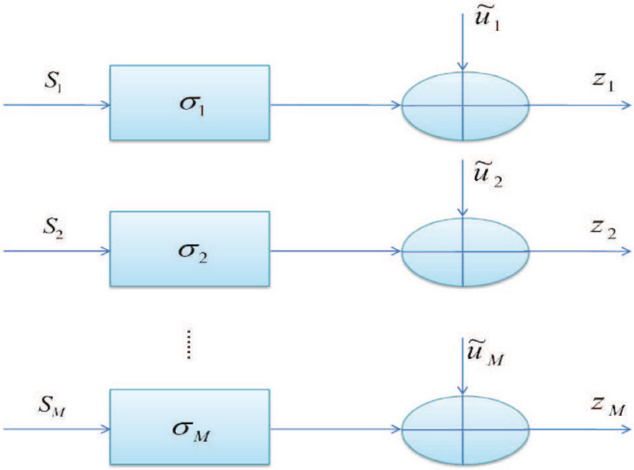

Fig. 12.8 A MIMO channel is equivalent to a set of parallel SISO channels.

Calculating $\frac{\partial J(\gamma_1,\cdots,\gamma_M)}{\partial \gamma_i}$ and setting $\frac{\partial J(\gamma_1,\cdots,\gamma_M)}{\partial \gamma_i} = 0$ for all i's, we obtain the water-filling solution

$$\gamma_i^* = \left(\mu - \frac{\sigma_u^2}{\lambda_i}\right)^+, \quad i = 1, \cdots, M,$$

where μ is the water level for which the equality power constraint is satisfied, $(x)^+ = x$ for $x \geq 0$, and $(x)^+ = 0$ for $x < 0$. With the water-filling power allocation, the channel capacity can then be calculated as

$$I_{\mathbf{H}}^{(\mathrm{WF})} = \sum_{i=1}^{M} \log_2 \left(1 + \frac{\lambda_i \gamma_i^*}{\sigma_u^2}\right).$$

For block fading channels, both $I_{\mathbf{H}}^{(\mathrm{EP})}$ and $I_{\mathbf{H}}^{(\mathrm{WF})}$ are random variables. If the distribution of the eigenvalues is known, we can then calculate the distributions of $I_{\mathbf{H}}^{(\mathrm{EP})}$ and $I_{\mathbf{H}}^{(\mathrm{WF})}$.

Taking expectations on $I_{\mathbf{H}}^{(\mathrm{EP})}$ and $I_{\mathbf{H}}^{(\mathrm{WF})}$ over the random matrix $\mathbf{H}$, we obtain the ergodic capacities of the MIMO fading channels as

$$C^{(\mathrm{EP})} = \mathrm{E}_{\mathbf{H}}[I_{\mathbf{H}}^{(\mathrm{EP})}] = \mathrm{E}_{\mathbf{\Lambda}}\left[\sum_{i=1}^{M} \log_2 \left(1 + \frac{\lambda_i \sigma_s^2}{M\sigma_u^2}\right)\right],$$

$$C^{(\mathrm{WF})} = \mathrm{E}_{\mathbf{H}}[I_{\mathbf{H}}^{(\mathrm{WF})}] = \mathrm{E}_{\mathbf{\Lambda}}\left[\sum_{i=1}^{M} \log_2 \left(1 + \frac{\lambda_i \gamma_i^*}{\sigma_u^2}\right)\right].$$

Since $I_{\mathbf{H}}^{(\mathrm{WF})} \geq I_{\mathbf{H}}^{(\mathrm{EP})}$, thus $C^{(\mathrm{WF})} \geq C^{(\mathrm{EP})}$. In the high SNR regime, however, these two capacities tend to be equal.

The expressions of the ergodic capacities can be derived based on the eigenvalue distributions [196]. Alternatively, we may quantify the performance gain of using multiple antennas by looking at the lower bound of the ergodic capacity. In fact, for the CSI-known case with equal power allocation, a lower bound of the ergodic capacity of MIMO Rayleigh fading channels is given by ([224])

$$C = C(\Gamma) \geq M \log_2 \left[1 + \frac{\Gamma}{M} \exp \left(\frac{1}{M} \sum_{j=1}^{M} \sum_{k=1}^{Q-j} \frac{1}{k} - \gamma \right) \right], \quad (12.1.12)$$

where $\gamma \approx 0.57721566$ is Euler's constant, $\Gamma = \frac{\sigma_s^2}{\sigma^2}$, and $Q = \max(n, p)$.

Let us define spatial multiplexing gain as $r = \lim_{\Gamma \to \infty} \frac{C(\Gamma)}{\log_2(\Gamma)}$. From (12.1.12), we can see that $r = \lim_{\Gamma \to \infty} \frac{C(\Gamma)}{\log_2(\Gamma)} = M$. Therefore, in the high SNR regime, the ergodic capacity increases by M bits/sec/Hz for every 3 dB increase in the average SNR, Γ. Recall that for SISO AWGN channels and SISO Rayleigh fading channels, the capacity increases by 1 bit/sec/Hz for every 3 dB increase of SNR in the high SNR regime. This shows the tremendous capacity gain by using MIMO antenna systems.

(ii). MIMO fading channels for CSI-unknown case

In Part 3, we have proved that, when the CSI is known at the transmitter side, the spatial multiplexing gain for a $p \times n$ MIMO channel is equal to M, which is the minimum number of transmit and receive antennas. In practice, the CSI may not be available at the transmitter, so can we still achieve a spatial multiplexing gain of M? We deal with this question in this part.

For MIMO flat fading channels, when the input signals are iid Gaussian, the instantaneous mutual information between $\mathbf{x}$ and $\mathbf{y}$ under channel state $\mathbf{H}$ is given by ([113], [274])

$$I_{\mathbf{H}} = \log_2 \{\det(\pi e \mathbf{R}_y)\} - \log_2 \{\det(\pi e \mathbf{R}_u)\}.$$

Since $\mathbf{R}_y = \mathrm{E}[\mathbf{y}\mathbf{y}^*] = \mathbf{R}_u + \mathbf{H}\mathbf{R}_x\mathbf{H}^*$ and $\mathbf{R}_u = \sigma_u^2 \mathbf{I}_p$, we have

$$I_{\mathbf{H}} = \log_2 \left\{ \det \left(\mathbf{I}_p + \frac{\mathbf{H}\mathbf{R}_x\mathbf{H}^*}{\sigma_u^2} \right) \right\}.$$

For the CSI-unknown case, we will choose the transmitted signal vector $\mathbf{x}$ such that $\mathbf{R}_x = \frac{\sigma_s^2}{n}\mathbf{I}_n$. Thus, if $n \geq p$,[1]

[1] Suppose matrices $\mathbf{A}$ and $\mathbf{B}$ are of dimensions $m \times q$ and $q \times m$, respectively. Then we have the equality $\det(\mathbf{I}_m + \mathbf{AB}) = \det(\mathbf{I}_q + \mathbf{BA})$.

$$\begin{aligned} I_{\mathbf{H}} &= \log_2 \left\{ \det \left(\mathbf{I}_p + \frac{\sigma_s^2}{K\sigma_u^2} \mathbf{H}\mathbf{H}^* \right) \right\} \\ &= \log_2 \left\{ \det \left(\mathbf{I}_p + \frac{\sigma_s^2}{K\sigma_u^2} \mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^* \mathbf{U}^* \right) \right\} \\ &= \log_2 \left\{ \det \left(\mathbf{I}_p + \frac{\sigma_s^2}{K\sigma_u^2} \boldsymbol{\Sigma}\boldsymbol{\Sigma}^* \mathbf{U}^* \mathbf{U} \right) \right\} \\ &= \sum_{i=1}^{M} \log_2 \left(1 + \frac{\sigma_s^2 \sigma_i^2}{n\sigma_u^2} \right) = \sum_{i=1}^{M} \log_2 \left(1 + \frac{\sigma_s^2 \lambda_i}{K\sigma_u^2} \right) = I_{\boldsymbol{\Lambda}}, \end{aligned} \tag{12.1.13}$$

where again σ_i is the i-th singular value of $\mathbf{H}$ and λ_i is the i-th eigenvalue of $\mathbf{H}\mathbf{H}^*$. On the other hand, if $n < p$,

$$\begin{aligned} I_{\mathbf{H}} &= \log_2 \left\{ \det \left(\mathbf{I}_p + \frac{\sigma_s^2}{n\sigma_u^2} \mathbf{H}\mathbf{H}^* \right) \right\} \\ &= \log_2 \left\{ \det \left(\mathbf{I}_n + \frac{\sigma_s^2}{n\sigma_u^2} \mathbf{H}^*\mathbf{H} \right) \right\} \\ &= \log_2 \left\{ \det \left(\mathbf{I}_n + \frac{\sigma_s^2}{n\sigma_u^2} \mathbf{V}\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}\mathbf{V}^* \right) \right\} \\ &= \log_2 \left\{ \det \left(\mathbf{I}_n + \frac{\sigma_s^2}{n\sigma_u^2} \boldsymbol{\Sigma}^*\boldsymbol{\Sigma}\mathbf{V}^*\mathbf{V} \right) \right\} \\ &= \sum_{i=1}^{M} \log_2 \left(1 + \frac{\sigma_s^2 \lambda_i}{n\sigma_u^2} \right) = I_{\boldsymbol{\Lambda}}. \end{aligned} \tag{12.1.14}$$

Combining (12.1.13) with (12.1.14) yields

$$I_{\mathbf{H}} = \sum_{i=1}^{M} \log_2 \left(1 + \frac{\sigma_s^2 \lambda_i}{K\sigma_u^2} \right).$$

When $n \le p$, the formula above is the same as (12.1.11). That is to say, if the number of transmit antennas is not greater than the number of receive antennas, even when the CSI is unknown at the transmitter, the capacity of the MIMO channel is the same as that for the CSI-known case when equal power allocation is applied.

Taking the expectation on $I_{\mathbf{H}}$ over the random matrix $\mathbf{H}$, we obtain the ergodic capacity of the MIMO fading channel as

$$C = \mathrm{E}_{\mathbf{H}}[I_{\mathbf{H}}] = \mathrm{E}_{\boldsymbol{\Lambda}} \left[\sum_{i=1}^{M} \log_2 \left(1 + \frac{\sigma_s^2 \lambda_i}{n\sigma_u^2} \right) \right].$$

For MIMO Rayleigh fading channels, similar to the CSI-known case, we have the inequality

$$C \geq M \log_2 \left[1 + \frac{\sigma_s^2}{n\sigma_u^2} \exp\left(\frac{1}{M} \sum_{j=1}^{M} \sum_{k=1}^{Q-j} \frac{1}{k} - \gamma \right) \right].$$

Thus a spatial multiplexing gain of M can be achieved for MIMO Rayleigh fading channels even when the CSI is unavailable at the transmitter side.

Figure 12.9 illustrates the ergodic capacities of MIMO fading channels with different antenna configurations for the CSI-unknown case. Here we also plot the lower bound of the ergodic capacity for 4×4 MIMO channels. It is seen that this lower bound is tight when the SNR is greater than 25 dB. Further, for asymmetric antenna configurations, for a given M and at high SNR regime, there exists a fixed SNR loss when the transmit antenna number is larger than the receive antenna number.

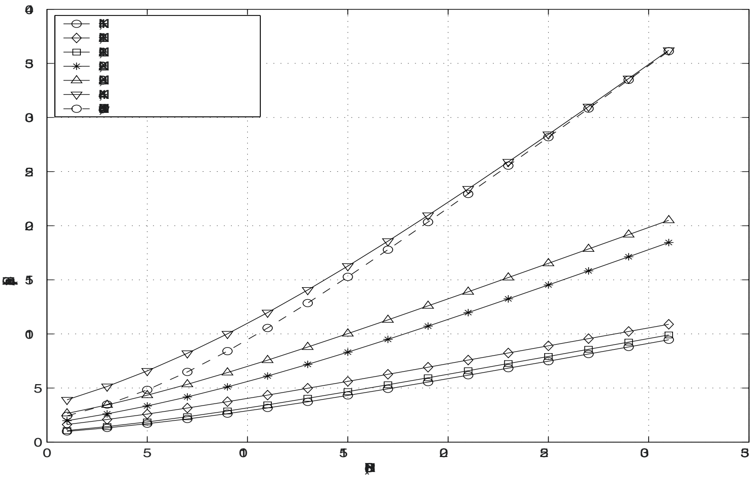

Fig. 12.9 The ergodic capacity comparison for MIMO Rayleigh fading channels with different numbers of antennas.

12.1.5 Limiting Capacity of Random MIMO Channels

In the previous section, we studied the ergodic capacity of MIMO fading channels with limited dimensions of n and p and showed that for a random matrix channel with Rayleigh fading coefficients, the MIMO channel achieves a spatial multiplexing gain of M, which is the minimum of the transmit antenna number and receive antenna number.

In this section, we are interested in the limiting performance of the instantaneous mutual information for any given random channel realization when

$n \to \infty$, $p \to \infty$ with $\frac{p}{n} \to y$ (constant). For the sake of brevity, we assume that n and p scale up with the same speed; i.e., $\frac{n}{p} = 1/y$ for every p.

According to the spectral theory of large random matrices, under assumption (A1) in Part 2 of Subsection 12.1.4 and for the limiting case, the empirical distribution of the eigenvalues of $\frac{1}{p}\mathbf{HH}^*$ converges almost surely to the M-P law (see Chapter 3) whose density function is given by

$$f_y(x) = (1 - 1/y)^+\delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi xy},$$

where

$$a = (1 - \sqrt{y})^2, \quad b = (1 + \sqrt{y})^2.$$

1. CSI-unknown case

Now, let us look at the limiting capacity of MIMO fading channels with CSI unavailable at the transmitter side. We are interested in the normalized mutual information

$$\begin{aligned}\tilde{I}_{\mathbf{H}} &= \frac{1}{p}\log_2\left\{\det\left(\mathbf{I}_p + \frac{\sigma_s^2}{n\sigma^2}\mathbf{HH}^*\right)\right\}\\ &= \frac{1}{p}\sum_{i=1}^{p}\log_2\left(1 + \frac{\sigma_s^2\lambda_i}{n\sigma_u^2}\right).\end{aligned}$$

In the previous subsection, we defined λ_i's as the eigenvalues of $\mathbf{HH}^*$. In the limiting case, the ESD of $\tilde{\lambda}_i = \frac{\lambda_i}{n}$, $i = 1, \cdots, p$, converges to $yf_y(x)$ almost surely for $x \in (a, b)$. Thus we have ([286])

$$\begin{aligned}\tilde{I}_{\mathbf{H}} &= \frac{1}{p}\sum_{i=1}^{p}\log_2\left(1 + \frac{\sigma_s^2\tilde{\lambda}_i}{\sigma_u^2}\right)\\ &\to \int_a^b \log_2(1 + \Gamma x)f_y(x)dx, \qquad (12.1.15)\\ &\triangleq \tilde{I}_1(\Gamma, y),\end{aligned}$$

where $\Gamma = \frac{\sigma_s^2}{\sigma_u^2}$. The limit above is termed the Shannon transform of the M-P law, and the closed-form expression for the limit can be found in Chapter 3.

2. CSI-known case

We are interested in the normalized mutual information

$$\tilde{I}_{\mathbf{H}} = \frac{1}{p}\log_2\left\{\det\left(\mathbf{I}_p + \frac{\sigma_s^2}{p\sigma_u^2}\mathbf{HH}^*\right)\right\}$$

$$
\begin{aligned}
&= \frac{1}{p}\sum_{i=1}^{p}\log_2\left(1+\frac{\sigma_s^2\lambda_i}{p\sigma^2}\right)\\
&\to \int_a^b \log_2(1+\frac{1}{y}\Gamma x)f_y(x)dx \qquad (12.1.16)\\
&\triangleq \tilde{I}_2(\Gamma, y).
\end{aligned}
$$

Note that $\tilde{I}_1(\Gamma, y)$ and $\tilde{I}_1(\Gamma, y)$ are only derived for the case $y < 1$. For the case $y \geq 1$, the formulas (12.1.15) and (12.1.16) still hold, although the M-P laws are different in the point mass at the origin for the two cases.

12.1.6 A General DS-CDMA Model

In this section, we return to the DS-CDMA system and consider a more general model where information is simultaneously transmitted to multiple antennas. We assume we have n users with p-dimensional spreading sequence $\mathbf{h}_i$ assigned to user i. We let $x_i \in \mathbb{R}$ denote user i's transmitted symbols, assumed to be iid standardized random variables across users, and $T_i \in \mathbb{R}^+$ user i's transmission power. We let L denote the number of antennas and $\gamma_i(\ell)$ the fading channel gain from user i to antenna ℓ. With $\mathbf{x}_\ell$ and $\mathbf{u}_\ell$ denoting, respectively, the received signal vector and the received noise vector to antenna ℓ, we have the matrix model in (12.1.1) for the transmission to the ℓ-th antenna with $s_i = s_i(\ell) = x_i\sqrt{T_i}\gamma_i(\ell)$; that is,

$$
\mathbf{x}_\ell = \sum_{i=1}^{n} x_i\sqrt{T_i}\gamma_i(\ell)\mathbf{h}_i + \mathbf{u}_\ell.
$$

The matrix $\mathbf{H}$ remains the same for all antennas. We assume the components of all the $\mathbf{u}_\ell$ are iid with mean zero and expected second absolute moment equal to σ^2.

Let $\mathbf{x} = [\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_L']'$. Estimating x_i for each user is done by taking the inner product of $\mathbf{x}$ with an appropriate vector $\mathbf{c}_i \in \mathbb{C}^{p\ell}$, called the linear receiver for user i. Define

$$
\hat{\mathbf{h}}_i = \sqrt{T}_i[\gamma_i(1)\mathbf{h}_i', \gamma_i(2)\mathbf{h}_i', \cdots, \gamma_i(\ell)\mathbf{h}_i']'.
$$

The quantity

$$
\frac{|\mathbf{c}_1^*\hat{\mathbf{h}}_1|^2}{\sigma^2\|\mathbf{c}_1\|^2 + \sum_{i=2}^{n}|\mathbf{c}_1^*\hat{\mathbf{h}}_i|^2}
$$

is called the *signal-to-interference ratio* associated with user 1 and is typically used as a measure for evaluating the performance of the linear receiver. It can be shown that the choice of $\mathbf{c}_1$ that minimizes $\mathrm{E}(\mathbf{c}_1^*\mathbf{x} - x_1)^2$ also maximizes the signal-to-interference ratio associated with user 1, the maximum value equal to

$$\mathrm{SIR}_1 = \hat{\mathbf{h}}_1^* \left(\sum_{i=2}^{K} \hat{\mathbf{h}}_i \hat{\mathbf{h}}_i^* + \sigma^2 \mathbf{I} \right)^{-1} \hat{\mathbf{h}}_1,$$

where $\mathbf{I}$ is the $pL \times pL$ identity matrix.

Although SIR_1 does not depend on eigenvalues of a random matrix, when n and p are large, tools used in RMT, notably Lemma B.26 and basic properties of matrices, can be implemented to prove a limit theorem as n and p approach infinity, while their ratio approaches a positive constant. The following is proven in Bai and Silverstein [28] (see also Cottatellucci and Müller [83] and Hanly and Tse [146], the latter containing a simplified version of this model):

Theorem 12.1. *Let* $\{h_{ij} : i, j = 1, 2, \ldots\}$ *be a doubly infinite array of iid complex random variables with* $\exp h_{11} = 0$, $\exp |h_{11}|^2 = 1$. *Define for* $i = 1, 2, \ldots, n$ $\mathbf{h}_i = \mathbf{h}_i(n) = (h_{1\,i}, h_{2\,i}, \ldots, h_{p\,i})'$. *We assume* $n = n(p)$ *and* $n/p \to y > 0$ *as* $p \to \infty$. *For each* p, *let* $\gamma_i(\ell) = \gamma_i^p(\ell) \in \mathbb{C}$, $T_i = T_i^p \in \mathbb{R}^+$, $i = 1, \ldots, n$, $\ell = 1, \ldots, L$ *be random variables, independent of* $\mathbf{h}_1, \ldots, \mathbf{h}_n$. *For each* p *and* i, *let*

$$\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_i^p = \sqrt{T_i}(\gamma_i(1), \ldots, \gamma_i(L))'.$$

Assume almost surely that the empirical distribution of $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ *weakly converges to a probability distribution* H *in* $\mathbb{C}^L$.

Let $\boldsymbol{\beta}_i = \boldsymbol{\beta}_i(p) = \sqrt{T_k}(\gamma_i(1)\mathbf{h}_i', \ldots, \gamma_i(L)\mathbf{h}_i')'$ *and*

$$\mathbf{C} = \mathbf{C}(p) = \frac{1}{p} \sum_{i=2}^{n} \boldsymbol{\beta}_i \boldsymbol{\beta}_i^*.$$

Define

$$SIR_1 = \frac{1}{p} \boldsymbol{\beta}_1^* (\mathbf{C} + \sigma^2 \mathbf{I})^{-1} \boldsymbol{\beta}_1.$$

Then, with probability 1,

$$\lim_{p \to \infty} SIR_1 = T_1 \sum_{\ell, \ell' = 1}^{L} \bar{\gamma}_1(\ell) \gamma_1(\ell') a_{\ell, \ell'},$$

where the $L \times L$ *matrix* $\mathbf{A} = (a_{\ell,\ell'})$ *is nonrandom, Hermitian positive definite, and is the unique Hermitian positive definite matrix satisfying*

$$\mathbf{A} = \left(y \,\mathrm{E} \frac{\boldsymbol{\alpha}\boldsymbol{\alpha}^*}{1 + \boldsymbol{\alpha}^* \mathbf{A} \boldsymbol{\alpha}} + \sigma^2 \mathbf{I}_L \right)^{-1},$$

where $\boldsymbol{\alpha} \in \mathbb{C}^L$ *has distribution* H *and* $\mathbf{I}_L$ *is the* $L \times L$ *identity matrix.*

The theorem assumes the entries of the spreading sequences are iid with mean zero and variance $1/p$. The scaling by $1/\sqrt{p}$ is removed from the definition of the $\mathbf{h}_i$'s.

Clearly SIR_1 defined in this theorem is the same as the one initially introduced, with the only difference in notation being the removal of the scaling by $1/\sqrt{n}$ in the definition of the $\mathbf{h}_i$'s.

Two separate assumptions are imposed in Hanly and Tse. One of them restricts applications to scenarios where all the antennas are near each other. The other assumptions imposed lift the restrictions but assume for each user that independent spreading sequences are going to the L antennas, which is completely unrealistic. Both assumptions assume the entries of the spreading sequences to be mean-zero complex Gaussian. Clearly Theorem 12.1 allows arbitrary scenarios to be considered. There is no restriction as to the placement of the antennas. Moreover, the general assumptions made on the entries of the $\mathbf{h}_i$'s can allow them to be ± 1, which is typically done in practice.

The proof of Theorem 12.1, besides relying on Lemma B.26, which essentially handles the random nature of SIR_1, uses identities involving inverses of matrices expressed in block form, most notably the following.

Lemma 12.2. *Suppose* $\mathbf{A}_1, ..., \mathbf{A}_L$ *are* $p \times n$*, and* $\sigma^2 > 0$*. Define the* ℓ, ℓ' *block of the* $pL \times pL$ *matrix* A *by* $\mathbf{A}_{\ell,\ell'} = \mathbf{A}_\ell \mathbf{A}_{\ell'}^*$ *and, splitting* $(\mathbf{A} + \sigma^2 \mathbf{I})^{-1}$ *into* L^2 $p \times p$ *matrices, let* $(\mathbf{A} + \sigma^2 \mathbf{I})_{\ell,\ell'}^{-1}$ *denote its* ℓ, ℓ' *block. Then*

$$(\mathbf{A} + \sigma^2 \mathbf{I})_{\ell,\ell'}^{-1} = \sigma^{-2} \left(\delta_{\ell,\ell'} \mathbf{I}_p - \mathbf{A}_\ell \left(\sum_{\underline{\ell}} \mathbf{A}_{\underline{\ell}}^* \mathbf{A}_{\underline{\ell}} + \sigma^2 \mathbf{I}_n \right)^{-1} \mathbf{A}_{\ell'}^* \right).$$

For further details, we refer the reader to Bai and Silverstein [28].

12.2 Application to Finance

Today, the financial environment is widely recognized to be riskier than it had been in past decades. The change was significant during the second half of the twentieth century. Price indices went up and the volatility of foreign exchange rates, interest rates, and commodity prices all increased. All firms and financial institutions are facing uncertainty due to changes in the financial markets. The markets for risk management products have grown dramatically since the 1980s. Risk management has become a key technique for all market participants. Risk should be carefully measured. Var (Value at risk) matrix and credit matrix have become popular terminologies in banks and fund management companies.

The wide adoption of modern computers in all financial institutions and markets has made it possible to do exchanges expeditiously and prices to vary abruptly. Also, the emergence of various mutual funds makes the investigation on finance global, and hence large dimensional data analysis has received tremendous attention in financial research. Over the last one or two decades, the application of RMT has appeared in many research papers and risk management institutions. For example, the correlation matrix and factor models that work on internal or external measures for financial risk have become well known in all financial institutions. In this section, we shall briefly introduce some applications of RMT to finance problems.

12.2.1 A Review of Portfolio and Risk Management

Optimal portfolio selection is a very useful strategy for investors. Since being proposed by Markowitz [205], it has received great attention and interest from both theoreticians and practitioners in finance.

The use of these criteria was defined in terms of the theory of rational behavior under risk and uncertainty as developed by von Neumann and Morgenstern [220] and Savage [250]. The relationship between many-period and single-period utility analyses was explained by Bellman [49], and algorithms were provided to compute portfolios that minimize variance or semivariance for various levels of expected returns once requisite estimates concerning securities are provided.

Portfolio theory refers to an investment strategy that seeks to construct an optimal portfolio by considering the relationship between risk and return. The fundamental issue of capital investment should no longer be to pick out good stocks but to diversify the wealth among different assets. The success of investment depends not only on return but also on risk. Risk is influenced by correlations between different assets such that the portfolio selection represents an optimization problem.

1. Optimal portfolio selection—mean-variance model

Suppose there are p assets whose returns $R_1, \cdots R_p$ are random variables with known means $\mathrm{E}R_i$ and covariances $\mathrm{Cov}(R_i, R_j)$. Denote $\mathbf{R} = (R_1, \cdots R_p)'$, $\mathbf{r} = \mathrm{E}\mathbf{R} = (r_1, \cdots r_p)'$, $\Sigma = \mathrm{Var}\mathbf{R} = \mathrm{E}(\mathbf{R}-\mathbf{r})(\mathbf{R}-\mathbf{r})' = (\sigma_{ij})$. Consider $\mathcal{P}$, a portfolio; i.e., a vector of weights (the ratio of different stocks in a portfolio, or loadings in some literature) $\mathbf{w} = (w_1, \cdots, w_p)'$. We impose a budget constraint

$$\sum_{i=1}^{p} w_i = \mathbf{w}'\mathbf{1} = 1,$$

where $\mathbf{1}$ is a vector of ones. If additionally $\forall i,\ w_i \geq 0$, the short sale is excluded.

If the return of a whole portfolio $\mathcal{P}$ is denoted by $R_{\mathcal{P}}$, then

$$R_{\mathcal{P}} = \sum_{i=1}^{p} w_i R_i = \mathbf{w}'\mathbf{R}$$

and

$$r_{\mathcal{P}} = \mathrm{E}R_{\mathcal{P}} = \sum_{i}^{p} w_i \mathrm{E}R_i = \sum_{i}^{p} w_i r_i = \mathbf{w}'\mathbf{r}.$$

The variance (or risk) of return is $\sigma_{\mathcal{P}}^2 = \mathbf{w}'\Sigma\mathbf{w}$.

According to Markowitz, a rational investor always searches for $\mathbf{w}$ that minimizes the risk at a given level of expected return R_0,

$$\min\left\{\mathbf{w}'\Sigma\mathbf{w}\middle|\mathbf{w}'\mathbf{r} \geq R_0 \text{ and } \mathbf{w}'\mathbf{1} = 1,\ w_i \geq 0\right\},$$

or its dual version to maximize the expected return under a given risk level σ_0^2,

$$\max\left\{\mathbf{w}'\mathbf{r}\middle|\mathbf{w}'\Sigma\mathbf{w} \leq \sigma_0^2 \text{ and } \mathbf{w}'\mathbf{1} = 1,\ w_i \geq 0\right\}.$$

When we use absolute deviation to measure risk, we get the mean absolute deviation model that minimizes $\mathrm{E}\left|\mathbf{w}'\mathbf{R} - \mathbf{w}'\mathbf{r}\right|$. If semivariance is considered, we minimize $\mathbf{w}'V_{-}\mathbf{w}$, where

$$V_{-} = \mathrm{Cov}((R_i - r_i)_{-}, (R_j - r_j)_{-})$$

$$(R_i - r_i)_{-} = [-(R_i - r_i)] \vee 0.$$

Sometimes a utility function is used to evaluate the investment performance, say $\ln x$. The utility of a portfolio $\mathcal{P}$ is $\sum_{i=1}^{p} \ln r_i$. Let $\widetilde{\Sigma} = (\widetilde{\sigma}_{ij})$ be the semivariance as a measure of risk, where

$$\widetilde{\sigma}_{ij} = \mathrm{Cov}((\ln R_i - \ln r_i)_{-},\ (\ln R_j - \ln r_j)_{-}).$$

Then, we come to the **log-utility model**:

$$\min\left\{\mathbf{w}'\widetilde{\Sigma}\mathbf{w}\middle|\sum_{i=1}^{p} w_i \ln r_i \geq R_0,\quad \mathbf{w}'\mathbf{1} = 1,\quad w_i \geq 0\right\}.$$

A portfolio is said to be legitimate if it satisfies constraints $\mathbf{A}\mathbf{w} = \mathbf{b}, \mathbf{w} \geq 0$. The reader should note that the expected return of a portfolio is denoted by E and the variance of the portfolio by V ($V = \mathbf{w}'\Sigma\mathbf{w}$ or $\mathbf{w}'\widetilde{\mathbf{\Sigma}}\mathbf{w}$ in different models). An E-V pair (E_0, V_0) is said to be obtainable if there is a legitimate portfolio $\mathbf{w}_0$ such that $E_0 = \mathbf{w}_0'\mathbf{r}$ and $V_0 = \mathbf{w}_0'\Sigma\mathbf{w}_0$. An E-V pair is said to be efficient if (1) the pair (E_0, V_0) is obtainable and (2) there is no obtainable (E_1, V_1) such that either $E_1 > E_0$ while $V_1 \leq V_0$ or $E_1 \geq E_0$ while $V_1 < V_0$. A portfolio $\mathbf{w}$ is efficient if it is legitimate and if its E-V pair is efficient. The problem is to find the set of all efficient E-V pairs and a legitimate portfolio for each efficient E-V pair. Kuhn and Tucker's results [179] on nonlinear programming are applicable in solving optimization problems. A simplex method is also applicable to the quadratic programming for portfolio selection, as shown by Wolfe [298].

2. Financial correlations and information extraction
Because the means and covariances of the return are practically unknown, to estimate the risk of a given portfolio it is natural to use the sample means vector and covariances of $R_{i,t}$ and $R_{j,t}$ of some historical data $\{\mathbf{R}_t = (R_{1,t}, \cdots, R_{p,t})'\}$ observed at discrete time instants $t = 1, \ldots, n$,

$$\hat{r}_i = \bar{R}_i = \frac{1}{n}\sum_{t=1}^{n} R_{i,t}, \quad \hat{\sigma}_{ij} = \widehat{\text{Cov}}(R_i, R_j) = \frac{1}{n}\sum_{t=1}^{n}(R_{i,t} - \bar{R}_i)(R_{j,t} - \bar{R}_j).$$

Theoretically, the covariances can be well estimated from the historical data. But in real practice it is not the case. The empirical covariance matrix from historical data is in fact random and noisy. That means the optimal risk and return of a portfolio are in fact neither well estimated nor controllable. We are facing the problems of covariance matrix cleaning in order to construct an efficient portfolio.

To estimate the correlation matrix $\mathbf{C} = (C_{ij})$, recalling $\Sigma = \mathbf{DCD}$, $\mathbf{D} = \text{diag}(\sigma_1, \ldots \sigma_p)$, where σ_i^2 is the variance of R_i, we need to determine the $p(p+1)/2$ coefficients from the p-dimensional time series of length n. Denoting $y = p/n$, only when $y \ll 1$ can we accurately determine the true correlation matrix. Denote $X_{i,t} = (R_{i,t} - \bar{R}_i)/\sigma_i$, and the empirical correlation matrix (ECM) is $\mathbf{H} = (h_{ij})$, where

$$h_{ij} = \frac{1}{n}\sum_{t=1}^{n} X_{i,t}X_{j,t}.$$

If $n < p$, the matrix $\mathbf{H}$ has $\text{rank}(\mathbf{H}) = n < p$ and thus has $p - n$ zero eigenvalues. The risk of a portfolio can then be measured by

$$\frac{1}{n}\sum_{i,j,t} w_i\sigma_i X_{i,t}X_{j,t}w_j\sigma_j.$$

It is expected to be close to $\sum_{i,j} w_i\sigma_j C_{ij}\sigma_j w_j$. The estimation above is unbiased with mean square error of order $\frac{1}{n}$. But the portfolio is not constructed by a linear function of $\mathbf{H}$, so the risk of a portfolio should be carefully evaluated.

Potters et al. [236] defined the in-sample, out-sample, and true minimum risk as

$$\Sigma_{in} = \mathbf{w}'_{\mathbf{H}}\mathbf{H}\mathbf{w}_{\mathbf{H}} = \frac{R_0^2}{\mathbf{r}'\mathbf{H}^{-1}\mathbf{r}},$$

$$\Sigma_{true} = \mathbf{w}'_{\mathbf{C}}\mathbf{C}\mathbf{w}_{\mathbf{C}} = \frac{R_0^2}{\mathbf{r}'\mathbf{C}^{-1}\mathbf{r}},$$

$$\Sigma_{out} = \mathbf{w}'_{\mathbf{H}}\mathbf{C}\mathbf{w}_{\mathbf{H}} = R_0^2 \frac{\mathbf{r}'\mathbf{H}^{-1}\mathbf{C}\mathbf{H}^{-1}\mathbf{r}}{(\mathbf{r}'\mathbf{H}^{-1}\mathbf{r})^2},$$

where

$$\mathbf{w}_{\mathbf{C}} = R_0 \frac{\mathbf{C}^{-1}\mathbf{r}}{\mathbf{r}'\mathbf{C}^{-1}\mathbf{r}}.$$

Since $\mathrm{E}(\mathbf{H}) = \mathbf{C}$, for large n and p we have approximately

$$\mathbf{r}'\mathbf{H}^{-1}\mathbf{r} \sim \mathrm{E}(\mathbf{r}'\mathbf{H}^{-1}\mathbf{r}) \geq \mathbf{r}'\mathbf{C}^{-1}\mathbf{r}.$$

So, with high probability, we have

$$\Sigma_{in} \leq \Sigma_{true} \leq \Sigma_{out}.$$

This indicates that the in-sample risk provides an underestimation of true risk, while the out-sample risk is an overestimation. Even when both the in-sample and out-sample risks are not unbiased estimators of the true risk, one might be thinking that the difference would become smaller when the sample size increased. However, that is not the case. When the true correlation matrix is the identity matrix $\mathbf{I}$ and $p/n \to y \in (0,1)$, Pafka et al. [225] showed that

$$\Sigma_{true} = \frac{R_0^2}{\mathbf{r}'\mathbf{r}}$$

and

$$\Sigma_{in} \simeq \Sigma_{true}\sqrt{1-y} \simeq \Sigma_{out}(1-y).$$

When and only when $y \to 0$ will all three coincide.

Denote by λ_k and $(V_{1,k}, \cdots, V_{p,k})'$ the eigenvalues and eigenvectors of the correlation matrix $\mathbf{H}$. Then the empirical loading weights are approximately

$$w_i \propto \sum_{k,j} \lambda_k^{-1} V_{i,k} V_{j,k} r_j = r_i + \sum_{k,j} (\lambda_k^{-1} - 1) V_{i,k} V_{j,k} r_j.$$

When $\sigma_i = 1$, the optimal portfolio should invest proportionally to get the expected return r_i, which is the first term of the RHS of the expression above. The second term is in fact an error caused by the estimation error of the eigenvalues of the correlation matrix according to $\lambda > 1$ or $\lambda < 1$. It is possible that the Markowitz solution will allocate a large weight to a small eigenvalue and cause the domination of measurement noise. To avoid the instability of empirical risk, people might use

$$w_i \propto r_i - \sum_{k \leq k^*; j} V_{i,k} V_{j,k} r_j,$$

projecting out the k^* eigenvectors corresponding to the largest eigenvalues.

3. Cleaning of ECM

Therefore, various methods of cleaning the ECM are developed in the literature; see Papp et al. [228], Sharifi et al. [253], and Conlon et al. [82], among others.

Shrinkage estimation is a way of correlation cleaning. Let $\mathbf{H}_c$ denote the cleaned correlation matrix

$$\mathbf{H}_c = \alpha\mathbf{H} + (1-\alpha)\mathbf{I},$$

$$\lambda_{c,k} = 1 + \alpha(\lambda_k - 1),$$

where $\lambda_{c,k}$ is the k-th eigenvalue of $\mathbf{H}$. The parameter α is related to the expected signal-to-noise ratio, $\alpha \in (0,1)$. That $\alpha \to 0$ means the noise is large. Laloux et al. [182] suggest the eigenvalue cleaning method

$$\lambda_{c,k} = \begin{cases} 1-\delta, & \text{if } k > k^*, \\ \lambda_k, & \text{if } k \le k^*, \end{cases}$$

where k^* is the number of meaningful sectors and δ is chosen to preserve the trace of the matrix. The choice of k^* is based on random matrix theory. The key point is to fix k^* such that eigenvalue λ_{k^*} of $\mathbf{H}$ is close to the theoretical left edge of the random part of the eigenvalue distribution.

4. Spectral Theory of ECM

The spectrum discussed in Chapter 3 set up the foundation for applications here. Consider an ECM $\mathbf{H}$ of p assets and n data points, both large with $y = p/n$ finite. Under existence of the second moments, the LSD of a correlation matrix

$$R = \frac{1}{n}\mathbf{A}\mathbf{A}'$$

is

$$P(x) = \frac{1}{2\pi y\sigma^2 x}\sqrt{(b-x)(x-a)} \qquad x \in (a,b) \qquad \text{(the M-P law)},$$

where $\mathbf{A}$ is a $p \times n$ matrix with iid entries of zero mean and unit variance, $a, b = \sigma^2(1 \mp \sqrt{y})^2$ being the bounds of the M-P law. Comparing the eigenvalues of ECM with $P(x)$, one can identify the deviating eigenvalues. These deviating eigenvalues are said to contain information about the system under consideration. If the correlation matrix $\mathbf{C}$ has one eigenvalue larger than $1+\sqrt{y}$, it has been shown by Baik et al. [44] that the largest eigenvalue of the ECM $\mathbf{H}$ will be Gaussian with a center outside the "M-P sea" and a width $\sim \frac{1}{\sqrt{n}}$, smaller than the uncertainty on the bulk eigenvalues (of order $\sim \sqrt{y}$). Then the number k^* can be determined by the expected edge of the bulk eigenvalues. The cleaned correlation matrix is used to construct the portfolio. Other cleaning methods include clustering analysis, by R. N. Mantegna. Empirical studies have reported that the risk of the optimized portfolio obtained using the cleaned correlation matrix is more reliable (see LaLoux et al. [182]), and less than 5% of the eigenvalues appear to carry most of the information.

To extract information from noisy time series, we need to assess the degree to which an ECM is noise-dominated. By comparing the eigenspectra properties, we identify the eigenstates of the ECM that contain genuine information content. Other remaining eigenstates will be noise-dominated and unstable. To analyze the structure of eigenvectors lying outside the 'M-P sea,' Ormerod [223], and Rojkova et al. [242], calculate the inverse participation ratio (IPR) (see Plerou et al. [235, 234]). Given the k-th eigenvalue λ_k and the corresponding eigenvector $\mathbf{V}_k$ with components $V_{k,i}$, the IPR is defined by

$$I_k = \sum_{i=1}^{p} (V_{k,i}^4).$$

It is commonly used in localization theory to quantify the contribution of different components of an eigenvector to the magnitude of that eigenvector. Two extreme cases are those where the eigenvector has identical components $V_{k,i} = \frac{1}{\sqrt{p}}$ or has only one nonzero component. For the two cases, we get $I_k = \frac{1}{p}$ and 1, respectively. When applied to finance, IPR is the reciprocal of the number of eigenvector components significantly different from zero (i.e., the number of economies contributing to that eigenvector). By analyzing the quarterly levels of GDP over the period 1977–2000 from the OECD database for EU economics, France, Germany, Italy, Spain and the UK, Ormerod shows that the co-movement over time between the growth rates of the EU economies does contain a large amount of information.

12.2.2 Enhancement to a Plug-in Portfolio

As mentioned in the last subsection, the plug-in procedure will cause the optimal portfolio selection to be strongly biased, and hence such a phenomenon is called "Markowitz's enigma" in the literature. In this subsection, we will introduce an improvement to the plug-in portfolio by using RMT. The main results are given in Bai it et al. [20].

1. Optimal Solution to the Portfolio Selection

As mentioned earlier, maximizing the return and minimizing the risk are complementary. Thus, we consider the maximization problem as

$$R = \max \mathbf{w}'\boldsymbol{\mu} \text{ subject to } \mathbf{w}'\mathbf{1} \le 1 \text{ and } \mathbf{w}'\Sigma\mathbf{w} \le \sigma_0^2. \tag{12.2.17}$$

We remark here that the condition $\mathbf{w}'\mathbf{1} = 1$ has been weakened to $\mathbf{w}'\mathbf{1} \le 1$ in order to prevent the maximization from having no solution if σ_0^2 is too small. If σ_0^2 is large enough, the optimization solution automatically satisfies $\mathbf{w}'\mathbf{1} = 1$.

The solution is given as follows:

1. If
$$\frac{\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}\sigma_0}{\sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}} \leq 1,$$
then the optimal return R and corresponding investment portfolio $\mathbf{w}$ will be
$$R = \sigma_0\sqrt{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}}$$
and
$$\mathbf{w} = \frac{\sigma_0}{\sqrt{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}}}\Sigma^{-1}\boldsymbol{\mu}.$$
2. If
$$\frac{\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}\sigma_0}{\sqrt{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}} > 1,$$
then the optimal return R and corresponding investment portfolio $\mathbf{w}$ will be
$$R = \frac{\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} + b\left(\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} - \frac{(\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu})^2}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\right)$$
and
$$\mathbf{w} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} + b\left(\Sigma^{-1}\boldsymbol{\mu} - \frac{\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\Sigma^{-1}\mathbf{1}\right),$$
where
$$b = \sqrt{\frac{\mathbf{1}'\sum^{-1}\mathbf{1}\sigma_0^2 - 1}{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}\,\mathbf{1}'\Sigma^{-1}\mathbf{1} - (\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu})^2}}.$$

2. Overprediction of the Plug-in Procedure

As mentioned earlier, the substitution of the sample mean and covariance matrix into Markowitz's optimal selection (called the plug-in procedure) will always cause the empirical return to be much higher than the theoretical optimal return. We call this phenomenon "overprediction" of the plug-in procedure. The following theorem theoretically proves this phenomenon under very mild conditions.

Theorem 12.3. *Assume that* $\mathbf{y}_1, \cdots, \mathbf{y}_n$ *are* n *independent random* p*-vectors of iid entries with mean zero and variance* 1*. Suppose that* $\mathbf{x}_k = \boldsymbol{\mu} + \mathbf{z}_k$ *with* $\mathbf{z}_k = \Sigma^{\frac{1}{2}}\mathbf{y}_k$*, where* $\boldsymbol{\mu}$ *is an unknown* p*-vector and* Σ *is an unknown* $p \times p$ *covariance matrix. Also, we assume that the entries of* $\mathbf{y}_k$*'s have finite fourth moments and that as* $p/n \to y \in (0, 1)$ *we have*
$$\frac{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}}{n} \to a_1\,, \quad \frac{\mathit{1}'\Sigma^{-1}\mathit{1}}{n} \to a_2\,, \quad \frac{\mathit{1}'\Sigma^{-1}\boldsymbol{\mu}}{n}\,, \to a_3,$$

satisfying $a_1 a_2 - a_3^2 > 0$. *Then, with probability* 1, *we have*

$$\lim_{n\to\infty} \frac{\hat{\hat{R}}_p}{\sqrt{n}} = \begin{cases} \sqrt{\gamma a_1} > \lim\limits_{n\to\infty} \dfrac{R^{(1)}}{\sqrt{n}} = \sqrt{a_1}, & \text{when } \; a_3 < 0, \\ \sigma_0 \sqrt{\frac{\gamma(a_1 a_2 - a_3^2)}{a_2}} > \lim\limits_{n\to\infty} \dfrac{R^{(2)}}{\sqrt{n}} = \sigma_0 \sqrt{\dfrac{a_1 a_2 - a_3^2}{a_2}}, & \text{when } \; a_3 > 0, \end{cases}$$

where $R^{(1)}$ *and* $R^{(2)}$ *are the returns for the two cases given in the last paragraph,* $\gamma = \int_a^b \frac{1}{x} dF_y(x) = \frac{1}{1-y} > 1$, $a = (1-\sqrt{y})^2$, *and* $b = (1+\sqrt{y})^2$.

Remark 12.4. The optimal return takes the form $R^{(1)}$ if $\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu} < \sqrt{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}}$. When $a_3 < 0$, for all large n, the condition for the first case holds, and hence we obtain the limit for the first case. If $a_3 > 0$, the condition $\mathbf{1}'\Sigma^{-1}\boldsymbol{\mu} < \sqrt{\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}}$ is eventually not true for all large n and hence the return takes the form $R^{(2)}$. When $a_3 = 0$, the case becomes very complicated. The return may attain the value in both cases and, hence, $\frac{\hat{\hat{R}}_p}{\sqrt{n}}$ may jump between the two limit points.

To illustrate the overprediction phenomenon, for simplicity we generate p-branch standardized security returns from a multivariate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)'$ and identity covariance matrix $\Sigma = \mathbf{I}$. Given the level of risk with the known population mean vector, $\boldsymbol{\mu}$, and known population covariance matrix, Σ, we can compute the theoretical optimal allocation $\mathbf{w}$ and thereafter compute the theoretical optimal return, R, for the portfolios. Using this data set, we compute the sample mean, $\overline{\mathbf{x}}$, and covariance matrix, $\mathbf{S}$, and then the plug-in return, $\hat{R}_p$, and its corresponding plug-in allocation, $\hat{\mathbf{w}}_p$. We finally plot the theoretical optimal returns R and the plug-in returns $\hat{R}_p$ against different values of p with the fixed sample size $n = 500$ in Fig. 12.10. We present the simulated theoretical optimal returns R and the plug-in returns $\hat{R}_p$ in Table 12.1 for two different cases: (A) for different values of p with the same dimension-to-sample-size ratio p/n $(= 0.5)$ and (B) for the same value of p $(= 252)$ but different dimension-to-sample-size ratios p/n.

From Fig. 12.10 and Table 12.1, we find the following: (1) the plug-in return $\hat{R}_p$ is close to the theoretical optimal return R when p is small (≤ 30); (2) when p is large (≥ 60), the difference between the theoretical optimal return R and the plug-in return $\hat{R}_p$ becomes dramatically large; (3) the larger the p, the greater the difference; and (4) when p is large, the plug-in return $\hat{R}_p$ is always larger than the theoretical optimal return R. These confirm the "Markowitz optimization enigma" that the plug-in return $\hat{R}_p$ should not be used in practice.

3. Enhancement by bootstrapping

Now, we construct a parametric bootstrap-corrected estimate $\hat{R}_p$ as follows.

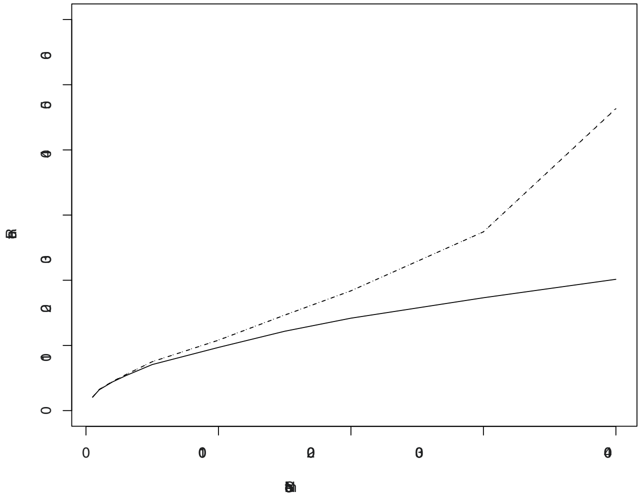

Fig. 12.10 Empirical and theoretical optimal returns. Solid line—the theoretic optimal return $R = \mathbf{w}'\boldsymbol{\mu}$; dashed line—the plug-in return $\hat{R}_p = \mathbf{w}'_{pl}\boldsymbol{\mu}$.

Table 12.1 Performance of $\hat{R}_p$ and $\hat{\hat{R}}_p$.

p	p/n	R	$\hat{R}_p$	$\hat{\hat{R}}_p$	p	p/n	R	$\hat{R}_p$	$\hat{\hat{R}}_p$
100	0.5	9.77	13.89	13.96	252	0.5	14.71	20.95	21.00
200	0.5	13.93	19.67	19.73	252	0.6	14.71	23.42	23.49
300	0.5	17.46	24.63	24.66	252	0.7	14.71	26.80	26.92
400	0.5	19.88	27.83	27.85	252	0.8	14.71	33.88	34.05
500	0.5	22.29	31.54	31.60	252	0.9	14.71	48.62	48.74

Note: In the table, $\hat{\hat{R}}_p = \hat{\mathbf{w}}'\overline{\mathbf{X}}$ is the estimated return. The table compares the performance between $\hat{R}_p$ and $\hat{\hat{R}}_p$ for the same p/n ratio with different numbers of assets, p, and for the same p with different p/n ratios where n is the number of samples and R is the optimal return defined in (12.2.17).

To avoid the singularity of the resampled covariance matrix, we employ the parametric bootstrap method. Suppose that $\chi = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ is the data set. Denote its sample mean and covariance matrix by $\bar{\mathbf{x}}$ and $\mathbf{S}$. First, draw a resample $\chi^* = \{\mathbf{x}_1^*, \cdots, \mathbf{x}_n^*\}$ from the p-variate normal distribution with mean vector $\overline{\mathbf{x}}$ and covariance matrix $\mathbf{S}$. Then, invoking Markowitz's optimization procedure again on the resample χ^*, we obtain the *bootstrapped "plug-in" allocation*, $\hat{\mathbf{w}}_p^*$, and the *bootstrapped "plug-in" return*, $\hat{R}_p^*$, such that

$$\hat{R}_p^* = \hat{\mathbf{c}}_p^{*T}\overline{\mathbf{x}}^*, \tag{12.2.18}$$

where $\overline{\mathbf{x}}^* = \frac{1}{n}\sum_1^n \mathbf{x}_k^*$.

We remind the reader that the bootstrapped "plug-in" allocation $\hat{\mathbf{w}}_p^*$ will be different from the original "plug-in" allocation $\hat{\mathbf{w}}_p$ and, similarly, the bootstrapped "plug-in" return $\hat{R}_p^*$ is different from the "plug-in" return $\hat{R}_p$, but by Theorem 12.3 one can easily prove the following theorem.

Theorem 12.5. *Under the conditions in Theorem 12.3 and using the bootstrapped plug-in procedure as described above, we have*

$$\sqrt{\gamma}(R - \hat{R}_p) \simeq \hat{R}_p - \hat{R}_p^*, \tag{12.2.19}$$

where γ is defined in Theorem 12.3, R is the theoretical optimal return, $\hat{R}_p$ is the plug-in return estimate obtained by the original sample χ, and $\hat{R}_p^*$ is the bootstrapped plug-in return obtained by the bootstrapped sample χ^*.

This theorem leads to the *bootstrap-corrected return estimate* $\hat{R}_b$ and the bootstrap-corrected portfolio $\hat{\mathbf{w}}_b$,

$$\begin{aligned} \hat{R}_b &= \hat{R}_p + \frac{1}{\sqrt{\gamma}}(\hat{R}_p - \hat{R}_p^*), \\ \hat{\mathbf{w}}_b &= \hat{\mathbf{w}}_p + \frac{1}{\sqrt{\gamma}}(\hat{\mathbf{w}}_p - \hat{\mathbf{w}}_p^*). \end{aligned} \tag{12.2.20}$$

4. Monte Carlo study

Now, we present some simulation results showing the superiority of both $\hat{R}_b$ and $\hat{\mathbf{w}}_b$ over their plug-in counterparts $\hat{R}_p$ and $\hat{\mathbf{w}}_p$. To this end, we first define the *bootstrap-corrected difference*, d_b^R, for the return as the difference between the bootstrap-corrected optimal return estimate $\hat{R}_b$ and the theoretical optimal return R; that is,

$$d_b^R = \hat{R}_b - R, \tag{12.2.21}$$

which will be used to compare with the *plug-in difference*,

$$d_p^R = \hat{R}_p - R. \tag{12.2.22}$$

To compare the bootstrapped allocation with the plug-in allocation, we define the *bootstrap-corrected difference norm* d_b^w and the *plug-in difference norm* d_p^w by

$$d_b^w = \|\hat{\mathbf{w}}_b - \mathbf{w}\| \quad \text{and} \quad d_p^w = \|\hat{\mathbf{w}}_p - \mathbf{w}\|. \tag{12.2.23}$$

In the Monte Carlo study, we resample 30 times to get the bootstrapped allocations and then use the average of the bootstrapped allocations to construct the bootstrap-corrected allocation and return for each case of $n = 500$ and $p = 100$, 200, and 300. The results are depicted in Fig. 12.11.

From Fig. 12.11, we find the desired property that d_b^R (d_b^w) is much smaller than d_p^R (d_p^w) for all cases. This suggests that the estimate obtained by uti-

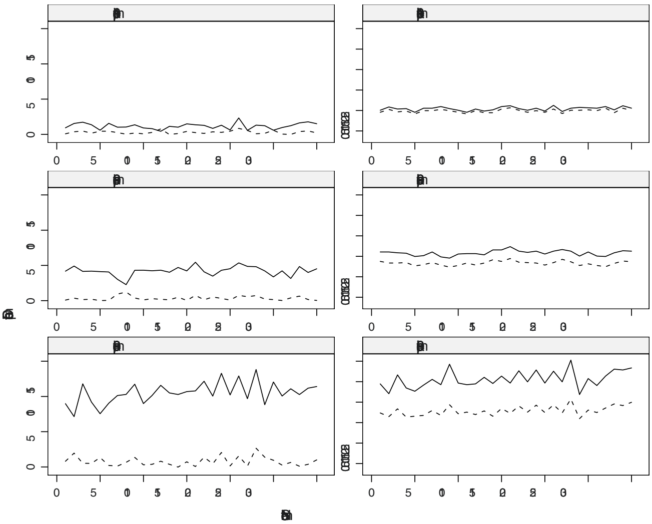

Fig. 12.11 Comparison of portfolio allocations and returns. Solid line—d_p^R and d_p^w, respectively; dashed line—d_b^R and d_b^w, respectively.

lizing the bootstrap-corrected method is much more accurate in estimating the theoretical value than that obtained by using the plug-in procedure. Furthermore, as p increases, the two lines of d_p^R and d_b^R (or d_p^w and d_b^w) on each level as shown in Fig. 12.11 separate further, implying that the magnitude of improvement from d_p^R (d_p^w) to d_b^R (d_b^w) is remarkable.

To further illustrate the superiority of our estimate over the traditional plug-in estimate, we simulated the mean square errors (MSEs) of the various estimates for different p and plot these values in Fig. 12.12. In addition, we define their *relative efficiencies* (REs) for both allocations and returns to be

$$RE_{p,b}^w = \frac{MSE(d_p^w)}{MSE(d_b^w)} \quad \text{and} \quad RE_{p,b}^R = \frac{MSE(d_p^R)}{MSE(d_b^R)} \tag{12.2.24}$$

and report their values in Table 12.2.

5. Comments and discussions

Comparing the MSE of d_b^R (d_b^w) with that of d_p^R (d_p^w) in Table 12.2 and Fig. 12.12, the MSEs of both d_b^R and d_b^w have been reduced dramatically from those of d_p^R and d_p^w, indicating that our proposed estimates are superior. We find that the MSE of d_b^R is only 0.04, improving 6.25 times over that of d_p^R when $p = 50$. When the number of assets increases, the improvement becomes much more substantial. For example, when $p = 350$, the MSE of d_b^R is only 1.59 but the MSE of d_p^R is 220.43, improving 138.64 times over that of d_p^R.

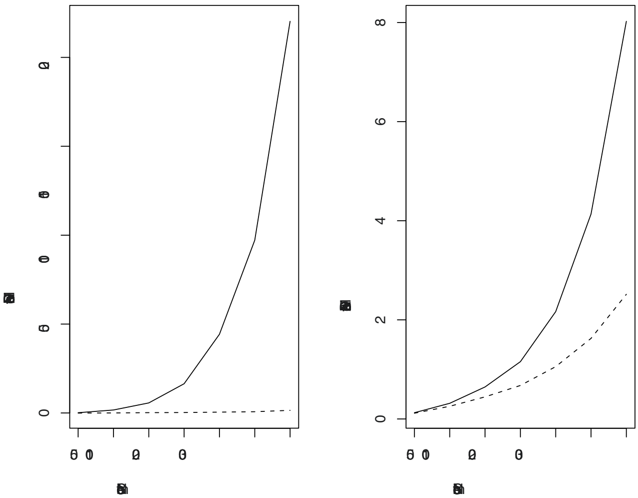

Fig. 12.12 MSE comparison between the empirical and corrected portfolio allocations/returns. Solid Line—the MSE of d_p^R and d_p^c, respectively; dashed line—the MSE of d_b^R and d_b^c, respectively.

Table 12.2 MSE and relative efficiency comparison.

p	MSE(d_p^R)	MSE(d_b^R)	MSE(d_p^w)	MSE(d_b^w)	$RE_{p,b}^R$	$RE_{p,b}^w$
$p = 50$	0.25	0.04	0.13	0.12	6.25	1.08
$p = 100$	1.79	0.12	0.32	0.26	14.92	1.23
$p = 150$	5.76	0.29	0.65	0.45	19.86	1.44
$p = 200$	16.55	0.36	1.16	0.68	45.97	1.71
$p = 250$	44.38	0.58	2.17	1.06	76.52	2.05
$p = 300$	97.30	0.82	4.14	1.63	118.66	2.54
$p = 350$	220.43	1.59	8.03	2.52	138.64	3.19

This is an unbelievable improvement. We note that when both n and p are bigger, the relative efficiency of our proposed estimate over the traditional plug-in estimate could be much larger. On the other hand, the improvement from d_p^c to d_b^w is also tremendous.

We illustrate the superiority of our approach by comparing the estimates of the bootstrap-corrected return and the plug-in return for daily S&P500 data. To match our simulation of $n = 500$ as shown in Table 12.2 and Fig. 12.12, we choose 500 daily data backward from December 30, 2005, for all companies listed in the S&P500 as the database for our estimation. We then choose the number of assets (p) from 5 to 400, and, for each p, we select p stocks from the S&P500 database randomly without replacement and compute the plug-in return and the corresponding bootstrap-corrected return. We plot the plug-in returns and the corresponding bootstrap-corrected returns in Fig. 12.13 and report these returns and their ratios in Table 12.3 for different p. We also repeat the procedure ($m =$) 10 and 100 times for checking. For each m and for each p, we first compute the bootstrap-corrected returns and the plug-in returns. Thereafter, we compute their averages for both the bootstrap-corrected returns and the plug-in returns and plot these values in Panels 2 and 3 of Fig. 12.13, respectively, for comparison with the results in Panel 1 for $m = 1$.

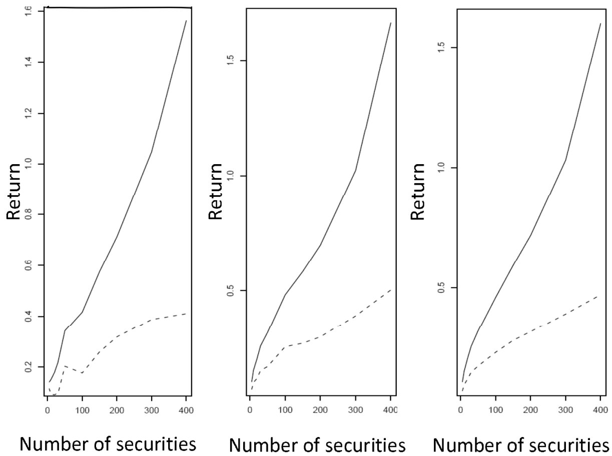

Fig. 12.13 Comparison in returns. Solid line—plug-in return; dashed line—bootstrap-corrected return.

From Table 12.2 and Fig. 12.13, we find that, as the number of assets increases, (1) the values of the estimates from both the bootstrap-corrected returns and the plug-in returns for the S&P500 database increase, and (2) the values of the estimates of the plug-in returns increase much faster than those of the bootstrap-corrected returns and thus their differences become wider. These empirical findings are consistent with the theoretical discovery of the "Markowitz optimization enigma," that the estimated plug-in return is always larger than its theoretical value and their difference becomes larger when the number of assets is large.

Comparing Figs. 12.12 and 12.13 (or Tables 12.3 and 12.1), one will find that the shapes of the graphs of both the bootstrap-corrected returns and the corresponding plug-in returns are similar to those in Fig. 12.10. This suggests that our empirical findings based on the S&P500 are consistent

Table 12.3 Plug-in returns and bootstrap-corrected returns.

p	m=1			m=10			m=100		
	$\hat{R}_p$	$\hat{R}_b$	$\hat{R}_b/\hat{R}_p$	$\hat{R}_p$	$\hat{R}_b$	$\hat{R}_b/\hat{R}_p$	$\hat{R}_p$	$\hat{R}_b$	$\hat{R}_b/\hat{R}_p$
5	0.142	0.116	0.820	0.106	0.074	0.670	0.109	0.072	0.632
10	0.152	0.092	0.607	0.155	0.103	0.650	0.152	0.097	0.616
20	0.179	0.09	0.503	0.204	0.120	0.576	0.206	0.121	0.573
30	0.218	0.097	0.447	0.259	0.154	0.589	0.254	0.148	0.576
50	0.341	0.203	0.597	0.317	0.171	0.529	0.319	0.174	0.541
100	0.416	0.177	0.426	0.482	0.256	0.530	0.459	0.230	0.498
150	0.575	0.259	0.450	0.583	0.271	0.463	0.592	0.279	0.469
200	0.712	0.317	0.445	0.698	0.298	0.423	0.717	0.315	0.438
300	1.047	0.387	0.369	1.023	0.391	0.381	1.031	0.390	0.377
400	1.563	0.410	0.262	1.663	0.503	0.302	1.599	0.470	0.293

with our theoretical and simulation results, which, in turn, confirms that our proposed bootstrap-corrected return performs better.

One may doubt the existence of bias in our sampling, as we choose only one sample in the analysis. To circumvent this problem, we also repeat the procedure m (=10, 100) times. For each m and for each p, we compute the bootstrap-corrected returns and the plug-in returns and then compute the averages for each. Thereafter, we plot the averages of the returns in Fig. 12.13 and report these averages and their ratios in Table 12.3 for $m = 10$ and 100. When comparing the values of the returns for $m = 10$ and 100 with $m = 1$, we find that the plots have basically similar values for each p but become smoother, suggesting that the sampling bias has been eliminated by increasing the value of m. The results for $m = 10$ and 100 are also consistent with the plot in Fig. 12.10 in our simulation, suggesting that our bootstrap-corrected return is a better estimate for the theoretical return in the sense that its value is much closer to the theoretical return when compared with the corresponding plug-in return.

Appendix A
Some Results in Linear Algebra

In this chapter, the reader is assumed to have a college-level knowledge of linear algebra. Therefore, we only introduce those results that will be used in this book.

A.1 Inverse Matrices and Resolvent

A.1.1 Inverse Matrix Formula

Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ matrix. Denote the cofactor of a_{ij} by A_{ij}. The Laplace expansion of the determinant states that, for any j,

$$\det(\mathbf{A}) = \sum_{i=1}^{n} a_{ij} A_{ij}. \tag{A.1.1}$$

Let $\mathbf{A}^a = (A_{ij})'$ denote the adjacent matrix of $\mathbf{A}$. Then, applying the formula above, one immediately gets

$$\mathbf{A}\mathbf{A}^a = \det(\mathbf{A})\mathbf{I}_n.$$

This proves the following theorems.

Theorem A.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix with a nonzero determinant. Then, it is invertible and*

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\mathbf{A}^a. \tag{A.1.2}$$

Theorem A.2. *We have*

$$\operatorname{tr}(\mathbf{A}^{-1}) = \sum_{k=1}^{n} A_{kk}/\det(\mathbf{A}). \tag{A.1.3}$$

A.1.2 Holing a Matrix

The following is known as Hua's holing method:

$$\begin{pmatrix} \mathbf{I} & \mathbf{O} \\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{O} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{pmatrix}. \tag{A.1.4}$$

In application, this formula can be considered as making a row Gaussian elimination on the matrix $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$ to eliminate the (2,1)-th block. A similar column transformation also holds. An important application of this formula is the following theorem.

Theorem A.3. *If* $\mathbf{A}$ *is a squared nonsingular matrix, then*

$$\det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \det(\mathbf{A})\det(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}). \tag{A.1.5}$$

This theorem follows by taking determinants on both sides of (A.1.4).

Note that the transformation (A.1.4) does not change the rank of the matrix. Therefore, it is frequently used to prove rank inequalities.

A.1.3 Trace of an Inverse Matrix

For $n \times n$ $\mathbf{A}$, define $\mathbf{A}_k$, called a major submatrix of order $n-1$, to be the matrix resulting from deleting the k-th row and column from $\mathbf{A}$. Applying (A.1.2) and (A.1.5), we obtain the following useful theorem.

Theorem A.4. *If both* $\mathbf{A}$ *and* $\mathbf{A}_k$, $k = 1, 2, \cdots, n$, *are nonsingular, and if we write* $\mathbf{A}^{-1} = \left[a^{k\ell}\right]$, *then*

$$a^{kk} = \frac{1}{a_{kk} - \boldsymbol{\alpha}_k' \mathbf{A}_k^{-1} \boldsymbol{\beta}_k}, \tag{A.1.6}$$

and hence

$$\operatorname{tr}(\mathbf{A}^{-1}) = \sum_{k=1}^{n} \frac{1}{a_{kk} - \boldsymbol{\alpha}_k' \mathbf{A}_k^{-1} \boldsymbol{\beta}_k}, \tag{A.1.7}$$

where a_{kk} *is the* k*-th diagonal entry of* $\mathbf{A}$, $\mathbf{A}_k$ *is defined above,* $\boldsymbol{\alpha}_k'$ *is the vector obtained from the* k*-th row of* $\mathbf{A}$ *by deleting the* k*-th entry, and* $\boldsymbol{\beta}_k$ *is the vector from the* k*-th column by deleting the* k*-th entry.*

If $\mathbf{A}$ is an $n \times n$ symmetric nonsingular matrix and all its major submatrices of order $(n-1)$ are nonsingular, then from (A.1.7) it follows immediately that

$$\operatorname{tr}(\mathbf{A}^{-1}) = \sum_{k=1}^{n} \frac{1}{a_{kk} - \boldsymbol{\alpha}_k' \mathbf{A}_k^{-1} \boldsymbol{\alpha}_k}. \tag{A.1.8}$$

If $\mathbf{A}$ is an $n \times n$ Hermitian nonsingular matrix and all its major submatrices of order $(n-1)$ are nonsingular, similarly we have

$$\operatorname{tr}(\mathbf{A}^{-1}) = \sum_{k=1}^{n} \frac{1}{a_{kk} - \boldsymbol{\alpha}_k^* \mathbf{A}_k^{-1} \boldsymbol{\alpha}_k},$$

where * denotes the complex conjugate transpose of matrices or vectors.

In this book, we shall frequently consider the resolvent of a Hermitian matrix $\mathbf{X} = (x_{jk})$ (i.e., $\mathbf{A} = (\mathbf{X} - z\mathbf{I})^{-1}$), where z is a complex number with positive imaginary part. In this case, we have

$$\operatorname{tr}((\mathbf{X} - z\mathbf{I})^{-1}) = \sum_{k=1}^{n} \frac{1}{x_{kk} - z - \mathbf{x}_k^* \mathbf{H}_k^{-1} \mathbf{x}_k}, \tag{A.1.9}$$

where $\mathbf{H}_k$ is the matrix obtained from $\mathbf{X} - z\mathbf{I}$ by deleting the k-th row and the k-th column and $\mathbf{x}_k$ is the k-th column of $\mathbf{X}$ with the k-th element removed.

A.1.4 Difference of Traces of a Matrix A *and Its Major Submatrices*

Suppose that the matrix $\boldsymbol{\Sigma}$ is positive definite and has the partition as given by $\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$. Then, the inverse of $\boldsymbol{\Sigma}$ has the form

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1} \\ -\boldsymbol{\Sigma}_{22.1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{\Sigma}_{22.1}^{-1} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{12}$. In fact, the formula above can be derived from the identity (by applying (A.1.4))

$$\begin{pmatrix} \mathbf{I} & \mathbf{O} \\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ \mathbf{O} & \mathbf{I} \end{pmatrix}$$
$$= \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \end{pmatrix}$$

and the fact that

$$\begin{pmatrix} \mathbf{I} & 0 \\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{O} \\ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix}.$$

Making use of this identity, we obtain the following theorem.

Theorem A.5. *If the matrix* $\mathbf{A}$ *and* $\mathbf{A}_k$*, the* k*-th major submatrix of* $\mathbf{A}$ *of order* $(n-1)$*, are both nonsingular and symmetric, then*

$$\operatorname{tr}(\mathbf{A}^{-1}) - \operatorname{tr}(\mathbf{A}_k^{-1}) = \frac{1 + \boldsymbol{\alpha}_k' \mathbf{A}_k^{-2} \boldsymbol{\alpha}_k}{a_{kk} - \boldsymbol{\alpha}_k' \mathbf{A}_k^{-1} \boldsymbol{\alpha}_k}. \tag{A.1.10}$$

If $\mathbf{A}$ *is Hermitian, then* α_k' *is replaced by* α_k^* *in the equality above.*

A.1.5 Inverse Matrix of Complex Matrices

Theorem A.6. *If Hermitian matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *are commutative and such that* $\mathbf{A}^2 + \mathbf{B}^2$ *is nonsingular, then the complex matrix* $\mathbf{A} + i\mathbf{B}$ *is nonsingular and*

$$(\mathbf{A} + i\mathbf{B})^{-1} = (\mathbf{A} - i\mathbf{B})(\mathbf{A}^2 + \mathbf{B}^2)^{-1}. \tag{A.1.11}$$

This can be directly verified.

Let $z = u + iv$, $v > 0$, and let $\mathbf{A}$ be an $n \times n$ Hermitian matrix. Then

$$\left|\operatorname{tr}(\mathbf{A} - z\mathbf{I}_n)^{-1} - \operatorname{tr}(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1}\right| \le v^{-1}. \tag{A.1.12}$$

Proof. By (A.1.10), we have

$$\operatorname{tr}(\mathbf{A} - z\mathbf{I}_n)^{-1} - \operatorname{tr}(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-1} = \frac{1 + \boldsymbol{\alpha}_k^*(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k}{a_{kk} - z - \boldsymbol{\alpha}_k^*(\mathbf{A} - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k}.$$

If we denote $\mathbf{A}_k = \mathbf{E}^* \operatorname{diag}[\lambda_1 \cdots \lambda_{n-1}]\mathbf{E}$ and $\boldsymbol{\alpha}_k^* \mathbf{E}^* = (y_1, \cdots, y_{n-1})$, where $\mathbf{E}$ is an $(n-1) \times (n-1)$ unitary matrix, then we have

$$\begin{aligned}
|1 + \boldsymbol{\alpha}_k^*(\mathbf{A}_k - z\mathbf{I}_{n-1})^{-2}\boldsymbol{\alpha}_k| &= |1 + \sum_{\ell=1}^{n-1} |y_\ell^2|(\lambda_\ell - z)^{-2}| \\
&\le 1 + \sum_{\ell=1}^{n-1} |y_\ell^2|((\lambda_\ell - u)^2 + v^2)^{-1} \\
&= 1 + \boldsymbol{\alpha}_k^*((\mathbf{A}_k - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k.
\end{aligned}$$

On the other hand, by (A.1.11) we have

$$\begin{aligned}
&\Im(a_{kk} - z - \boldsymbol{\alpha}_k^*(\mathbf{A} - z\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k) \\
&= v(1 + \boldsymbol{\alpha}_k^*((\mathbf{A}_k - u\mathbf{I}_{n-1})^2 + v^2\mathbf{I}_{n-1})^{-1}\boldsymbol{\alpha}_k).
\end{aligned} \tag{A.1.13}$$

From these estimates, (A.1.12) follows.

A.2 Inequalities Involving Spectral Distributions

In this section, we shall establish some inequalities to bound the differences between spectral distributions in terms of characteristics of the matrices, say norms or ranks. These inequalities are important in the truncation and centralization techniques.

A.2.1 Singular-Value Inequalities

If $\mathbf{A}$ is a $p \times n$ matrix of complex entries, then its singular values $s_1 \geq \cdots \geq s_q \geq 0$, $q = \min(p, n)$, are defined as the square roots of the q largest eigenvalues of the nonnegative definite Hermitian matrix $\mathbf{A}\mathbf{A}^*$. If $\mathbf{A}$ $(n \times n)$ is Hermitian, then let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ denote its eigenvalues. The following results are well known and are referred to as the *singular decomposition* and *spectral decomposition*, respectively.

Theorem A.7. *Let* $\mathbf{A}$ *be a* $p \times n$ *matrix. Then there exist* q *p-dimensional orthonormal vectors* $\mathbf{u}_1, \cdots \mathbf{u}_q$ *and* q *n-dimensional orthonormal vectors* $\mathbf{v}_1, \cdots, \mathbf{v}_q$ *such that*

$$\mathbf{A} = \sum_{j=1}^{q} s_j \mathbf{u}_j \mathbf{v}_j^*. \tag{A.2.1}$$

From this expression, we immediately get the well-known Courant-Fischer formula

$$s_k = \min_{\mathbf{w}_1, \cdots, \mathbf{w}_{k-1}} \max_{\substack{\|\mathbf{v}\|_2 = 1 \\ \mathbf{v} \perp \mathbf{w}_1, \cdots, \mathbf{w}_{k-1}}} \|\mathbf{A}\mathbf{v}\|_2. \tag{A.2.2}$$

If $\mathbf{A}$ *is an* $n \times n$ *Hermitian matrix, then there exist* n *n-dimensional orthonormal vectors* $\mathbf{u}_1, \cdots \mathbf{u}_n$ *such that*

$$\mathbf{A} = \sum_{j=1}^{n} \lambda_j \mathbf{u}_j \mathbf{u}_j^*. \tag{A.2.3}$$

Similarly, we have the formula

$$\lambda_k = \min_{\mathbf{w}_1, \cdots, \mathbf{w}_{k-1}} \max_{\substack{\|\mathbf{v}\|_2 = 1 \\ \mathbf{v} \perp \mathbf{w}_1, \cdots, \mathbf{w}_{k-1}}} \mathbf{v}^* \mathbf{A} \mathbf{v}. \tag{A.2.4}$$

The following theorem due to Fan [103] is useful for establishing rank inequalities, which will be discussed in the next section.

Theorem A.8. *Let* $\mathbf{A}$ *and* $\mathbf{C}$ *be two* $p \times n$ *complex matrices. Then, for any nonnegative integers* i *and* j, *we have*

$$s_{i+j+1}(\mathbf{A} + \mathbf{C}) \leq s_{i+1}(\mathbf{A}) + s_{j+1}(\mathbf{C}). \tag{A.2.5}$$

Proof. Let $\mathbf{w}_1, \cdots, \mathbf{w}_i$ be the left eigenvectors of $\mathbf{A}$, corresponding to the singular values $s_1(\mathbf{A}), \cdots, s_i(\mathbf{A})$, and let $\mathbf{w}_{i+1}, \cdots, \mathbf{w}_{i+j}$ be the left eigenvectors of $\mathbf{C}$, corresponding to the singular values $s_1(\mathbf{C}), \cdots, s_j(\mathbf{C})$. Then, by (A.2.2), we obtain

$$\begin{aligned} s_{i+j+1}(\mathbf{A}+\mathbf{C}) &\le \max_{\substack{\|\mathbf{v}\|_2=1 \\ \mathbf{v}\perp\mathbf{w}_1,\cdots,\mathbf{w}_{i+j}}} \|(\mathbf{A}+\mathbf{C})\mathbf{v}\|_2 \\ &\le \max_{\substack{\|\mathbf{v}\|_2=1 \\ \mathbf{v}\perp\mathbf{w}_1,\cdots,\mathbf{w}_{i+j}}} [\|\mathbf{A}\mathbf{v}\|_2 + \|\mathbf{C}\mathbf{v}\|_2] \\ &\le \max_{\substack{\|\mathbf{v}\|_2=1 \\ \mathbf{v}\perp\mathbf{w}_1,\cdots,\mathbf{w}_{i}}} \|\mathbf{A}\mathbf{v}\|_2 \\ &\quad + \max_{\substack{\{\|\mathbf{v}\|_2=1 \\ \mathbf{v}\perp\mathbf{w}_{i+1},\cdots,\mathbf{w}_{i+j}\}}} \|\mathbf{C}\mathbf{v}\|_2 \\ &= s_{i+1}(\mathbf{A}) + s_{j+1}(\mathbf{C}). \end{aligned}$$

The proof is complete.

In the language of functional analysis, the largest singular value is referred to as the operator norm of the linear operator (matrix) in a Hilbert space. The following theorem states that the norm of the product of linear transformations is not greater than the product of the norms of the linear transformations.

Theorem A.9. *Let* $\mathbf{A}$ *and* $\mathbf{C}$ *be complex matrices of order* $p \times n$ *and* $n \times m$. *We have*

$$s_1(\mathbf{AC}) \le s_1(\mathbf{A}) s_1(\mathbf{C}). \tag{A.2.6}$$

This theorem follows from the simple fact that

$$\begin{aligned} s_1(\mathbf{AC}) &= \sup_{\|\mathbf{x}\|=1} \|\mathbf{ACx}\| = \sup_{\|\mathbf{x}\|=1} \left\| \mathbf{A}\frac{\mathbf{Cx}}{\|\mathbf{Cx}\|} \right\| \|\mathbf{Cx}\| \\ &\le \sup_{\|\mathbf{y}\|=1} \|\mathbf{Ay}\| \sup_{\|\mathbf{x}\|=1} \|\mathbf{Cx}\| = s_1(\mathbf{A}) s_1(\mathbf{C}). \end{aligned}$$

There are some extensions to Theorem A.9 that are very useful in the theory of spectral analysis of large dimensional random matrices.

The first is the following due to Fan Ky [103].

Theorem A.10. *Let* $\mathbf{A}$ *and* $\mathbf{C}$ *be complex matrices of order* $p \times n$ *and* $n \times m$. *For any* $i, j \ge 0$, *we have*

$$s_{i+j+1}(\mathbf{AC}) \le s_{i+1}(\mathbf{A}) s_{j+1}(\mathbf{C}), \tag{A.2.7}$$

where when $i > \operatorname{rank}(\mathbf{A})$, *define* $s_i(\mathbf{A}) = 0$.

Proof. First we consider the case where $\mathbf{C}$ is an invertible square matrix. Then, we have

$$\begin{aligned}
s_{i+j+1}(\mathbf{AC}) &= \inf_{\mathbf{w}_1,\cdots,\mathbf{w}_{i+j}} \sup_{\substack{\mathbf{x}\perp\{\mathbf{w}_1,\cdots,\mathbf{w}_{i+j}\}\\ \|\mathbf{x}\|=1}} \|\mathbf{ACx}\| \\
&= \inf_{\mathbf{w}_1,\cdots,\mathbf{w}_{i+j}} \sup_{\substack{\mathbf{x}\perp\{(\mathbf{C}^*)\mathbf{w}_1,\cdots,(\mathbf{C}^*)\mathbf{w}_i,\mathbf{w}_{i+1},\cdots,\mathbf{w}_{i+j}\}\\ \|\mathbf{x}\|=1}} \|\mathbf{ACx}\| \\
&= \inf_{\mathbf{w}_1,\cdots,\mathbf{w}_{i+j}} \sup_{\substack{\mathbf{Cx}\perp\{\mathbf{w}_1,\cdots,\mathbf{w}_i\}\\ \mathbf{x}\perp\{\mathbf{w}_{i+1},\cdots,\mathbf{w}_{i+j}\}\\ \|\mathbf{x}\|=1}} \frac{\|\mathbf{ACx}\|}{\|\mathbf{Cx}\|}\|\mathbf{Cx}\| \\
&\le \inf_{\mathbf{w}_1,\cdots,\mathbf{w}_{i+j}} \sup_{\substack{\mathbf{y}\perp\{\mathbf{w}_1,\cdots,\mathbf{w}_i\}\\ \mathbf{x}\perp\{\mathbf{w}_{i+1},\cdots,\mathbf{w}_{i+j}\}\\ \|\mathbf{x}\|=1,\|\mathbf{y}\|=1}} \|\mathbf{Ay}\|\|\mathbf{Cx}\| \\
&= s_{i+1}(\mathbf{A})s_{j+1}(\mathbf{C}).
\end{aligned}$$

For the general case, let the singular decomposition of $\mathbf{C}$ be given by

$$\mathbf{C} = \mathbf{EDF},$$

where $\mathbf{D}$ is the $r \times r$ diagonal matrix of positive singular values of $\mathbf{C}$ and $\mathbf{E}$ $(n \times r)$ and $\mathbf{E}$ $(r \times m)$ are such that $\mathbf{E}^*\mathbf{E} = \mathbf{FF}^* = \mathbf{I}_r$. Then, by what has been proved,

$$\begin{aligned}
s_{i+j+1}(\mathbf{AC}) &= s_{i+j+1}(\mathbf{AED}) \\
&\le s_{i+1}(\mathbf{AE})s_{j+1}(\mathbf{D}) \\
&\le s_{i+1}(\mathbf{A})s_{j+1}(\mathbf{C}).
\end{aligned}$$

Here, in the last step, we have used the simple fact that

$$\begin{aligned}
s_{i+1}(\mathbf{AE}) &= \inf_{\mathbf{w}_1,\cdots,\mathbf{w}_i} \sup_{\substack{\mathbf{x}\perp\{\mathbf{w}_1,\cdots,\mathbf{w}_i\}\\ \|\mathbf{x}\|=1}} \|\mathbf{x}^*\mathbf{AE}\| \\
&\le \inf_{\mathbf{w}_1,\cdots,\mathbf{w}_i} \sup_{\substack{\mathbf{x}\perp\{\mathbf{w}_1,\cdots,\mathbf{w}_i\}\\ \|\mathbf{x}\|=1}} \|\mathbf{x}^*\mathbf{A}\| = s_{i+1}(\mathbf{A}). \qquad \text{(A.2.8)}
\end{aligned}$$

To prove another extension, we need the following lemma.

Lemma A.11. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix with singular values* $s_i(\mathbf{A})$, $i = 1, 2, \cdots, p = \min(m, n)$, *arranged in decreasing order. Then, for any integer* k $(1 \le k \le p)$,

$$\sum_{i=1}^{k} s_i(\mathbf{A}) = \sup_{\{\mathbf{E}^*\mathbf{E}=\mathbf{F}^*\mathbf{F}=\mathbf{I}_k\}} |\mathrm{tr}(\mathbf{E}^*\mathbf{AF})|, \qquad \text{(A.2.9)}$$

where the orders of $\mathbf{E}$ *are* $m \times k$ *and those of* $\mathbf{F}$ *are* $n \times k$.

Proof of Lemma A.11. By Theorem A.7, if we choose $\mathbf{E} = (\mathbf{u}_1, \cdots, \mathbf{u}_k)$ and $\mathbf{F} = (\mathbf{v}_1, \cdots, \mathbf{v}_k)$, then we have

$$\operatorname{tr}(\mathbf{E}^*\mathbf{A}\mathbf{F}) = \sum_{i=1}^{k} s_i(\mathbf{A}).$$

Therefore, to finish the proof of (A.2.9), one needs only to show that

$$|\operatorname{tr}(\mathbf{E}^*\mathbf{A}\mathbf{F})| \leq \sum_{i=1}^{k} s_i(\mathbf{A})$$

for any $\mathbf{E}^*\mathbf{E} = \mathbf{F}^*\mathbf{F} = \mathbf{I}_k$.

In fact, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
|\operatorname{tr}(\mathbf{E}^*\mathbf{A}\mathbf{F})| &= \left|\sum_{i=1}^{p} s_i(\mathbf{A})\mathbf{v}_i^*\mathbf{F}\mathbf{E}^*\mathbf{u}_i\right| \\
&\leq \left(\sum_{i=1}^{p} s_i(\mathbf{A})\mathbf{v}_i^*\mathbf{F}\mathbf{F}^*\mathbf{v}_i\right)^{1/2} \left(\sum_{i=1}^{p} s_i(\mathbf{A})\mathbf{u}_i^*\mathbf{E}\mathbf{E}^*\mathbf{u}_i\right)^{1/2}.
\end{aligned}$$

Because $\mathbf{F}^*\mathbf{F} = \mathbf{I}_k$ and $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ forms an orthonormal basis in $\mathbb{C}^n$, we have

$$0 \leq \mathbf{v}_i^*\mathbf{F}\mathbf{F}^*\mathbf{v}_i \leq 1 \tag{A.2.10}$$

and

$$\sum_{i=1}^{n} \mathbf{v}_i^*\mathbf{F}\mathbf{F}^*\mathbf{v}_i = k. \tag{A.2.11}$$

From these two facts, it follows that

$$\sum_{i=1}^{p} s_i(\mathbf{A})\mathbf{v}_i^*\mathbf{F}\mathbf{F}^*\mathbf{v}_i \leq \sum_{i=1}^{k} s_i(\mathbf{A}).$$

Similarly, we have

$$\sum_{i=1}^{p} s_i(\mathbf{A})\mathbf{u}_i^*\mathbf{E}\mathbf{E}^*\mathbf{u}_i \leq \sum_{i=1}^{k} s_i(\mathbf{A}).$$

The proof of the lemma is complete.

In (A.2.9), letting $k = m = n$ and taking $\mathbf{E} = \mathbf{F} = \mathbf{I}_n$, we immediately get the following corollary.

Corollary A.12. *For any $n \times n$ complex matrix $\mathbf{A}$,*

$$|\operatorname{tr}(\mathbf{A})| \leq \sum_{j=1}^{n} s_j(\mathbf{A}).$$

Similar to the proof of Lemma A.11, the conclusion of Corollary A.12 can be extended to the following theorem.

Theorem A.13. *Let* $\mathbf{A} = (a_{ij})$ *be a complex matrix of order* n *and* f *be an increasing and convex function. Then we have*

$$\sum_{j=1}^{n} f(|a_{jj}|) \leq \sum_{j=1}^{n} f(s_j(\mathbf{A})). \tag{A.2.12}$$

Note that when $\mathbf{A}$ is Hermitian, $s_j(\mathbf{A})$ can be replaced by eigenvalues and f need not be increasing.

Proof. By singular-value decomposition, we can write

$$a_{jj} = \sum_{k=1}^{n} s_k(\mathbf{A}) u_{kj} v_{kj},$$

where u_{kj} and v_{kj} satisfy

$$\sum_{j=1}^{n} |u_{kj}|^2 = \sum_{j=1}^{n} |v_{kj}|^2 = 1.$$

By applying the Jensen inequality, we obtain

$$\begin{aligned}\sum_{j=1}^{n} f(|a_{jj}|) &\leq \sum_{j=1}^{n} f\left(\frac{1}{2}\sum_{k=1}^{n} s_k(\mathbf{A})\big(|u_{kj}|^2 + |v_{kj}|^2\big)\right) \\ &\leq \sum_{j=1}^{n}\left(\frac{1}{2}\sum_{k=1}^{n} f(s_k(\mathbf{A}))\big(|u_{kj}|^2 + |v_{kj}|^2\big)\right) = \sum_{k=1}^{n} f(s_k(\mathbf{A})).\end{aligned}$$

This completes the proof of the theorem.

The extension to Theorem A.9 is stated as follows.

Theorem A.14. *Let* $\mathbf{A}$ *and* $\mathbf{C}$ *be complex matrices of order* $p \times n$ *and* $n \times m$. *We have*

$$\sum_{j=1}^{k} s_j(\mathbf{AC}) \leq \sum_{j=1}^{k} s_j(\mathbf{A}) s_j(\mathbf{C}). \tag{A.2.13}$$

Before proving this theorem, we first prove an important special case of Theorem A.14 due to von Neumann [219].

Theorem A.15. *Let* $\mathbf{A}$ *and* $\mathbf{C}$ *be complex matrices of order* $p \times n$. *We have*

$$\sum_{j=1}^{p\wedge n} s_j(\mathbf{A}^*\mathbf{C}) \leq \sum_{j=1}^{p\wedge n} s_j(\mathbf{A}) s_j(\mathbf{C}), \tag{A.2.14}$$

where $p \wedge n = \min\{p, n\}$*. The following immediate consequence of the inequality above is the famous von Neumann inequality:*

$$|\mathrm{tr}(\mathbf{A}^*\mathbf{C})| \le \sum_{j=1}^{p\wedge n} s_j(\mathbf{A})s_j(\mathbf{C}).$$

Proof. Without loss of generality, we can assume that $p \le n$. Also, without change of the singular values of the matrices of $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{A}^*\mathbf{C}$, we can assume the two matrices are

$$\mathbf{A} = \begin{pmatrix} s_1(\mathbf{A}) & \cdots & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & s_p(\mathbf{A}) & \cdots & 0 \end{pmatrix} \mathbf{U}$$

and

$$\mathbf{C} = \mathbf{V}^* \begin{pmatrix} s_1(\mathbf{C}) & \cdots & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & s_p(\mathbf{C}) & \cdots & 0 \end{pmatrix},$$

where $\mathbf{U}$ $(n \times n)$ and $\mathbf{V}$ $(p \times p)$ are unitary matrices. In the expression below, $\mathbf{E}$ and $\mathbf{F}$ are $n \times n$ unitary. Write $\mathbf{F}\mathbf{E}^*\mathbf{U}^* = \mathbf{Q} = (q_{ij})$, which is an $n \times n$ unitary matrix, and $\mathbf{V}^* = (v_{ij})$ $(p \times p)$. Then, by Lemma A.11, we have

$$\begin{aligned} \sum_{j=1}^{p} s_j(\mathbf{A}^*\mathbf{C}) &= \sup_{\mathbf{E},\mathbf{F}} |\mathrm{tr}(\mathbf{E}^*\mathbf{A}^*\mathbf{C}\mathbf{F})| \\ &\le \sup_{\mathbf{Q}} \left| \sum_{i=1}^{p}\sum_{j=1}^{p} s_i(\mathbf{A})s_j(\mathbf{C})q_{ji}v_{ij} \right| \\ &\le \sup_{\mathbf{Q}} \left| \sum_{i=1}^{p}\sum_{j=1}^{p} s_i(\mathbf{A})s_j(\mathbf{C})|q_{ij}|^2 \right|^{1/2} \sup_{\mathbf{V}} \left| \sum_{i=1}^{p}\sum_{j=1}^{p} s_i(\mathbf{A})s_j(\mathbf{C})|v_{ji}|^2 \right|^{1/2}. \end{aligned}$$

Noting that both $\mathbf{Q}$ and $\mathbf{V}$ are unitary matrices, we have the following relations:

$$\sum_{i=1}^{p} |q_{ij}|^2 \le 1, \quad \sum_{j=1}^{p} |q_{ij}|^2 \le 1,$$
$$\sum_{i=1}^{p} |v_{ij}|^2 = 1, \quad \sum_{j=1}^{p} |v_{ij}|^2 = 1.$$

By linear programming, one can prove that

$$\sup_{\mathbf{Q}} \left| \sum_{i=1}^{p} \sum_{j=1}^{p} s_i(\mathbf{A}) s_j(\mathbf{C}) |q_{ij}|^2 \right| \le \sum_{i=1}^{p} s_i(\mathbf{A}) s_i(\mathbf{C})$$

and

$$\sup_{\mathbf{W}} \left| \sum_{i=1}^{p} \sum_{j=1}^{p} s_i(\mathbf{A}) s_j(\mathbf{C}) |w_{ji}|^2 \right| \le \sum_{i=1}^{p} s_i(\mathbf{A}) s_i(\mathbf{C}).$$

The proof of the theorem is then complete.

To prove Theorem A.14, we also need the following lemma, which is a trivial consequence of (A.2.8).

Lemma A.16. *Let* $\mathbf{A}$ *be a* $p \times n$ *complex matrix and* $\mathbf{U}$ *be an* $n \times m$ *complex matrix with* $\mathbf{U}^*\mathbf{U} = \mathbf{I}_m$. *Then, for any* $k \le p$,

$$\sum_{j=1}^{k} s_j(\mathbf{AU}) \le \sum_{j=1}^{k} s_j(\mathbf{A}).$$

Proof of Theorem A.14. By Lemma A.11, Theorem A.15, and Lemma A.16,

$$\begin{aligned}
\sum_{j=1}^{k} s_j(\mathbf{AC}) &= \sup_{\mathbf{E}^*\mathbf{E}=\mathbf{F}^*\mathbf{F}=\mathbf{I}_k} |\mathrm{tr}(\mathbf{E}^*\mathbf{A}^*\mathbf{CF})| \\
&\le \sup_{\mathbf{E}^*\mathbf{E}=\mathbf{F}^*\mathbf{F}=\mathbf{I}_k} \sum_{j=1}^{k} s_j(\mathbf{AE}) s_j(\mathbf{CF}) \\
&= \sup_{\mathbf{E}^*\mathbf{E}=\mathbf{F}^*\mathbf{F}=\mathbf{I}_k} \sum_{i=1}^{k-1} [s_i(\mathbf{AE}) - s_{i+1}(\mathbf{AE})] \sum_{j=1}^{i} s_j(\mathbf{CF}) \\
&\qquad + s_k(\mathbf{AE}) \sum_{j=1}^{k} s_j(\mathbf{CF}) \\
&\le \sup_{\mathbf{E}^*\mathbf{E}=\mathbf{F}^*\mathbf{F}=\mathbf{I}_k} \sum_{i=1}^{k-1} [s_i(\mathbf{AE}) - s_{i+1}(\mathbf{AE})] \sum_{j=1}^{i} s_j(\mathbf{C}) \\
&\qquad + s_k(\mathbf{AE}) \sum_{j=1}^{k} s_j(\mathbf{C}) \\
&= \sup_{\mathbf{E}^*\mathbf{E}=\mathbf{I}_k} \sum_{j=1}^{k} s_j(\mathbf{AE}) s_j(\mathbf{C}) \\
&\le \sum_{j=1}^{k} s_j(\mathbf{A}) s_j(\mathbf{C}).
\end{aligned}$$

Here, the last inequality follows by arguments similar to those of the proof of removing $\mathbf{F}$. The proof of the theorem is complete.

A.3 Hadamard Product and Odot Product

Definition. Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be two $m \times n$ matrices. Then, the $m \times n$ matrix $(a_{ij}b_{ij})$ is called the Hadamard product and denoted by $\mathbf{A} \circ \mathbf{B}$.

In this section, we shall quote some results useful for this book. For more details about Hadamard products, the reader is referred to the book by Horn and Johnson [154].

Lemma A.17. *Let $\mathbf{x} = (x_1, \cdots, x_n)'$ and $\mathbf{y} = (y_1, \cdots, y_n)'$ be two independent random vectors with mean zero and covariance matrices $\Sigma_{\mathbf{x}}$ and $\Sigma_{\mathbf{y}}$, respectively. Then, the covariance matrix of $\mathbf{x} \circ \mathbf{y}$ is $\Sigma_{\mathbf{x}} \circ \Sigma_{\mathbf{y}}$. In fact,*

$$\Sigma_{\mathbf{x}\circ\mathbf{y}} = \mathrm{E}(x_i y_i \overline{x}_j \overline{y}_j) = (\mathrm{E}(x_i \overline{x}_j)\mathrm{E}(y_j \overline{y}_j)) = \Sigma_{\mathbf{x}} \circ \Sigma_{\mathbf{y}}. \tag{A.3.1}$$

By (A.3.1), it is easy to derive the Schur product theorem.

Theorem A.18. *If $\mathbf{A}$ and $\mathbf{B}$ are two $n \times n$ nonnegative definite matrices, then so is $\mathbf{A} \circ \mathbf{B}$. If $\mathbf{A}$ is positive definite and $\mathbf{B}$ nonnegative definite with no zero diagonal elements, then $\mathbf{A} \circ \mathbf{B}$ is positive definite. In particular, when the two matrices are both positive definite, then so is $\mathbf{A} \circ \mathbf{B}$.*

Proof. Let $\mathbf{A}$ and $\mathbf{B}$ be the covariance matrices of the random vectors $\mathbf{x}$ and $\mathbf{y}$. Then $\mathbf{A} \circ \mathbf{B}$ is the covariance matrix of $\mathbf{x} \circ \mathbf{y}$. Therefore, $\mathbf{A} \circ \mathbf{B}$ is nonnegative definite.

Suppose that $\mathbf{A}$ is positive definite and $\mathbf{B}$ is nonnegative definite with no zero diagonal elements. Let $\mathbf{x}$ be distributed as $N(\mathbf{O}, \mathbf{A})$ and let $\mathbf{y}$ be distributed as $N(\mathbf{O}, \mathbf{B})$ and independent of $\mathbf{x}$. Since the distribution of $\mathbf{x}$ is absolutely continuous and $\mathbf{y}$ has no zero entries, we conclude that the distribution of $\mathbf{x} \circ \mathbf{y}$ is absolutely continuous. Therefore, its covariance matrix $\mathbf{A} \circ \mathbf{B}$ is positive definite.

Next, we introduce an inequality concerning singular values of Hadamard products due to Fan [104].

Theorem A.19. *Let $\mathbf{A}$ and $\mathbf{B}$ be two $m \times n$ matrices with singular values $s_i(\mathbf{A})$ and $s_i(\mathbf{B})$, $i = 1, 2, \cdots, p = \min(m, n)$, arranged in decreasing order. Denote the singular values of $\mathbf{A} \circ \mathbf{B}$ by $s_i(\mathbf{A} \circ \mathbf{B})$, $i = 1, 2, \cdots, p$. Then, for any integer k $(1 \le k \le p)$,*

$$\sum_{i=1}^{k} s_i(\mathbf{A} \circ \mathbf{B}) \le \sum_{i=1}^{k} s_i(\mathbf{A}) s_i(\mathbf{B}). \tag{A.3.2}$$

Proof of Theorem A.19. Suppose that the singular decompositions of $\mathbf{A}$ and $\mathbf{B}$ are given by

$$\mathbf{A} = \sum_{i=1}^{p} s_i(\mathbf{A})\mathbf{u}_i\mathbf{v}_i^*$$

and

$$\mathbf{B} = \sum_{i=1}^{p} s_i(\mathbf{B})\mathbf{x}_i\mathbf{y}_i^*.$$

Then, by Lemma A.11, we have

$$\begin{aligned}
&\sum_{i=1}^{k} s_i(\mathbf{A}\circ\mathbf{B}) \\
&= \sup_{\{\mathbf{E}^*\mathbf{E}=\mathbf{F}^*\mathbf{F}=\mathbf{I}_k\}} |\mathrm{tr}(\mathbf{E}^*(\mathbf{A}\circ\mathbf{B})\mathbf{F})| \\
&\le \sum_{i=1}^{p}\sum_{j=1}^{p} s_i(\mathbf{A})s_j(\mathbf{B}) \left|\mathrm{tr}\left(\mathbf{E}^*((\mathbf{u}_i\mathbf{v}_i^*)\circ(\mathbf{x}_j\mathbf{y}_j^*))\mathbf{F}\right)\right| \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p} s_i(\mathbf{A})s_j(\mathbf{B}) \left|(\mathbf{v}_i\circ\mathbf{y}_j)^*\mathbf{F}\mathbf{E}^*(\mathbf{u}_i\circ\mathbf{x}_j)\right| \\
&\le \left(\sum_{i=1}^{p}\sum_{j=1}^{p} s_i(\mathbf{A})s_j(\mathbf{B})(\mathbf{v}_i\circ\mathbf{y}_j)^*\mathbf{F}\mathbf{F}^*(\mathbf{v}_i\circ\mathbf{y}_j)\right. \\
&\qquad \left.\sum_{i=1}^{p}\sum_{j=1}^{p} s_i(\mathbf{A})s_j(\mathbf{B})(\mathbf{u}_i\circ\mathbf{x}_j)^*\mathbf{E}\mathbf{E}^*(\mathbf{u}_i\circ\mathbf{x}_j)\right)^{1/2}.
\end{aligned}$$

Thus, to finish the proof of Theorem A.19, it is sufficient to show that

$$\sum_{i=1}^{p}\sum_{j=1}^{p} s_i(\mathbf{A})s_j(\mathbf{B})(\mathbf{v}_i\circ\mathbf{y}_j)^*\mathbf{F}\mathbf{F}^*(\mathbf{v}_i\circ\mathbf{y}_j) \le \sum_{i=1}^{k} s_i(\mathbf{A})s_i(\mathbf{B}).$$

This inequality then follows easily from the following observations:

$$0 \le (\mathbf{v}_i\circ\mathbf{y}_j)^*\mathbf{F}\mathbf{F}^*(\mathbf{v}_i\circ\mathbf{y}_j) \le 1,$$

$$\sum_{i=1}^{m}(\mathbf{v}_i\circ\mathbf{y}_j)^*\mathbf{F}\mathbf{F}^*(\mathbf{v}_i\circ\mathbf{y}_j) \le 1,$$

$$\sum_{j=1}^{n}(\mathbf{v}_i\circ\mathbf{y}_j)^*\mathbf{F}\mathbf{F}^*(\mathbf{v}_i\circ\mathbf{y}_j) \le 1,$$

and

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(\mathbf{v}_i \circ \mathbf{y}_j)^*\mathbf{F}\mathbf{F}^*(\mathbf{v}_i \circ \mathbf{y}_j) = k.$$

Corollary A.20. *Let* $\mathbf{A}_1, \cdots, \mathbf{A}_\ell$ *be* ℓ $m \times n$ *matrices whose singular values are denoted by* $s_i(\mathbf{A}_1), \cdots, s_i(\mathbf{A}_\ell)$, $i = 1, 2, \cdots, p = \min(m, n)$, *arranged in decreasing order. Denote the singular values of* $\mathbf{A}_1 \circ \cdots \circ \mathbf{A}_\ell$ *by* $s_i(\mathbf{A}_1 \circ \cdots \circ \mathbf{A}_\ell)$, $i = 1, 2, \cdots, p$. *Then, for any integer* k $(1 \le k \le p)$,

$$\sum_{i=1}^{k} s_i(\mathbf{A}_1 \circ \cdots \circ \mathbf{A}_\ell) \le \sum_{i=1}^{k} s_i(\mathbf{A}_1) \cdots s_i(\mathbf{A}_\ell). \tag{A.3.3}$$

Proof. When $\ell = 2$, the conclusion is already proved in Theorem A.19. Suppose that (A.3.3) is true for ℓ. Then, by Theorem A.19 and the induction hypothesis, we have

$$\begin{aligned}
&\sum_{i=1}^{k} s_i(\mathbf{A}_1 \circ \cdots \circ \mathbf{A}_{\ell+1}) \\
&\le \sum_{i=1}^{k} s_i(\mathbf{A}_1 \circ \cdots \circ \mathbf{A}_\ell) s_i(\mathbf{A}_{\ell+1}) \\
&= \sum_{j=1}^{k-1}\sum_{i=1}^{j} s_i(\mathbf{A}_1 \circ \cdots \circ \mathbf{A}_\ell)[s_j(\mathbf{A}_{\ell+1}) - s_{j+1}(\mathbf{A}_{\ell+1})] \\
&\qquad + \sum_{i=1}^{k} s_i(\mathbf{A}_1 \circ \cdots \circ \mathbf{A}_\ell) s_k(\mathbf{A}_{\ell+1}) \\
&\le \sum_{j=1}^{k-1}\sum_{i=1}^{j} s_i(\mathbf{A}_1) \cdots s_i(\mathbf{A}_\ell)[s_j(\mathbf{A}_{\ell+1}) - s_{j+1}(\mathbf{A}_{\ell+1})] \\
&\qquad + \sum_{i=1}^{k} s_i(\mathbf{A}_1) \cdots s_i(\mathbf{A}_\ell) s_k(\mathbf{A}_{\ell+1}) \\
&= \sum_{i=1}^{k} s_i(\mathbf{A}_1) \cdots s_i(\mathbf{A}_{\ell+1}).
\end{aligned}$$

This completes the proof of Corollary A.20.

Taking $k = 1$ in the corollary above, we immediately obtain the following norm inequality for Hadamard products.

Corollary A.21. *Let* $\mathbf{A}_1, \cdots, \mathbf{A}_\ell$ *be* ℓ $m \times n$ *matrices. We have*

$$\|\mathbf{A}_1 \circ \cdots \circ \mathbf{A}_\ell\| \le \|\mathbf{A}_1\| \cdots \|\mathbf{A}_\ell\|. \tag{A.3.4}$$

Note that the singular values of a Hermitian matrix are the absolute values of its eigenvalues. Applying Corollary A.20, we obtain the following corollary.

Corollary A.22. *Suppose that* $\mathbf{A}_j$, $j = 1, 2, \cdots, \ell$, *are* ℓ $p \times p$ *Hermitian matrices whose eigenvalues are bounded by* M_j*; i.e.,* $|\lambda_i(\mathbf{A}_j)| \leq M_j$, $i = 1, 2, \cdots, p$, $j = 1, 2, \cdots, \ell$. *Then,*

$$|\mathrm{tr}\,(\mathbf{A}_1 \circ \cdots \circ \mathbf{A}_\ell)| \leq pM_1 \cdots M_\ell. \tag{A.3.5}$$

Definition. Let $\mathbf{T}_j = (t_{i\ell}^{(j)})$, $j = 1, 2, \cdots, k$, be k complex matrices with dimensions $n_j \times n_{j+1}$, respectively. Define the **Odot product** of the k matrices by

$$\mathbf{T}_1 \odot \cdots \odot \mathbf{T}_k = \left(\sum t_{ai_2}^{(1)} t_{i_2 i_3}^{(2)} \cdots t_{i_{k-1} i_k}^{(k-1)} t_{i_k b}^{(k)} \right), \tag{A.3.6}$$

where the summation runs for $i_j = 1, 2, \cdots, n_j$, $j = 2, \cdots, k$, subject to restrictions $i_3 \neq a$, $i_4 \neq i_2$, $\cdots$, $i_k \neq i_{k-2}$, and $i_{k-1} \neq b$. If $k = 2$, we require $a \neq b$, namely, the diagonal elements of $\mathbf{T}_1 \odot \mathbf{T}_2$ are zero.

The dimensions of the odot product are $n_1 \times n_{k+1}$. The following theorem will be needed in establishing the limit of smallest eigenvalues of large sample covariance matrices.

Theorem A.23. *Let* $\mathbf{T}_j = (t_{i\ell}^{(j)})$, $j = 1, 2, \cdots, k$, *be* k *complex matrices with dimensions* $n_j \times n_{j+1}$, *respectively. Then, we have*

$$\|\mathbf{T}_1 \odot \cdots \odot \mathbf{T}_k\| \leq 2^{k-1} \|\mathbf{T}_1\| \cdots \|\mathbf{T}_k\|.$$

Proof. When $k = 1$, Theorem A.23 is trivially true. When $k = 2$, Theorem A.23 follows from the fact that $\mathbf{T}_1 \odot \mathbf{T}_2 = \mathbf{T}_1 \mathbf{T}_2 - \mathrm{diag}(\mathbf{T}_1 \mathbf{T}_2)$, where $\mathrm{diag}(\mathbf{A})$ is the diagonal matrix of the diagonal elements of the matrix $\mathbf{A}$. Let $k > 2$. Note that

$$\begin{aligned} &\mathbf{T}_1 \odot \cdots \odot \mathbf{T}_k \\ &= \mathbf{T}_1 (\mathbf{T}_2 \odot \cdots \odot \mathbf{T}_k) - \mathrm{diag}(\mathbf{T}_1 \mathbf{T}_2) (\mathbf{T}_3 \odot \cdots \odot \mathbf{T}_k) \\ &\quad + (\mathbf{T}_1 \diamond \mathbf{T}_2' \diamond \mathbf{T}_3) \odot \mathbf{T}_4 \odot \cdots \odot \mathbf{T}_k, \end{aligned}$$

where $\mathbf{T}_1 \diamond \mathbf{T}_2' \diamond \mathbf{T}_3 = \left(t_{ab}^{(1)} t_{ba}^{(2)} t_{ab}^{(3)} \right)$ with dimensions $n_1 \times n_4$. Here, the (a, b) entry of the matrix $\mathbf{T}_1 \diamond \mathbf{T}_2' \diamond \mathbf{T}_3$ is zero if $b > n_2$ or $a > n_3$. By Lemma A.16 and Corollary A.21, we have $\|\mathbf{T}_1 \diamond \mathbf{T}_2' \diamond \mathbf{T}_3\| \leq \|\mathbf{T}_1\| \|\mathbf{T}_2\| \|\mathbf{T}_3\|$.

Then, the conclusion of the theorem follows by induction.

A.4 Extensions of Singular-Value Inequalities

In this section, we shall extend the concepts of vectors and matrices to multiple vectors and matrices, especially graph-associated multiple matrices, which will be used in deriving the LSD of products of random matrices.

A.4.1 Definitions and Properties

Definition A.24. A collection of ordered numbers

$$\boldsymbol{\alpha}(\mathbf{i}) = \{a_{i_1,\cdots,i_t};\ i_1 = 1, \cdots, n_1, \cdots, i_t = 1, \cdots, n_t\}$$

is called a multiple vector (MV) with dimensions $\mathbf{n} = \{n_1, \cdots, n_t\}$, where $\mathbf{i} = \{i_1, \cdots, i_t\}$ and $t \geq 1$ is an integer.

Its norm is defined by

$$\|\boldsymbol{\alpha}\|^2 = \sum_{\mathbf{i}} |a_{\mathbf{i}}|^2.$$

Definition A.25. A multiple matrix (MM) is defined as a collection of ordered numbers

$$\mathbf{A} = \{a_{\mathbf{i};\mathbf{j}}, i_1 = 1, 2, \cdots, m_1, \cdots, i_s = 1, 2, \cdots, m_s,$$

$$\text{and } j_1 = 1, 2, \cdots, n_1, \cdots, j_t = 1, 2, \cdots, n_t\},$$

where $\mathbf{i} = \{i_1, \cdots, i_s\}$ and $\mathbf{j} = \{j_1, \cdots, j_t\}$. The integer vectors $\mathbf{m} = \{m_1, \cdots, m_s\}$ and $\mathbf{n} = \{n_1, \cdots, n_t\}$ are called its dimensions.

Similarly, its norm is defined as

$$\|\mathbf{A}\| = \sup \left| \sum_{\mathbf{j}} a_{\mathbf{i};\mathbf{j}} g_{\mathbf{i}} h_{\mathbf{j}} \right|,$$

where the supremum is taken subject to $\sum_{\mathbf{i}} |g_{\mathbf{i}}|^2 = 1$ and $\sum_{\mathbf{j}} |h_{\mathbf{j}}|^2 = 1$.

The domains of g and h are both compact sets. Thus, the supremum in the definition of $\|\mathbf{A}\|$ is attainable. By choosing

$$h_{\mathbf{j}} = \frac{\sum_{\mathbf{i}} a_{\mathbf{i};\mathbf{j}} g_{\mathbf{i}}}{\sqrt{\sum_{\mathbf{v}} \left|\sum_{\mathbf{u}} a_{\mathbf{u};\mathbf{v}} g_{\mathbf{u}}\right|^2}}$$

or

$$g_{\mathbf{i}} = \frac{\sum_{\mathbf{j}} a_{\mathbf{i};\mathbf{j}} h_{\mathbf{j}}}{\sqrt{\sum_{\mathbf{u}} \left|\sum_{\mathbf{v}} a_{\mathbf{u};\mathbf{v}} h_{\mathbf{v}}\right|^2}},$$

we know that the definition of $\|\mathbf{A}\|$ is equivalent to

$$\|\mathbf{A}\|^2 = \sup_{\|g\|=1} \sum_{\mathbf{j}} \left| \sum_{\mathbf{i}} a_{\mathbf{i};\mathbf{j}} g_{\mathbf{i}} \right|^2 \tag{A.4.1}$$

$$= \sup_{\|h\|=1} \sum_{\mathbf{i}} \left| \sum_{\mathbf{j}} a_{\mathbf{i};\mathbf{j}} h_{\mathbf{j}} \right|^2. \tag{A.4.2}$$

We define a product of two MMs as follows.

Definition A.26. If an MM $\mathbf{B} = \{b_{\mathbf{j}';\boldsymbol{\ell}}\}$, where $\mathbf{j}' = \{j'_1, \cdots, j_{t'_1}) \subset \mathbf{j}$, $\boldsymbol{\ell} = (\ell_1, \cdots, \ell_u)$, and $\ell_1 = 1, 2, \cdots, p_1, \cdots, \ell_u = 1, 2, \cdots, p_u$, then the product of $\mathbf{A}$ and $\mathbf{B}$ is defined by

$$\mathbf{A} \cdot \mathbf{B}_{\mathbf{i},\tilde{\boldsymbol{\ell}}} = \sum_{\mathbf{j}'} a_{\mathbf{i};\mathbf{j}} b_{\mathbf{j}';\boldsymbol{\ell}},$$

where $\boldsymbol{\ell} = \{\ell_1, \cdots, \ell_u\}$ and $\tilde{\boldsymbol{\ell}} = \boldsymbol{\ell} \cup \mathbf{j} \backslash \mathbf{j}'$. The product is then an MM of dimensions $\mathbf{m} \times \mathbf{p}$, where $\mathbf{p}$ contains all p's as well as those n's corresponding to the j-indices not contained in $\mathbf{j}'$.

Theorem A.27. *Using the notation defined above, we have*

$$\|\mathbf{A} \cdot \mathbf{B}\| \le \|\mathbf{A}\| \cdot \|\mathbf{B}\|. \tag{A.4.3}$$

Proof. Using definition (A.4.1), we have

$$\begin{aligned}
\|\mathbf{A} \cdot \mathbf{B}\|^2 &= \sup_{\|h\|=1} \left(\sum_{\mathbf{i}} \left| \sum_{\mathbf{j};\boldsymbol{\ell}} a_{\mathbf{i};\mathbf{j}} b_{\mathbf{j}';\boldsymbol{\ell}} h_{\tilde{\boldsymbol{\ell}}} \right|^2 \right) \\
&\le \|\mathbf{A}\|^2 \sup_{\|h\|=1} \left(\sum_{\mathbf{j}} \left| \sum_{\boldsymbol{\ell}} b_{\mathbf{j}';\boldsymbol{\ell}} h_{\tilde{\boldsymbol{\ell}}} \right|^2 \right) \\
&\le \|\mathbf{A}\|^2 \sup_{\|h\|=1} \left(\sum_{\mathbf{j}''} \frac{\sum_{\mathbf{j}'} \left| \sum_{\boldsymbol{\ell}} b_{\mathbf{j}';\boldsymbol{\ell}} h_{\tilde{\boldsymbol{\ell}}} \right|^2}{\sum_{\boldsymbol{\ell}} |h_{\tilde{\boldsymbol{\ell}}}|^2} \sum_{\boldsymbol{\ell}} |h_{\tilde{\boldsymbol{\ell}}}|^2 \right) \\
&\le \|\mathbf{A}\|^2 \cdot \|\mathbf{B}\|^2 \sup_{\|h\|=1} \sum_{\mathbf{j}''} \sum_{\boldsymbol{\ell}} |h_{\tilde{\boldsymbol{\ell}}}|^2 \\
&= \|\mathbf{A}\|^2 \cdot \|\mathbf{B}\|^2,
\end{aligned}$$

where $\mathbf{j}'' = \mathbf{j} \backslash \mathbf{j}'$.

Conclusion (A.4.3) then follows.

A.4.2 Graph-Associated Multiple Matrices

Now, we describe a kind of graph-associated MM as follows.

Suppose that $G = (V, E, F)$ is a directional graph where $V = V_1 + V_2 + V_3$, $V_1 = \{1, \cdots, s\}$, $V_2 = \{s+1, \cdots, t_1\}$, $V_3 = \{t_1 + 1, \cdots, t\}$, $E = \{e_1, \cdots, e_k\}$, and $F = (f_i, f_e)$ is a function from E into $V \times V_3$; i.e., for each edge, its initial vertex can be in V_1, V_2, or V_3 and its end vertex can only be in V_3. We assume that the graph G is V_1-based; i.e., each vertex in V_1 is an initial vertex of at least one edge and each vertex in V_3 is an end vertex of at least one edge that starts from V_1. Between V_2 and V_3 there may be some edges or no edges at all. We call the edges initiated from V_1, V_2, or V_3 the **first**, **second**, or **third type** edges, respectively. Furthermore, assume that there are k matrices $\mathbf{T}^{(j)} = (t_{uv}^{(j)})$ of dimensions $m_j \times n_j$, corresponding to the k edges, subject to the consistent dimension restriction (that is, coincident vertices corresponding to equal dimensions); e.g., if e_i and e_j have a coincident initial vertex, then $m_i = m_j$, if they have a coincident end vertex, then $n_i = n_j$, and if the initial vertex of e_i coincides with the end vertex of e_j, then $m_i = n_j$, etc. In other words, each vertex corresponds to a dimension. In what follows, the dimension corresponding to the vertex j is denoted by p_j.

Without loss of generality, assume that the first k_1 edges of the graph G are of the first type and the next k_2 edges are of the second type. Then the last $k - k_1 - k_2$ are of the third type. Define an MM $T(G)$ by

$$\mathbf{T}(G)_{\mathbf{u};\mathbf{v}} = \prod_{j=1}^{k_1} t^{(j)}_{u_{f_i(e_j)},v_{f_e(e_j)}} \prod_{j=k_1+1}^{k} t^{(j)}_{v_{f_i(e_j)},v_{f_e(e_j)}},$$

where $\mathbf{u} = \{u_1, u_2, \cdots, u_s\}$, $\mathbf{v}_1 = \{v_{s+1}, v_{s+2}, \cdots, v_{t_1}\}$, $\mathbf{v}_2 = \{v_{t_1+1}, \cdots, v_t\}$, and $\mathbf{v} = \{\mathbf{v}_1, \mathbf{v}_2\}$.

Theorem A.28. *Using the notation defined above, let* $\mathbf{A} = \{a_{\mathbf{i};(\mathbf{u},\mathbf{v}_1)}\}$ *be an MM. Define a product MM of the MM* $\mathbf{A}$ *and the quasi-MM* $\mathbf{T}(G)$ *as given by*

$$\mathbf{A} \cdot \mathbf{T}(G) = \left\{ \sum_{\mathbf{u}} a_{\mathbf{i};(\mathbf{u},\mathbf{v}_1)} T(G)_{\mathbf{u};\mathbf{v}} \right\}.$$

Then, we have

$$\|\mathbf{A} \cdot \mathbf{T}(G)\| \le \|\mathbf{A}\| \prod_{j=1}^{k} \|\mathbf{T}^{(j)}\|. \tag{A.4.4}$$

Proof. Using definition (A.4.1) and noting that $|t^{(j)}_{u,v}| \le \|\mathbf{T}^{(j)}\|$, we have

$$\begin{aligned}\|\mathbf{A} \cdot \mathbf{T}(G)\|^2 &= \sup_{\|g\|=1} \sum_{\mathbf{v}} \left| \sum_{\mathbf{i}} g_{\mathbf{i}} \sum_{\mathbf{u}} a_{\mathbf{i};(\mathbf{u},\mathbf{v}_1)} T(G)_{\mathbf{u};\mathbf{v}} \right|^2 \\ &\le \prod_{j=k_1+1}^{k} \|\mathbf{T}^{(j)}\|^2 \sup_{\|g\|=1} \sum_{\mathbf{v}} \left| \sum_{\mathbf{i},\mathbf{u}} g_{\mathbf{i}} a_{\mathbf{i};(\mathbf{u},\mathbf{v}_1)} \prod_{j=1}^{k_1} t^{(j)}_{u_{f_i(e_j)},v_{f_e(e_j)}} \right|^2.\end{aligned}$$

By the singular decomposition of $\mathbf{T}^{(j)}$ (see Theorem A.7), we have

$$t^{(j)}_{uv} = \sum_{\ell=1}^{r_j} \lambda^{(j)}_{\ell} \eta^{(j)}_{u\ell} \xi^{(j)}_{v\ell},$$

where $r_j = \min(m_j, n_j)$, $j = 1, \cdots, k_1$. By the definition of the graph G, all noncoincident u vertices in $\{u_{f_i(e_j)}, j = 1, \cdots, k_1\}$ are the indices in $\mathbf{u}$ and all noncoincident v vertices in $\{v_{f_e(e_j)}, j = 1, \cdots, k_1\}$ are the indices in $\mathbf{v}_2$. Let $\tilde{\mathbf{v}}_2 = (v_1, \cdots, v_{k_1})$, where v_j runs over $1, \cdots, n_j$ independently; that is, not restricted by the graph G. Similarly, define $\tilde{\boldsymbol{\ell}} = (\tilde{\ell}_1, \cdots, \tilde{\ell}_{k_1})$, where $\tilde{\ell}_j$ runs $1, \cdots, m_j$ independently. Then, substituting these expressions into the inequality above, we obtain

$$\begin{aligned}
&\|\mathbf{A} \cdot \mathbf{T}(G)\|^2 \\
&\le \prod_{j=k_1+1}^{k} \|\mathbf{T}^{(j)}\|^2 \sup_{\|g\|=1} \sum_{\mathbf{v}_1, \mathbf{v}_2} \left| \sum_{\mathbf{i},\mathbf{u}} g_{\mathbf{i}} a_{\mathbf{i};(\mathbf{u},\mathbf{v}_1)} \prod_{j=1}^{k_1} t^{(j)}_{u_{f_i(e_j)}, v_{f_e(e_j)}} \right|^2 \\
&\le \prod_{j=k_1+1}^{k} \|\mathbf{T}^{(j)}\|^2 \sup_{\|g\|=1} \sum_{\mathbf{v}_1, \mathbf{v}_2} \left| \sum_{\mathbf{i},\mathbf{u}} g_{\mathbf{i}} a_{\mathbf{i};(\mathbf{u},\tilde{\mathbf{v}}_1)} \prod_{j=1}^{k_1} t^{(j)}_{u_{f_i(e_j)}, v_j} \right|^2 \\
&= \prod_{j=k_1+1}^{k} \|\mathbf{T}^{(j)}\|^2 \sup_{\|g\|=1} \sum_{\mathbf{v}_1, \tilde{\mathbf{v}}_2} \left| \sum_{\boldsymbol{\ell}} \lambda^{(1)}_{\ell_1} \cdots \lambda^{(k_1)}_{\ell_{k_1}} \sum_{\mathbf{i},\mathbf{u}} g_{\mathbf{i}} a_{\mathbf{i};(\mathbf{u},\mathbf{v}_1)} \eta_{\mathbf{u};\boldsymbol{\ell}} \xi_{\tilde{\mathbf{v}}_2;\boldsymbol{\ell}} \right|^2 \\
&= \prod_{j=k_1+1}^{k} \|\mathbf{T}^{(j)}\|^2 \sup_{\|g\|=1} \sum_{\mathbf{v}_1} \sum_{\boldsymbol{\ell}} (\lambda^{(1)}_{\ell_1} \cdots \lambda^{(k_1)}_{\ell_{k_1}})^2 \left| \sum_{\mathbf{i},\mathbf{u}} g_{\mathbf{i}} a_{\mathbf{i};(\mathbf{u},\mathbf{v}_1)} \eta_{\mathbf{u};\boldsymbol{\ell}} \right|^2 \\
&\le \prod_{j=1}^{k} \|\mathbf{T}^{(j)}\|^2 \sup_{\|g\|=1} \sum_{\mathbf{v}_1} \sum_{\boldsymbol{\ell}} \left| \sum_{\mathbf{i},\mathbf{u}} g_{\mathbf{i}} a_{\mathbf{i};(\mathbf{u},\mathbf{v}_1)} \eta_{\mathbf{u};\boldsymbol{\ell}} \right|^2 \\
&\le \prod_{j=1}^{k} \|\mathbf{T}^{(j)}\|^2 \sup_{\|g\|=1} \sum_{\mathbf{v}_1} \sum_{\tilde{\boldsymbol{\ell}}} \left| \sum_{\mathbf{i},\mathbf{u}} g_{\mathbf{i}} a_{\mathbf{i};(\mathbf{u},\mathbf{v}_1)} \eta_{\mathbf{u};\tilde{\boldsymbol{\ell}}} \right|^2 \\
&= \prod_{j=1}^{k} \|\mathbf{T}^{(j)}\|^2 \sup_{\|g\|=1} \sum_{\mathbf{u},\mathbf{v}_1} \left| \sum_{\mathbf{i}} g_{\mathbf{i}} a_{\mathbf{i};(\mathbf{u},\mathbf{v}_1)} \right|^2 \\
&= \prod_{j=1}^{k} \|\mathbf{T}^{(j)}\|^2 \|\mathbf{A}\|, \qquad \text{(A.4.5)}
\end{aligned}$$

where

$$\boldsymbol{\ell} = (\ell_1, \cdots, \ell_{k_1})',$$

$$\eta_{\mathbf{u};\boldsymbol{\ell}} = \prod_{j=1}^{k_1} \eta^{(j)}_{u_{f_i(e_j)}, \ell_j},$$

$$\xi_{\tilde{\mathbf{v}}_2;\boldsymbol{\ell}} = \prod_{j=1}^{k_1} \xi^{(j)}_{v_j,\ell_j}.$$

Here, the second identity in (A.4.5) follows from the fact that

$$\sum_{\tilde{\mathbf{v}}_2} \xi_{\tilde{\mathbf{v}}_2,\boldsymbol{\ell}} \bar{\xi}_{\tilde{\mathbf{v}}_2,\boldsymbol{\ell}'} = \delta_{\boldsymbol{\ell},\boldsymbol{\ell}'}$$

and the third identity from

$$\sum_{\tilde{\boldsymbol{\ell}}} \eta_{\mathbf{u},\tilde{\boldsymbol{\ell}}} \bar{\eta}_{\mathbf{u}',\tilde{\boldsymbol{\ell}}} = \delta_{\mathbf{u},\mathbf{u}'},$$

and $\delta_{a,b}$ is the Kronecker delta; i.e., $\delta_{a,a} = 1$ and $\delta_{a,b} = 0$ for $a \neq b$.

This completes the proof of Theorem A.28.

Remark A.29. When $V_1 = V_3 = \{1\}$ and $V_2 = \emptyset$, Theorem A.28 reduces to Corollary A.22.

A.4.3 Fundamental Theorem on Graph-Associated MMs

Definition A.30. A graph $G = (V, E, F)$ is called two-edge connected if, removing any one edge from G, the resulting subgraph is still connected.

The following theorem is fundamental in finding the existence of the LSD of a product of two random matrices. It was initially proved in Yin and Krishnaiah [304] for a common nonnegative definite matrix $\mathbf{T}$. Now, it is extended to any complex matrices with consistent dimensions.

Theorem A.31. (Fundamental theorem for graph-associated matrix). *Suppose that $G = (V, E, F)$ is a two-edge connected graph with t vertices and k edges. Each vertex i corresponds to an integer $m_i \geq 2$ and each edge e_j corresponds to a matrix $\mathbf{T}^{(j)}$, $j = 1, \cdots, k$, with consistent dimensions; that is, if $F(e_j) = (g, h)$, then the matrix $\mathbf{T}^{(j)}$ has dimensions $m_g \times m_h$. Define $\mathbf{v} = (v_1, \cdots, v_t)$ and*

$$T = \sum_{\mathbf{v}} \prod_{j=1}^{k} t^{(j)}_{v_{f_i(e_j)}, v_{f_e(e_j)}},$$

where the summation $\sum_{\mathbf{v}}$ is taken for $v_i = 1, 2, \cdots, m_i$, $i = 1, 2, \cdots, t$. Then, for any $i \leq t$,

$$|T| \leq m_i \prod_{j=1}^{k} \|\mathbf{T}^{(j)}\|. \tag{A.4.6}$$

Because the graph $G = (V, E, F)$ is two-edge connected, the degree of vertex 1 is at least 2. Divide the edges connecting vertex 1 into two (non-empty) sets. Define a new graph G^* by splitting vertex 1 into two vertices $1'$ and $1''$ and connecting the edges of the two sets to the vertices $1'$ and $1''$, respectively. The correspondence between the edges and the matrices remains unchanged (both vertices $1'$ and $1''$ correspond to the integer m_1). For brevity, we denote the vertices $1'$ and $1''$ by 1 and 2 and other vertices by $3, \cdots, t+1$. Define an $m_1 \times m_1$ matrix $\mathbf{T}(G^*)$ with entries

$$\mathbf{T}_{v_1,v_2}(G^*) = \sum_{\mathbf{v}^*} \prod_{j=1}^{k} t^{(j)}_{v_{f_i(e_j)}, v_{f_e(e_j)}},$$

where $\mathbf{v}^* = (v_3, \cdots, v_{t+1})$.

One finds that T is the trace of the matrix $\mathbf{T}$ and hence Theorem A.31 is an easy consequence of the following theorem.

Theorem A.32.

$$\|\mathbf{T}\| \le \prod_{j=1}^{k} \|\mathbf{T}^{(j)}\|. \tag{A.4.7}$$

To prove Theorem A.32, we need to define a new graph G_p of $\tilde{t}$ vertices and $\tilde{k}$ edges associated with a new class of matrices $\widetilde{\mathbf{T}}^{(j)}$ with consistent dimensions such that the similarly defined matrix $\mathbf{T}(G_p) = \mathbf{T}(G^*)$, where

$$\widetilde{\mathbf{T}}(G_p) = \left(\sum_{\tilde{\mathbf{v}}} \prod_{j=1}^{k} \tilde{t}^{(j)}_{v_{f_i(\tilde{e}_j)}, v_{f_e(\tilde{e}_j)}} \right)$$

and $\tilde{\mathbf{v}} = (v_3, \cdots, v_{\tilde{t}})$.

The graph G_p is directional and satisfies the following properties:

1. Every edge of G_p is on a directional path from vertex 1 to 2 (a path is a proper chain without cycles).

2. The graph G_p is direction-consistent; that is, it has no directional cycles.

3. Vertex 1 meets with only arrow tails and vertex 2 meets with only arrow heads, and all other vertices connect with both arrow heads and tails.

Remark A.33. Due to the second property, we have in fact established a partial order on the vertex set of G_p; in other words, we say that a vertex u is prior to vertex w if there is a directional path from u to w.

Construction of graph $\mathbf{G_p}$

Arbitrarily choose a circuit passing through vertex 1 of G, and split the edges connecting to vertex 1 into two sets so that the two edges connecting to vertex 1 do not belong to one set simultaneously (a circuit is a cycle without proper subcycles). Then, this circuit becomes a chain of G^* starting from vertex

1 and ending at vertex 2. (We temporarily denote the vertices of G^* by $1, \cdots, t+1$, and the numbering of vertices will automatically shift forward when new vertices are added.) Then, we may mark each edge of the chain as an arrow so that the chain forms a directional path from vertex 1 to vertex 2. Then, the directional chain, regarded as the directionalized subgraph of G^*, satisfies the three properties above.

Suppose we have directionalized or extended (if necessary) the graph G^* to G^{*d} with a directionalized subgraph G^d (of G^{*d}) starting from vertex 1 to vertex 2 and satisfying the three properties above.

If this subgraph G^d does not contain all edges of G^{*d}, we can find a simple path $\mathcal{P}$ with two distinct ends at two different vertices $a \prec b$ (say, of G^d or a circuit $\mathcal{C}$ with only one vertex A on G^d) since G is two-edge-connected. Consider the following cases.

Case 1. Suppose that path $\mathcal{P}$ ends at two distinct vertices $a \prec c$ of a directional path of the directionalized subgraph G^d. As an example, see Fig. A.1. Then, we mark arrows on the edges of $\mathcal{P}$ as a directional path from a to c. As shown in Fig. A.1(left), suppose that the undirectionalized path $\mathcal{P} = adec$ intersects the directionalized path abc at vertices a and c. Since the arrows on the path abc are from a to c, we mark arrows on path $adec$ from a to c.

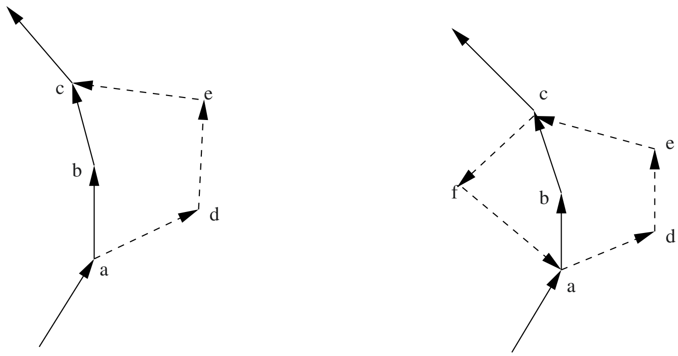

Fig. A.1 Directionalize a path attaching to a directional chain.

Now, let us show that the new subgraph $G^d \cup \overline{\mathcal{P}}$ satisfies the three conditions given above, where $\overline{\mathcal{P}}$ is the directionalized path $\mathcal{P}$.

Since G^d satisfies condition 1, there is a directional path from 1 to a and a directional path from c to 2, so we conclude that the directional graph $G^d \cup \overline{\mathcal{P}}$ satisfies condition 1.

If $G^d \cup \overline{\mathcal{P}}$ contains a directional cycle (say $\mathcal{F}$), then $\mathcal{F}$ must contain an edge of $\overline{\mathcal{P}}$ and an edge of G^d because G^d and $\overline{\mathcal{P}}$ have no directional cycles. Since $\overline{\mathcal{P}}$ is a simple path, $\mathcal{F}$ must contain the whole path $\overline{\mathcal{P}}$. Thus, the remaining part of $\mathcal{F}$ contained in G^d forms a directional chain from c to a. As shown in the right graph of Fig. A.1(right), the directional chains (cfa) and (abc) form a directional cycle of G^d. Because G^d also contains a directional path from a to c, we reach a contradiction to the assumption that G^d is direction-consistent. Thus, $G^d \cup \overline{\mathcal{P}}$ satisfies condition 2.

Condition 3 is trivially seen.

Case 2. Suppose that path $\mathcal{P}$ meets G^d at two distinct vertices a and c between which there are no directional paths in the directionalized subgraph G^d. As an example, see Fig. A.2. Then, we mark arrows on the edges of $\mathcal{P}$ as a directional path from a to c, say. As shown in the left graph of Fig. A.2(left), suppose that the undirectionalized path $\mathcal{P} = abc$ intersects G^d at vertices a and c. We make arrows on path abc from a to c (or the other direction without any harm).

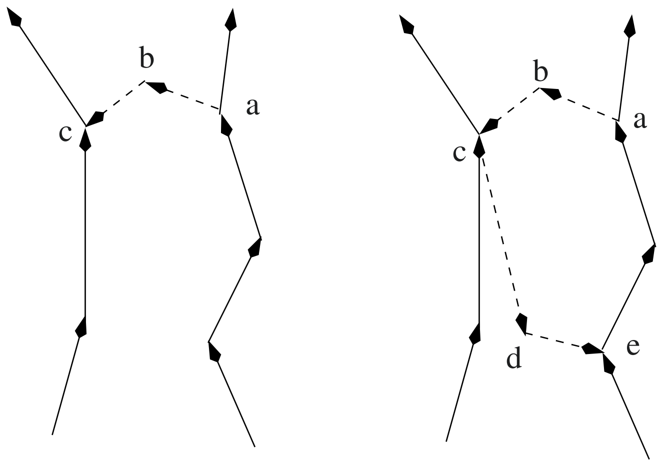

Fig. A.2 Directionalize a path attaching to incomparable vertices.

Because G^d satisfies condition 1 and contains an edge leading to vertex a if $a \neq 1$, there is a directional path from 1 to a. Similarly, there is a directional path from c to the vertex 2 if $c \neq 2$. Thus, the directionalized graph $G^d \cup \overline{\mathcal{P}}$ satisfies condition 1.

As shown in the right graph of Fig. A.2, if there is a directional cycle in the extended directional subgraph, then there would be a directional path $cdefa$ from c to a in G^d, which violates our assumption that a and c are not comparable. Thus, $G^d \cup \overline{\mathcal{P}}$ satisfies condition 2. Condition 3 is trivially satisfied.

Case 3. Suppose that there is a simple cycle $\mathcal{C}$ (or a loop), say. As an example (shown in Fig. A.3), the cycle $abcd$ intersects G^d at vertex a. Cut the graph G^{*d} off at a, separate vertex a as a' and a'' and add an arrowed edge from a'' to a', connect edges with a as initial vertices to a' and edges with a as end vertices to a'', stretch the cycle $\mathcal{C}$ ($abcd$ in Fig. A.3(left)) as a directional path from a'' to a', and finally connect other undirectionalized edges with a as end vertices to a'' or a' arbitrarily.

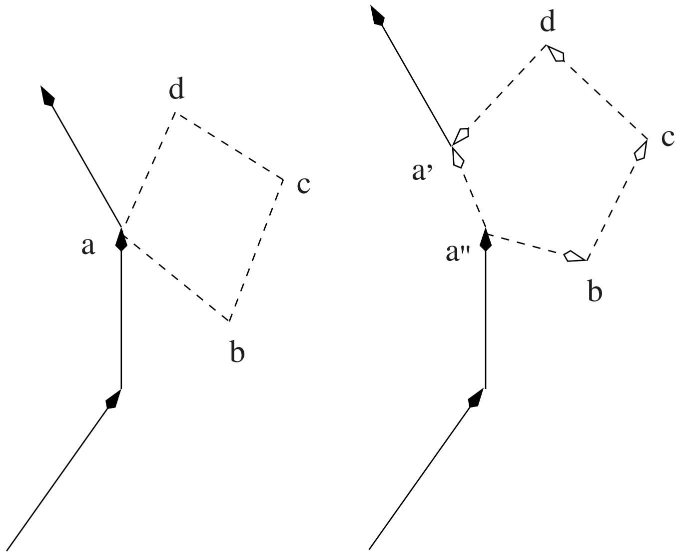

Fig. A.3 Directionalize a loop.

We can similarly show that the resulting directionalized subgraph satisfies the three conditions. If the newly added edge is made to correspond to an identity matrix of dimension m_a, the matrix $\mathbf{T}$ defined by the extended graph is the same as defined by the original graph.

By induction, we have eventually directionalized the graph G^*, with possible extensions, to a directional graph G_p, which satisfies the three properties.

Proof of Theorem A.32. By the argument above, we may assume that G^* is a directional graph and satisfies the properties above.

Now, we define a function g mapping the vertex set of G^* to nonnegative integers. We first define $g(1) = 0$. For a given vertex $u > 1$, there must be one but may be several directional paths from 1 to u. We define $g(u)$ to be the maximum number of edges among the directional paths from 1 to u.

For each nonnegative integer ℓ, define a vertex subset $V(\ell) = \{u \in V; g(u) = \ell\}$ with $V(0) = \{1\}$. If k_0 is the maximum number such that $V(k_0) \neq \emptyset$, then $V(k_0) = \{2\}$.

Note that the vertex sets $V(\ell)$ are disjoint and there are no edges connecting vertices of $V(\ell)$. Fixing an integer $\ell < k_0$, for each vertex $b \in V(\ell + 1)$, there is at least one vertex $a \in V(\ell)$ such that $(a, b) \in E$.

For each $0 < \ell \leq k_0$, define an MM $\mathbf{T}_{(\ell)}$ by

$$\mathbf{T}_{(\ell)} = \left(\prod_{\substack{f_i(e_j) \in V(0)+\cdots+V(\ell-1) \\ f_e(e_j) \in V(\ell)}} t^{(j)}_{i_{f_i(e_j)}, i_{f_e(e_j)}} \right)$$

and an MM $\mathbf{A}_{(\ell)}$ by

$$\mathbf{A}_{(\ell)} = \left(\sum_{\mathbf{i}} \prod_{\substack{f_i(e_j) \in V(0)+\cdots+V(\ell-1) \\ f_e(e_j) \in V(1)+\cdots+V(\ell)}} t^{(j)}_{i_{f_i(e_j)}, i_{f_e(e_j)}} \right),$$

where $\mathbf{i} = \{i_a : \; g(a) \leq \ell \; \& \; \forall g(b) > \ell, \;\; (a, b) \notin E\}$.

Intuitively, $\mathbf{t}_{(\ell)}$ is the MM defined by the subgraph of all edges starting from $V(0)+\cdots+V(\ell-1)$ and ending in $V(\ell)$ and their corresponding matrices, while $\mathbf{A}_{(\ell)}$ is the MM defined by the subgraph of all edges starting from $V(0) + \cdots + V(\ell - 1)$ and endeding in $V(0) + \cdots + V(\ell)$.

The left index of $\mathbf{A}_{(\ell)}$ is i_1 and its right indices are

$$\mathbf{v}^{\ell} = \{i_a : g(a) \leq \ell, \& \; \exists g(b) > \ell, (a, b) \in E\}.$$

The left indices of $\mathbf{T}_{(\ell)}$ are

$$\mathbf{u}^{\ell} = \{ \; i_a : \; g(a) \leq \ell - 1, \& \; \exists g(b) = \ell,$$

$$(a, b) \in E \& \; \forall g(c) > \ell, \; (a, c) \notin E\}$$

and its right indices are

$$\mathbf{v}^{\ell} = \{i_b : b \in V(\ell)\} \cup \{ \; i_a : g(a) \leq \ell - 1, \& \; \exists g(b) = \ell,$$

$$g(c) > \ell, \; (a, b), (a, c) \in E\}.$$

It is obvious that $\mathbf{u}^{\ell} \subset \mathbf{v}^{\ell-1}$ and

$$\mathbf{A}_{(\ell)} = \mathbf{A}_{(\ell-1)} \cdot \mathbf{T}_{(\ell)} = \sum_{\mathbf{u}} \mathbf{A}_{(\ell-1)}(i_1, \mathbf{v}^{(\ell-1)}) \mathbf{T}_{(\ell)}(\mathbf{u}, \mathbf{v}^{\ell}_1).$$

Applying Theorem A.28, we obtain

$$\|\mathbf{A}_{(\ell)}\| \leq \|\mathbf{A}_{(\ell-1)}\| \prod_{\substack{f_i(e_j) \in V(0)+\cdots+V(\ell-1) \\ f_e(e_j) \in V(\ell)}} \|\mathbf{T}^{(j)}\|.$$

For the case $\ell = 1$, we have $\mathbf{A}_{(1)} = \big(A_{(1)}(i_1; \mathbf{v}^1)\big) = \mathbf{T}_{(1)}$. It is easy to see that

$$\|\mathbf{A}_{(1)}\| = \left\| \left(\prod_{\substack{f_i(e_j)=1 \\ f_e(e_j)\in V(1)}} t^{(j)}_{i_1, i_{f_e(e_j)}} \right) \right\| \leq \prod_{\substack{f_i(e_j)=1 \\ f_e(e_j)\in V(1)}} \|\mathbf{T}^{(j)}\|.$$

Applying induction, it is proven that

$$\|\mathbf{A}_{(\ell)}\| \prod_{f_e(e_j)\in V(1)+\cdots+V(\ell)} \|T_j\|.$$

Especially, for $\ell = k_0$,

$$\|\mathbf{T}\| = \|\mathbf{A}_{(k_0)}\| \leq \prod_{j=1}^{k} \|T_j\|. \tag{A.4.8}$$

This completes the proof of the theorem.

Now, let us consider a connected graph G of k edges.

Definition A.34. An edge e is called a cutting edge if removing this edge will result in a disconnected subgraph.

Whether an edge is a cutting edge or not remains the same when a cutting edge is removed. Now, removing all cutting edges, the resulting subgraph consists of disjoint two-edge connected subgraphs, isolated loops, and/or isolated vertices, which we call the MC blocks. On the other hand, contracting these two-edge connected subgraphs results in a tree of cutting edges and their vertices. Suppose that corresponding to each edge e_ℓ there is a matrix $\mathbf{T}_\ell$ and dimensions of the matrices are consistent.

Theorem A.35. *Suppose that the edge set $E = E_1 + E_2$, where $E_1 = E - E_2$ and E_2 is the set of all cutting edges. If G is connected, then we have*

$$\left| \sum_{i_w \in V} \prod_{j=1}^{k} t_{i_{f_i(e_j)}, i^{(j)}_{f_e(e_j)}} \right| \leq p_0 \prod_{e_j\in E_1} \|\mathbf{T}^{(j)}\| \prod_{e_j\in E_2} \|\mathbf{T}^{(j)}\|_0, \tag{A.4.9}$$

where $p_0 = \min\{n_\ell;\ \ell \in V\}$, $\|\mathbf{T}^{(j)}\|_0 = \overline{n}(e_j)\max_{gh}|t^{(j)}_{g,h}|$, and $\overline{n}(e_j) = \max(m_j, n_j)$ is the maximum of the dimensions of the $\mathbf{T}^{(j)}$.

Furthermore, let V_2^* be a subset of the vertex set V. Denote by $\sum_{\{-V_2^*\}}$ the summation running for $i_w = 1, \cdots, m_w$ subject to the restriction that $i_{w_1} \neq i_{w_2}$ if both $w_1, w_2 \in V_2^*$. Then, we have

$$\left| \sum_{\{-V_2^*\}} \prod_{j=1}^{k} t_{i_{f_i(e_j)}, i_{f_e(e_j)}, j} \right| \leq C_k p_0 \prod_{e_j\in E_1} \|\mathbf{T}_j\| \prod_{e_j\in E_2} \|\mathbf{T}_j\|_0, \tag{A.4.10}$$

where C_k is a constant depending on k only.

Remark A.36. The second part of the theorem will be used for the inclusion-exclusion principle to estimate the values of MMs associated with graphs. The useful case is for $V_2^* = V$; that is, the indices for noncoincident vertices are not allowed to take equal values.

Proof. If the graph contains only one MC block, Theorem A.35 reduces to Theorem A.31. We shall prove the theorem by induction with respect to the number of MC blocks. Suppose that Theorem A.35 is true when the number of MC blocks of the graph G is less than u.

Now, we consider the case where G contains u (> 1) MC blocks. Select a vertex v_0 such that it corresponds to the smallest dimension of the matrices. Since the cutting edges form a tree if the MC blocks are contracted, we can select an MC block B that connects with only one cutting edge, say $e_c = (v_1, v_2)$, and does not contain the vertex v_0. Suppose that $v_1 \in B$ and $v_2 \in G - B - e_c$. Remove the MC block B and the cutting edge e_c from G and add a loop attached at the vertex v_2. Write the resulting graph as G'. Let the added loop correspond to the diagonal matrix

$$\mathbf{T}_0 = \mathrm{diag}\Bigg[\sum_{i_w,\, w \in B} t^{(c)}_{i_{f_i(e_c)},1} \prod_{e_j \in B} t^{(j)}_{i_{f_i(e_j)}, i_{f_e(e_j)}}, \cdots, \\ \cdots, \sum_{i_w,\, w \in B} t^{(c)}_{i_{f_i(e_c)},n_{v_2}} \prod_{e_j \in B} t^{(j)}_{i_{f_i(e_j)}, i_{f_e(e_j)}} \Bigg].$$

By Theorem A.31, we have

$$\|\mathbf{T}_0\| \le \overline{n}(e_c) \max_{ij} |t^{(c)}_{i,j}| \prod_{e_j \in B} \|\mathbf{T}^{(j)}\| = \|\mathbf{T}_c\|_0 \prod_{e_j \in B} \|\mathbf{T}^{(j)}\|.$$

Note that graph G' has $u - 1$ MC blocks. Then, by induction, we have

$$\left| \sum_{i_w \in V} \prod_{j=1}^{k} t_{i_{f_i(e_j)}, i_{f_e(e_j)}, j} \right| \le p_0 \prod_{e_j \in E_1 - B} \|\mathbf{T}_j\| \prod_{e_j \in E_2 - e_c} \|\mathbf{T}_j\|_0 \|\mathbf{T}_0\| \\ = p_0 \prod_{e_j \in E_1} \|\mathbf{T}_j\| \prod_{e_j \in E_2} \|\mathbf{T}_j\|_0.$$

The proof of (A.4.9) is complete.

Note that (A.4.9) is a special case of (A.4.10) when V_2^* is empty. We shall prove (A.4.10) by induction with respect to the cardinality of the set V_2^*. We have already proved that (A.4.10) is true when $\|V_2^*\| = 0$. Now, assume that (A.4.10) is true for $\|V_2^*\| \le a - 1 \ge 0$. We shall show that (A.4.10) is true for $\|V_2^*\| = a$.

Suppose that $w_1, w_2 \in V_2^*$ and $w_1 \ne w_2$. Write $\widetilde{V}_2^* = V_2^* - \{w_2\}$. Let $\widehat{G}$ denote the graph obtained from G by gluing the vertices w_1 and w_2 as one vertex, still denoted by w_1. Then, we have $\left\|\widetilde{V}_2^*\right\| = a-1$. Without loss of generality, let vertex w_1 correspond to a smaller dimension, say p_1. If the edge $\hat{e}_j$ of $\widehat{G}$ is obtained from the edge e_j of G with w_2 as a vertex, then, corresponding to $\hat{e}_j$, we define a matrix $\widehat{\mathbf{T}}^{(j)}$ by the first p_1 rows (or columns) of the matrix $\mathbf{T}^{(j)}$ when w_2 is the initial (or end, respectively) vertex of e_j. For all other edges, we define the associated matrices by $\widehat{\mathbf{T}}^{(j)} = \mathbf{T}^{(j)}$. Note that

$$\left\|\widehat{\mathbf{T}}^{(j)}\right\| \le \left\|\mathbf{T}^{(j)}\right\| \le \left\|\mathbf{T}^{(j)}\right\|_0$$

and

$$\left\|\widehat{\mathbf{T}}^{(j)}\right\|_0 \le \left\|\mathbf{T}^{(j)}\right\|_0.$$

For definiteness, write $\sum_{\{G,-V_2^*\}} = \sum_{\{-V_2^*\}}$. Then, we have

$$\sum_{G,\{-V_2^*\}} = \sum_{G,\{-\widetilde{V}_2^*\}} - \sum_{\widehat{G},\{-\widehat{V}_2^*\}}.$$

By the induction hypothesis, we have

$$\left|\sum_{G,\{-\widetilde{V}_2^*\}} \prod_{j=1}^{k} t_{i_{f_i(e_j)},i_{f_e(e_j)},j}\right| \le C_{k,1} p_0 \prod_{e_j\in E_1} \|\mathbf{T}_j\| \prod_{e_j\in E_2} \|\mathbf{T}_j\|_0. \tag{A.4.11}$$

When constructing the graph $\widehat{G}$, some cutting edge of G may be changed to a noncutting edge of $\widehat{G}$, while the noncutting edge of G remains a noncutting edge of $\widehat{G}$. By induction, we also have

$$\left|\sum_{\widehat{G},\{-\widehat{V}_2^*\}} \prod_{j=1}^{k} t_{i_{f_i(e_j)},i_{f_e(e_j)},j}\right| \le C_{k,2} p_0 \prod_{e_j\in E_1} \|\mathbf{T}_j\| \prod_{e_j\in E_2} \|\mathbf{T}_j\|_0. \tag{A.4.12}$$

Combining (A.4.11) and (A.4.12) and by induction, we complete the proof of (A.4.10) and hence the remaining part of the theorem.

A.5 Perturbation Inequalities

Theorem A.37. (i) *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $n \times n$ *normal matrices with eigenvalues* λ_k *and* δ_k, $k = 1, 2, \cdots, n$, *respectively. Then*

$$\min_\pi \sum_{k=1}^{n} |\lambda_k - \delta_{\pi(k)}|^2 \le \operatorname{tr}[(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^*] \le \max_\pi \sum_{k=1}^{n} |\lambda_k - \delta_{\pi(k)}|^2, \tag{A.5.1}$$

where $\pi = (\pi(1), \cdots, \pi(n))$ *is a permutation of* $1, 2, \cdots, n$.

(ii) *In* (i), *if* $\mathbf{A}$ *and* $\mathbf{B}$ *are two* $n \times p$ *matrices and* λ_k *and* δ_k, $k = 1, 2, \cdots, n$, *denote their singular values, then the conclusion in* (*A.5.1*) *remains true. If the singular values are arranged in descending order, then we have*

$$\sum_{k=1}^{\nu} |\lambda_k - \delta_k|^2 \le \mathrm{tr}[(\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^*],$$

where $\nu = \min\{p, n\}$.

Proof. Because a normal matrix is similar to a diagonal matrix through a unitary matrix, without loss of generality, we can assume that $\mathbf{A} = \mathrm{diag}(\lambda_k)$ and assume $\mathbf{B} = \mathbf{U}\Delta\mathbf{U}^*$, where $\Delta = \mathrm{diag}(\delta_k)$ and $\mathbf{U} = (u_{kj})$ is a unitary matrix. Then we have

$$\mathrm{tr}(\mathbf{A}\mathbf{A}^*) = \sum_{k=1}^{n} |\lambda_k|^2,$$

$$\mathrm{tr}(\mathbf{B}\mathbf{B}^*) = \sum_{k=1}^{n} |\delta_k|^2,$$

$$2\Re[\mathrm{tr}(\mathbf{A}\mathbf{B}^*)] = 2\Re\left(\sum_{kj} \lambda_k \bar{\delta}_j |u_{kj}|^2\right).$$

From these, we obtain

$$\mathrm{tr}[(\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^*] = \sum_{k=1}^{n} |\lambda_k|^2 + \sum_{k=1}^{n} |\delta_k|^2 - 2\Re\left(\sum_{kj} \lambda_k \bar{\delta}_j |u_{kj}|^2\right). \quad \text{(A.5.2)}$$

The proof of the first assertion of the theorem will be complete if one can show that there are two permutations π_j, $j = 1, 2$, of $1, 2, \cdots, n$ such that

$$\Re\sum_{k=1}^{n} \lambda_k \bar{\delta}_{\pi_1(k)} \le \Re\left(\sum_{kj} \lambda_k \bar{\delta}_j |u_{kj}|^2\right) \le \Re\sum_{k=1}^{n} \lambda_k \bar{\delta}_{\pi_2(k)}. \quad \text{(A.5.3)}$$

Assertion (A.5.3) is a trivial consequence of the following real linear programming problem:

$$\begin{array}{lll} \max \ \sum_{k,j} a_{kj} x_{kj} & \text{subject to constraints;} & \\ a_{ij} & \text{real;} & \\ x_{kj} \ge 0 & \text{for all} & 1 \le k, j \le n; \\ \sum_{k=1}^{n} x_{kj} = 1 & \text{for all} & 1 \le j \le n; \\ \sum_{j=1}^{n} x_{kj} = 1 & \text{for all} & 1 \le k \le n. \end{array} \quad \text{(A.5.4)}$$

In fact, we can show that

$$\min_{\pi} \sum_{i=1}^{n} a_{i,\pi(i)} \leq \sum_{ij} a_{ij} x_{ij} \leq \max_{\pi} \sum_{i=1}^{n} a_{i,\pi(i)}. \tag{A.5.5}$$

If (x_{ij}) forms a permutation matrix (i.e., each row (and column) has one element 1 and others 0), then for this permutation π^0 (i.e., for all i, $x_{i,\pi^0(i)} = 1$)

$$\min_{\pi} \sum_{i=1}^{n} a_{i,\pi(i)} \leq \sum_{ij} a_{ij} x_{ij} = \sum_{i=1}^{n} a_{i,\pi^0(i)} \leq \max_{\pi} \sum_{i=1}^{n} a_{i,\pi(i)}.$$

That is, assertion (A.5.5) holds. If (x_{ij}) is not a permutation matrix, then we can find a pair of integers i_1, j_1 such that $0 < x_{i_1,j_1} < 1$. By the condition that the rows sum up to 1, there is an integer $j_2 \neq j_1$ such that $0 < x_{i_1,j_2} < 1$. By the condition that the columns sum up to 1, there is an $i_2 \neq i_1$ such that $0 < x_{i_2,j_2} < 1$. Continuing this procedure, we can find integers $i_1, j_1, i_2, j_2, \cdots, i_k, j_k$ such that

$$\begin{aligned}
&i_1 \neq i_2, i_2 \neq i_3, \cdots, i_{k-1} \neq i_k, \\
&j_1 \neq j_2, j_2 \neq j_3, \cdots, j_{k-1} \neq j_k, \\
&0 < x_{i_t,j_t} < 1,\ 0 < x_{i_t,j_{t+1}} < 1, t = 1, 2, \cdots, k.
\end{aligned}$$

During the process, there must be a k such that $j_{k+1} = j_s$ for some $1 \leq s \leq k$ and hence we find a cycle on whose vertices the x-values are all positive. Such an example is shown in Fig. A.4(right), where we started from (i_1, j_2), stopped at $(i_5, j_5) = (i_2, j_2)$, and obtain a cycle

$$(i_2, j_2) \to (i_2, j_3) \to \cdots \to (i_4, j_5) \to (i_2, j_2).$$

Consider the cycle

$$(i_s, j_s) \ \to \ (s_1, j_{s+1}) \ \to \ (i_k, j_k) \ \to \ (i_k, j_s) \ \to \ (i_s, j_s),$$

which has the property that at the vertices of this route, all x_{ij}'s take positive values.

If

$$a_{i_s,j_s} + a_{i_{s+1},j_{s+1}} + \cdots + a_{i_k,j_k} \geq a_{i_s,j_{s+1}} + a_{i_{s+1},j_{s+2}} + \cdots + a_{i_k,j_{k+1}},$$

define

$$\begin{aligned}
\tilde{x}_{i_t,j_t} &= x_{i_t,j_t} + \delta,\ t = s, s+1, \cdots, k, \\
\tilde{x}_{i_t,j_{t+1}} &= x_{i_t,j_{t+1}} - \delta,\ t = s, s+1, \cdots, k, \\
\tilde{x}_{ij} &= x_{ij},\ \text{for other elements},
\end{aligned}$$

where $\delta = \min\{x_{i_t,j_{t+1}},\ t = s, s+1, \cdots, k\} > 0$.

If

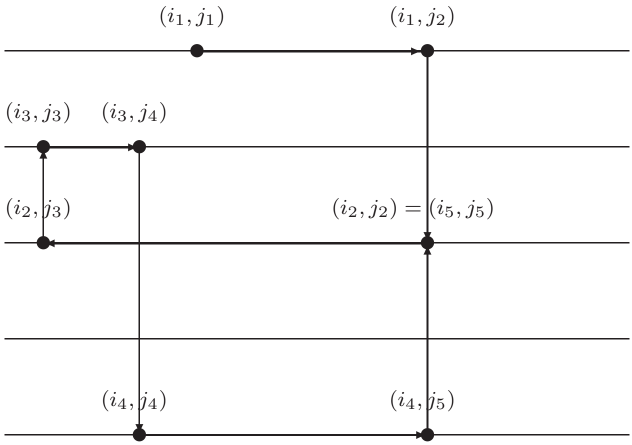

Fig. A.4 Find a cycle of positive x_{ij}'s.

$$a_{i_s,j_s} + a_{i_{s+1},j_{s+1}} + \cdots + a_{i_k,j_k} < a_{i_s,j_{s+1}} + a_{i_{s+1},j_{s+2}} + \cdots + a_{i_k,j_{k+1}},$$

define

$$\tilde{x}_{i_t,j_t} = x_{i_t,j_t} - \delta, \ t = s, s+1, \cdots, k,$$
$$\tilde{x}_{i_t,j_{t+1}} = x_{i_t,j_{t+1}} + \delta, \ t = s, s+1, \cdots, k,$$
$$\tilde{x}_{ij} = x_{ij}, \ \text{for other elements},$$

where $\delta = \min\{x_{i_t,j_t}, \ t = s, s+1, \cdots, k\} > 0$.

For both cases, it is easy to see that

$$\sum_{ij} a_{ij} x_{ij} \le \sum_{ij} a_{ij} \tilde{x}_{ij}$$

and $\{\tilde{x}_{ij}\}$ still satisfies condition (A.5.4). Note that the set $\{\tilde{x}_{ij}\}$ has at least one more 0 entry than $\{x_{ij}\}$. If $(\tilde{x}_{ij})$ is still not a permutation matrix, repeat the procedure above until the matrix is transformed to a permutation matrix. The inequality on the right-hand side of (A.5.5) follows. The inequality on the left-hand side follows from the inequality on the right-hand side by considering $\sum_{ij}(-a_{ij})x_{ij}$. Consequently, conclusion (i) of the theorem is proven.

In applying the linear programming above to our maximization problem, $a_{kj} = \Re(\lambda_k \bar{\delta}_j)$ and $x_{kj} = |u_{kj}|^2$.

As for the proof of the second part of the theorem, by the singular decomposition theorem, we may assume that $\mathbf{A} = \text{diag}[\lambda_1, \cdots, \lambda_\nu]$ and $\mathbf{B}^* = \mathbf{U}\text{diag}[\delta_1, \cdots, \delta_\nu]\mathbf{V}$, where $\mathbf{U} = (u_{ij})$ $(p \times \nu)$ and $\mathbf{V} = (v_{ij})$ $(n \times \nu)$ satisfy $\mathbf{U}^*\mathbf{U} = \mathbf{V}^*\mathbf{V} = \mathbf{I}_\nu$. Also, we may assume that $\lambda_1 \ge \cdots \ge \lambda_\nu \ge 0$ and

$\delta_1 \geq \cdots \geq \delta_\nu \geq 0$. Similarly, we have

$$\begin{aligned}
&\operatorname{tr}[(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^*] \\
&= \operatorname{tr}\mathbf{A}\mathbf{A}^* + \operatorname{tr}\mathbf{B}\mathbf{B}^* - 2\Re\operatorname{tr}\mathbf{A}\mathbf{B}^* \\
&= \sum_{k=1}^{n} \lambda_k^2 + \sum_{k=1}^{n} \delta_k^2 - 2\left(\sum_{k,j=1}^{\nu} \lambda_i \delta_j \Re(u_{ij} v_{ji})\right) \\
&\geq \sum_{k=1}^{n} \lambda_k^2 + \sum_{k=1}^{n} \delta_k^2 - 2\sum_{i,j=1}^{\nu} \lambda_i \delta_j |u_{ij} v_{ji}|.
\end{aligned}$$

Thus, the second conclusion follows if one can show that

$$\sum_{i,j=1}^{\nu} \lambda_i \delta_j |u_{ij} v_{ji}| \leq \sum_{i=1}^{\nu} \lambda_i \delta_i. \tag{A.5.6}$$

Note that

$$\sum_{i=1}^{\nu} |u_{ij} v_{ji}| \leq \left(\sum_{i=1}^{\nu} |u_{ij}|^2 \sum_{i=1}^{\nu} |v_{ji}|^2\right)^{1/2} \leq 1$$

and similarly

$$\sum_{i=1}^{\nu} |u_{ij} v_{ji}| \leq 1.$$

Thus, (A.5.6) is a special case of the problem

$$\max \sum_{i,j=1}^{\nu} \lambda_i \delta_j x_{ij} = \sum_{i=1}^{\nu} \lambda_i \delta_i \tag{A.5.7}$$

under the constraints

$$\begin{aligned}
&x_{ij} \geq 0, \\
&\sum_{i=1}^{\nu} x_{ij} \leq 1, \text{ for all } j, \\
&\sum_{j=1}^{\nu} x_{ij} \leq 1, \text{ for all } i.
\end{aligned}$$

Now, let

$$\begin{aligned}
u_1 &= \lambda_1 - \lambda_2 \geq 0, & v_1 &= \delta_1 - \delta_2 \geq 0, \\
&\ldots\ldots & &\ldots\ldots \\
u_{\nu-1} &= \lambda_{\nu-1} - \lambda_\nu \geq 0, & v_{\nu-1} &= \delta_{\nu-1} - \delta_\nu \geq 0, \\
u_\nu &= \lambda_\nu \geq 0, & v_\nu &= \delta_\nu \geq 0, \\
a_{s,t} &= \textstyle\sum_{i=1}^{s} \sum_{j=1}^{t} x_{ij} \leq \min(s,t).
\end{aligned}$$

Then,

$$\sum_{i,j=1}^{\nu} \lambda_i \delta_j x_{ij} = \sum_{i,j=1}^{\nu} \sum_{s=i}^{\nu} u_s \sum_{t=j}^{\nu} u_s v_t x_{ij}$$
$$= \sum_{s,t}^{\nu} u_s v_t a_{st}.$$

From this, it is easy to see that the maximum is attained when $a_{s,t} = \min(s,t)$, which implies that $x_{ii} = 1$ and $x_{ij} = 0$. This completes the proof of the theorem.

Theorem A.38. *Let $\{\lambda_k\}$ and $\{\delta_k\}$, $k = 1, 2, \cdots, n$, be two sets of complex numbers and their empirical distributions be denoted by F and $\overline{F}$. Then, for any $\alpha > 0$, we have*

$$L(F, \overline{F})^{\alpha+1} \le \min_{\pi} \frac{1}{n} \sum_{k=1}^{n} |\lambda_k - \delta_{\pi(k)}|^{\alpha}, \tag{A.5.8}$$

where L is the Levy distance between two two-dimensional distribution functions F and G defined by

$$L(F, G) = \inf\{\varepsilon : \ F(x-\varepsilon, y-\varepsilon)-\varepsilon \le G(x, y) \le F(x+\varepsilon, y+\varepsilon)+\varepsilon\}. \tag{A.5.9}$$

Remark A.39. For one-dimensional distribution functions F and G, we may regard them as two-dimensional distributions in the following manner:

$$\widetilde{F}(x, y) = \begin{cases} F(x), & \text{if } y \ge 0, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\widetilde{G}(x, y) = \begin{cases} G(x), & \text{if } y \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$

Then, the Levy distance $L(\widetilde{F}(x, y), \widetilde{G}(x, y))$ reduces to the usual definition of the Levy distance for one-dimensional distributions $L(F, G)$.

Remark A.40. It is not difficult to show that convergence in the metric L implies convergence in distribution.

Proof. To prove (A.5.8), we need only show that

$$L(F, \overline{F})^{\alpha+1} \le \frac{1}{n} \sum_{k=1}^{n} |\lambda_k - \delta_k|^{\alpha}. \tag{A.5.10}$$

Inequality (A.5.10) is trivially true if $d = \frac{1}{n} \sum_{k=1}^{n} |\lambda_k - \delta_k|^{\alpha} \ge 1$. Therefore, we need only consider the case where $d < 1$. Take ε such that $1 > \varepsilon^{\alpha+1} > d$. For fixed x and y, let $m = \#(A(x, y) \backslash B(x, y)$, where

$$A(x,y) = \{k \le n; \Re(\lambda_k) \le x,\ \Im(\lambda_k) \le y\}$$

and

$$B(x,y) = \{k \le n;\ \Re(\delta_k) \le x+\varepsilon, \Im(\delta_k) \le y+\varepsilon\}.$$

Then, we have

$$\begin{aligned} F(x,y) - \overline{F}(x+\varepsilon, y+\varepsilon) &\le \frac{m}{n} \\ &\le \frac{1}{n\varepsilon^\alpha} \sum_{k=1}^n |\lambda_k - \delta_k|^\alpha \\ &\le \varepsilon. \end{aligned}$$

Here the first inequality follows from the fact that the elements $k \in A(x,y) \backslash B(x,y)$ contribute to $F(x,y)$ but not to $\overline{F}(x,y)$, and the second inequality from the fact that for each $k \in A(x,y)\backslash B(x,y)$, $|\lambda_k - \delta_k| \ge \varepsilon$.

Similarly, we may prove that

$$\overline{F}(x-\varepsilon, y-\varepsilon) - F(x,y) \le \varepsilon.$$

Therefore, $L(F, \overline{F}) \le \varepsilon$, which implies the assertion of the lemma.

Combining Theorems A.37 and A.38 with $\alpha = 2$, we obtain the following corollaries.

Corollary A.41. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $n \times n$ *normal matrices with their ESDs* $F^{\mathbf{A}}$ *and* $F^{\mathbf{B}}$*. Then,*

$$L^3(F^{\mathbf{A}},\, F^{\mathbf{B}}) \le \frac{1}{n} \mathrm{tr}[(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^*]. \tag{A.5.11}$$

Corollary A.42. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $p \times n$ *matrices and the ESDs of* $\mathbf{S} = \mathbf{A}\mathbf{A}^*$ *and* $\overline{\mathbf{S}} = \mathbf{B}\mathbf{B}^*$ *be denoted by* $F^{\mathbf{S}}$ *and* $F^{\overline{\mathbf{S}}}$*. Then,*

$$L^4(F^{\mathbf{S}},\, F^{\overline{\mathbf{S}}}) \le \frac{2}{p^2} \left(\mathrm{tr}(\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*)\right) \left(\mathrm{tr}[(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^*]\right). \tag{A.5.12}$$

Proof. Denote the singular values of the matrices $\mathbf{A}$ and $\mathbf{B}$ by λ_k and δ_k, $k = 1, 2, \cdots, p$. Applying Theorems A.37 and A.38 with $\alpha = 1$, we have

$$\begin{aligned} L^2(F^{\mathbf{S}},\, F^{\overline{\mathbf{S}}}) &\le \frac{1}{p} \sum_{k=1}^p |\lambda_k^2 - \delta_k^2| \\ &\le \frac{1}{p} \left(\sum_{k=1}^p (\lambda_k + \delta_k)^2 \right)^{1/2} \left(\sum_{k=1}^p |\lambda_k - \delta_k|^2 \right)^{1/2} \\ &\le \frac{1}{p} \left(2\sum_{k=1}^p (\lambda_k^2 + \delta_k^2) \right)^{1/2} \left(\sum_{k=1}^p |\lambda_k - \delta_k|^2 \right)^{1/2} \end{aligned}$$

$$\leq \left(\frac{2}{p}\text{tr}(\mathbf{AA}^* + \mathbf{BB}^*)\right)^{1/2} \left(\frac{1}{p}\text{tr}[(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^*]\right)^{1/2}. \quad (A.5.13)$$

A.6 Rank Inequalities

In cases where the underlying variables are not iid, Corollaries A.41 and A.42 are not convenient in showing strong convergence. The following theorems are powerful in this case.

Theorem A.43. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $n \times n$ *Hermitian matrices. Then,*

$$\|F^{\mathbf{A}} - F^{\mathbf{B}}\| \leq \frac{1}{n}\text{rank}(\mathbf{A}-\mathbf{B}). \quad (A.6.1)$$

Throughout this book, $\|f\| = \sup_x |f(x)|$.

Proof. Since both sides of (A.6.1) are invariant under a common unitary transformation on $\mathbf{A}$ and $\mathbf{B}$, we may transform $\mathbf{A}-\mathbf{B}$ as $\begin{pmatrix} \mathbf{C} & 0 \\ 0 & 0 \end{pmatrix}$, where $\mathbf{C}$ is a full rank matrix. To prove (A.6.1), we may assume

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

where the order of $\mathbf{A}_{22}$ is $(n-k)\times(n-k)$ and $\text{rank}(\mathbf{A}-\mathbf{B}) = \text{rank}(\mathbf{A}_{11} - \mathbf{B}_{11}) = k$. Denote the eigenvalues of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A}_{22}$ by $\lambda_1 \leq \cdots \leq \lambda_n$, $\eta_1 \leq \cdots \leq \eta_n$, and $\tilde{\lambda}_1 \leq \cdots \leq \tilde{\lambda}_{(n-k)}$, respectively. By the interlacing theorem,[1] we have the relation that $\max(\lambda_j, \eta_j) \leq \tilde{\lambda}_j \leq \min(\lambda_{(j+k)}, \eta_{(j+k)})$, and we conclude that, for any $x \in (\tilde{\lambda}_{(j-1)}, \tilde{\lambda}_j)$,

$$\frac{j-1}{n} \leq F^{\mathbf{A}}(x) \quad (\text{and} \quad F^{\mathbf{B}}(x)) \quad < \frac{j+k}{n},$$

which implies (A.6.1).

Theorem A.44. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $p \times n$ *complex matrices. Then,*

$$\|F^{\mathbf{AA}^*} - F^{\mathbf{BB}^*}\| \leq \frac{1}{p}\text{rank}(\mathbf{A}-\mathbf{B}). \quad (A.6.2)$$

More generally, if $\mathbf{F}$ *and* $\mathbf{D}$ *are Hermitian matrices of orders* $p\times p$ *and* $n \times n$*, respectively, then we have*

[1] The interlacing theorem says that if $\mathbf{C}$ is an $(n-1)\times(n-1)$ major sub-matrix of the $n \times n$ Hermitian matrix $\mathbf{A}$, then $\lambda_1(\mathbf{A}) \geq \lambda_1(\mathbf{C}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_{n-1}(\mathbf{C}) \geq \lambda_n(\mathbf{A})$, where $\lambda_i(\mathbf{A})$ denotes the i-th largest eigenvalues of the Hermitian matrix $\mathbf{A}$. A reference for this theorem may be found in Rao and Rao [237]. In fact, this theorem may be easily proven by the formula $\lambda_i(\mathbf{A}) = \inf_{\mathbf{y}_1,\cdots,\mathbf{y}_{i-1}} \sum_{\mathbf{x}\perp\mathbf{y}_1,\cdots,\mathbf{y}_{i-1}} \mathbf{x}^*\mathbf{A}\mathbf{x}/\mathbf{x}^*\mathbf{x}$.

$$\|F^{\mathbf{F}+\mathbf{ADA}^*} - F^{\mathbf{F}+\mathbf{BDB}^*}\| \le \frac{1}{p}\mathrm{rank}(\mathbf{A}-\mathbf{B}). \tag{A.6.3}$$

Proof. Let $\mathbf{C} = \mathbf{B} - \mathbf{A}$. Write $\mathrm{rank}(\mathbf{C}) = k$. Then, applying Theorem A.8, it follows that for any nonnegative integer $i \le p - k$,

$$\sigma_{i+k+1}(\mathbf{A}) \le \sigma_{i+1}(\mathbf{B}) \ \text{ and } \ \sigma_{i+k+1}(\mathbf{B}) \le \sigma_{i+1}(\mathbf{A}).$$

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

$$\begin{aligned} F^{\mathbf{BB}^*}(x) &= 1 - \frac{i}{p} = 1 - \frac{i+k}{p} + \frac{k}{p} \\ &\le F^{\mathbf{AA}^*}(x) + \frac{k}{p}. \end{aligned}$$

This has in fact proved that, for all x,

$$F^{\mathbf{BB}^*}(x) - F^{\mathbf{AA}^*}(x) \le \frac{k}{p}.$$

Similarly, we have

$$F^{\mathbf{AA}^*}(x) - F^{\mathbf{BB}^*}(x) \le \frac{k}{p}.$$

This completes the proof of (A.6.2).

The proof of (A.6.3) follows from the interlacing theorem and the following observation. If $\mathrm{rank}(\mathbf{A}-\mathbf{B}) = k$, then we may choose a $p \times p$ unitary matrix $\mathbf{U}$ such that

$$\mathbf{U}(\mathbf{A}-\mathbf{B}) = \begin{pmatrix} \mathbf{C}_1 & : k \times n \\ 0 & : (p-k) \times n \end{pmatrix}.$$

Write $\widetilde{F} = \mathbf{UFU}^* = \begin{pmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{pmatrix}$,

$$\widetilde{\mathbf{A}} = \mathbf{UA} = \begin{pmatrix} \mathbf{A}_1 & : k \times n \\ \mathbf{A}_2 & : (p-k) \times n \end{pmatrix},$$

and

$$\widetilde{\mathbf{B}} = \mathbf{UB} = \begin{pmatrix} \mathbf{B}_1 & : k \times n \\ \mathbf{A}_2 & : (p-k) \times n \end{pmatrix},$$

with $\mathbf{A}_1 - \mathbf{B}_1 = \mathbf{C}_1$. Then,

$$F^{\mathbf{F}+\mathbf{ADA}^*} = F^{\widetilde{\mathbf{F}}+\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}\widetilde{\mathbf{A}}^*} \ \text{ and } \ F^{\mathbf{F}+\mathbf{BDB}^*} = F^{\widetilde{\mathbf{F}}+\widetilde{\mathbf{B}}\widetilde{\mathbf{D}}\widetilde{\mathbf{B}}^*}.$$

Note that

$$\widetilde{\mathbf{F}} + \widetilde{\mathbf{A}}\widetilde{\mathbf{D}}\widetilde{\mathbf{A}}^* = \begin{pmatrix} \mathbf{F}_{11} + \mathbf{A}_1\mathbf{DA}_1^* & \mathbf{F}_{12} + \mathbf{A}_1\mathbf{DA}_2^* \\ \mathbf{F}_{21} + \mathbf{A}_2\mathbf{DA}_1^* & \mathbf{F}_{22} + \mathbf{A}_2\mathbf{DA}_2^* \end{pmatrix}$$

and

$$\widetilde{\mathbf{F}} + \widetilde{\mathbf{B}}\widetilde{\mathbf{D}}\widetilde{\mathbf{B}}^* = \begin{pmatrix} \mathbf{F}_{11} + \mathbf{B}_1\mathbf{D}\mathbf{A}_B^* & \mathbf{F}_{12} + \mathbf{B}_1\mathbf{D}\mathbf{A}_2^* \\ \mathbf{F}_{21} + \mathbf{A}_2\mathbf{D}\mathbf{B}_1^* & \mathbf{F}_{22} + \mathbf{A}_2\mathbf{D}\mathbf{A}_2^* \end{pmatrix}.$$

The bound (A.6.3) can be proven by similar arguments in the proof of Theorem A.43 and the comparison of eigenvalues of $\widetilde{\mathbf{F}} + \widetilde{\mathbf{A}}\widetilde{\mathbf{D}}\widetilde{\mathbf{A}}^*$, $\widetilde{\mathbf{F}} + \widetilde{\mathbf{B}}\widetilde{\mathbf{D}}\widetilde{\mathbf{B}}^*$, and $\mathbf{F}_{22} + \mathbf{A}_2\mathbf{D}\mathbf{A}_2^*$. The theorem is proved.

A.7 A Norm Inequality

The following theorems will be used to remove the diagonal elements of a random matrix or the mean matrix due to truncation in establishing the convergence rate of the ESDs.

Theorem A.45. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $n \times n$ *Hermitian matrices. Then,*

$$L(F^{\mathbf{A}}, F^{\mathbf{B}}) \le \|\mathbf{A} - \mathbf{B}\|. \tag{A.7.1}$$

The proof of the theorem follows from $L(F^{\mathbf{A}}, F^{\mathbf{B}}) \le \max_k |\lambda_k(\mathbf{A}) - \lambda_k(\mathbf{B})|$ and a theorem due to Horn and Johnson [154] given as follows.

Theorem A.46. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $n \times p$ *complex matrices. Then,*

$$\max_k |s_k(\mathbf{A}) - s_k(\mathbf{B})| \le \|\mathbf{A} - \mathbf{B}\|. \tag{A.7.2}$$

If $\mathbf{A}$ *and* $\mathbf{B}$ *are Hermitian, then the singular values can be replaced by eigenvalues; i.e.,*

$$\max_k |\lambda_k(\mathbf{A}) - \lambda_k(\mathbf{B})| \le \|\mathbf{A} - \mathbf{B}\|. \tag{A.7.3}$$

Proof. By (A.2.2), the first conclusion follows from

$$s_k(\mathbf{A}) = \min_{\mathbf{y}_1,\cdots,\mathbf{y}_{k-1}} \max_{\substack{\mathbf{x}\perp\mathbf{y}_1,\cdots,\mathbf{y}_{k-1} \\ \|\mathbf{x}\|=1}} \|\mathbf{A}\mathbf{x}\|$$
$$\begin{cases} \le \min\limits_{\mathbf{y}_1,\cdots,\mathbf{y}_{k-1}} \max\limits_{\substack{\mathbf{x}\perp\mathbf{y}_1,\cdots,\mathbf{y}_{k-1} \\ \|\mathbf{x}\|=1}} \|\mathbf{B}\mathbf{x}\| + \|\mathbf{A} - \mathbf{B}\| = s_k(\mathbf{B}) + \|\mathbf{A} - \mathbf{B}\|, \\ \ge \min\limits_{\mathbf{y}_1,\cdots,\mathbf{y}_{k-1}} \max\limits_{\substack{\mathbf{x}\perp\mathbf{y}_1,\cdots,\mathbf{y}_{k-1} \\ \|\mathbf{x}\|=1}} \|\mathbf{B}\mathbf{x}\| - \|\mathbf{A} - \mathbf{B}\| = s_k(\mathbf{B}) - \|\mathbf{A} - \mathbf{B}\|. \end{cases}$$

Similarly, the second conclusion follows from

$$\lambda_k(\mathbf{A}) = \min_{\mathbf{y}_1,\cdots,\mathbf{y}_{k-1}} \max_{\substack{\mathbf{x}\perp\mathbf{y}_1,\cdots,\mathbf{y}_{k-1} \\ \|\mathbf{x}\|=1}} \mathbf{x}^*\mathbf{A}\mathbf{x}$$
$$\begin{cases} \le \min\limits_{\mathbf{y}_1,\cdots,\mathbf{y}_{k-1}} \max\limits_{\substack{\mathbf{x}\perp\mathbf{y}_1,\cdots,\mathbf{y}_{k-1} \\ \|\mathbf{x}\|=1}} \mathbf{x}^*\mathbf{B}\mathbf{x} + \|\mathbf{A} - \mathbf{B}\| = \lambda_k(\mathbf{B}) + \|\mathbf{A} - \mathbf{B}\|, \\ \ge \min\limits_{\mathbf{y}_1,\cdots,\mathbf{y}_{k-1}} \max\limits_{\substack{\mathbf{x}\perp\mathbf{y}_1,\cdots,\mathbf{y}_{k-1} \\ \|\mathbf{x}\|=1}} \mathbf{x}^*\mathbf{B}\mathbf{x} - \|\mathbf{A} - \mathbf{B}\| = \lambda_k(\mathbf{B}) - \|\mathbf{A} - \mathbf{B}\|. \end{cases}$$

Theorem A.47. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $p \times n$ *complex matrices. Then,*

$$L(F^{\mathbf{AA}^*}, F^{\mathbf{BB}^*}) \leq 2\|\mathbf{A}\|\|\mathbf{A}-\mathbf{B}\| + \|\mathbf{A}-\mathbf{B}\|^2. \tag{A.7.4}$$

This theorem is a simple consequence of Theorem A.45 or Theorem A.46.

Appendix B
Miscellanies

B.1 Moment Convergence Theorem

One of the most popular methods in RMT is the moment method, which uses the moment convergence theorem (MCT). That is, suppose $\{F_n\}$ denotes a sequence of distribution functions with finite moments of all orders. The MCT investigates under what conditions the convergence of moments of all fixed orders implies the weak convergence of the sequence of the distributions $\{F_n\}$. In this chapter, we introduce Carleman's theorem.

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

$$\beta_{n,k} = \beta_k(F_n) := \int x^k dF_n(x). \tag{B.1.1}$$

Lemma B.1. (Unique limit). *A sequence of distribution functions $\{F_n\}$ converges weakly to a limit if the following conditions are satisfied:*

1. *Each F_n has finite moments of all orders.*

2. *For each fixed integer $k \geq 0$, $\beta_{n,k}$ converges to a finite limit β_k as $n \to \infty$.*

3. *If two right-continuous nondecreasing functions F and G have the same moment sequence $\{\beta_k\}$, then $F = G + const$.*

Proof. By Helly's theorem, $\{F_n\}$ has a subsequence $\{F_{n_i}\}$ vaguely convergent to (i.e., convergent at each continuity point of) a right-continuous nondecreasing function F.

Let $k \geq 0$ be an integer. We have the inequality

$$\begin{aligned}\left|\int_{|x|\geq K} x^k dF_{n_i}(x)\right| &\leq \frac{1}{K^{k+2}} \int_{|x|\geq K} x^{2k+2} dF_{n_i}(x) \\ &\leq \frac{1}{K^{k+2}} \sup_n \beta_{n,2k+2} < \infty.\end{aligned}$$

From this inequality, we can conclude that $\int_{|x|\geq K} x^k dF_{n_i} \to 0$ uniformly in i as $K \to \infty$, and

$$\int x^k dF_{n_i} \to \int x^k dF(x).$$

Thus, $\int x^k dF(x) = \beta_k$, and F is a distribution function (set $k = 0$).

If G is the vague limit of another vaguely convergent subsequence, then G must also be a distribution function and the moment sequence of G is also $\{\beta_k\}$. So, applying (3), $F = G$. Therefore, the whole sequence F_n converges vaguely to F. Since F is a distribution function, F_n converges weakly to F.

When we apply Lemma B.1, one needs to verify condition (3) of the lemma. The following lemmas give conditions that imply (3).

Lemma B.2. (M. Riesz). *Let $\{\beta_k\}$ be the sequence of moments of the distribution function F. If*

$$\liminf_{k\to\infty} \frac{1}{k}\beta_{2k}^{\frac{1}{2k}} < \infty, \tag{B.1.2}$$

then F is uniquely determined by the moment sequence $\{\beta_k,\, k = 0, 1, \cdots\}$.

This lemma is a corollary of the next lemma due to Carleman. However, we give a proof of Lemma B.2 because its proof is much easier than the latter and it is powerful enough in spectral analysis of large dimensional random matrices. The uninterested reader may skip the proof of Carleman's theorem.

Proof. Let F and G be two distributions with common moments β_k for all integers $k \geq 0$. Denote their characteristic functions by $f(t)$ and $g(t)$ (Fourier-Stieltjes transforms). We need only show that $f(t) = g(t)$ for all $t \geq 0$. Since F and G have common moments, we have, for all $j = 0, 1, \cdots,$

$$f^{(j)}(0) = g^{(j)}(0) = i^j \beta_j.$$

Define

$$t_0 = \sup\{\ t \geq 0;\ f^{(j)}(s) = g^{(j)}(s),$$

$$\text{for all}\ \ 0 \leq s \leq t\ \ \text{and}\ \ j = 0, 1, \cdots\}.$$

Then Lemma B.2 follows if $t_0 = \infty$. Suppose that $t_0 < \infty$. We have, for any j,

$$\int_{-\infty}^{\infty} x^j e^{it_0 x}[F(dx) - G(dx)] = 0.$$

By condition (B.1.2), there is a constant $M > 0$ such that

$$\beta_{2k} \leq (Mk)^{2k} \text{ for infinitely many } k.$$

Choosing $s \in (0, 1/(eM))$, applying the inequality that $k! > (k/e)^k$, and

$$|e^{ia} - 1 - ia - \cdots - (ia)^k/k!| \leq |a|^{k+1}/(k+1)! \tag{B.1.3}$$

(see Loève [200]), for any fixed $j \geq 0$, we have

$$
\begin{aligned}
&|f^{(j)}(t_0+s) - g^{(j)}(t_0+s)| \\
&= \left| \int_{-\infty}^{\infty} x^j e^{i(t_0+s)x} [F(dx) - G(dx)] \right| \\
&= \left| \int_{-\infty}^{\infty} x^j e^{it_0 x} \left[e^{isx} - 1 - isx - \cdots - \frac{(isx)^{2k-j-1}}{(2k-j-1)!} \right] \right. \\
&\qquad \left. \times [F(dx) - G(dx)] \right| \\
&\leq 2 \frac{s^{2k-j} \beta_{2k}}{(2k-j)!} \leq 2 \frac{(sMk)^{2k}}{s^j (2k-j)!} \\
&\leq 2(esMk/(2k-j))^{2k} (2k/s)^j \to 0,
\end{aligned}
$$

as $k \to \infty$ along those k such that $\beta_{2k} \leq (Mk)^{2k}$. This violates the definition of t_0. The proof of Lemma B.2 is complete.

Lemma B.3. (Carleman). *Let $\{\beta_k = \beta_k(F)\}$ be the sequence of moments of the distribution function F. If the Carleman condition*

$$\sum \beta_{2k}^{-1/2k} = \infty \tag{B.1.4}$$

is satisfied, then F is uniquely determined by the moment sequence $\{\beta_k,\ k = 0, 1, \cdots\}$.

Proof. Let F and G be two distribution functions with the common moment sequence $\{\beta_k\}$ satisfying condition (B.1.4). Let $f(t)$ and $g(t)$ be the characteristic functions of F and G, respectively. By the uniqueness theorem for characteristic functions, we need only prove that $f(t) = g(t)$ for all $t > 0$.

By the relation $\beta_{2k}^{1/2k} \leq \beta_{2k+2}^{1/(2k+2)}$, it is easy to see that Carleman's condition is equivalent to

$$\sum_{k=1}^{\infty} 2^k \beta_{2^k}^{-2^{-k}} = \infty. \tag{B.1.5}$$

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

$$h_{n,k} = n^{-1} 2^k \left(\beta_{2^k}^4 / \beta_{2^{k+1}}^{5/2} \right)^{2^{-k}}.$$

We first show that, for any n,

$$\sum_{k=1}^{\infty} h_{n,k} = \infty. \tag{B.1.6}$$

Let $c < 1/2$ be a positive constant and define

$$\mathcal{K}_1 = \{1\} \bigcup \{k: \ \beta_{2^k}^{2^{-k}} \geq c\beta_{2^{k+1}}^{2^{-k-1}}\}$$

and

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

We first show that

$$\sum_{k \in \mathcal{K}_1} 2^k \beta_{2^{k+1}}^{-2^{-k-1}} = \infty. \tag{B.1.7}$$

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

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

From this and the fact that $\mathcal{K}_1$ is nonempty, one can easily derive that

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

from which, along with condition (B.1.5), assertion (B.1.7) follows.

For each $k \in \mathcal{K}_1$, we have

$$h_{n,k} \geq c^4 n^{-1} 2^k \beta_{2^{k+1}}^{-2^{-k-1}}. \tag{B.1.8}$$

Then, by (B.1.7), for each fixed n, we have

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

Thus, for any $t > 0$, there is an integer m such that $t_{n,m-1} \leq t < t_{n,m}$, where $t_{n,j} = h_{n,1} + \cdots + h_{n,j}$, $j = 1, 2, \cdots, m-1$.

For simplicity of notation, we write $h_{n,m} = t - t_{n,m-1}$, $t_{n,0} = 0$, and $t_{n,m} = t$. Write $H = F - G$, $q_{n,1}(x) = \exp(ih_{n,1}x) - 1 - ih_{n,1}x$ and

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

For $k \leq m$, by inequality (B.1.3), we have

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

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

$$\begin{aligned}|f(t)-g(t)| &= \left|\int_{-\infty}^{\infty} e^{itx} H(dx)\right| \\ &= \left|\sum_{k\le m}\int_{-\infty}^{\infty} \exp[i(t-t_{n,k})x]q_{n,k}(x)H(dx)\right| \\ &\le \sum_{k\le m}\int_{-\infty}^{\infty} Q_{n,k}(x)(F(dx)+G(dx)) \\ &= 2\sum_{k\le m}\int_{-\infty}^{\infty} Q_{n,k}(x)F(dx). \end{aligned} \tag{B.1.9}$$

Expanding $Q_{n,k}(x)$, the general terms have the form

$$\frac{h_{n,1}^{\nu_1}}{\nu_1!}\cdots\frac{h_{n,k-1}^{\nu_{k-1}}}{\nu_{k-1}!}\frac{h_{n,k}^{2^k}|x|^{\nu}}{(2^k)!},$$

where $\nu=\nu_1+\cdots+\nu_{k-1}+2^k$ and $0\le\nu_j\le 2^j-1$. By the definition of $h_{n,k}$, the integral of this term is bounded by

$$\begin{aligned}&\frac{h_{n,1}^{\nu_1}}{\nu_1!}\cdots\frac{h_{n,k-1}^{\nu_{k-1}}}{\nu_{k-1}!}\frac{h_{n,k}^{2^k}\beta_\nu}{(2^k)!} \\ &\le \frac{n^{-\nu}2^{\mu}\beta_2^{2\nu_1}\beta_4^{4^{-1}(4\nu_2-5\nu_1)}\cdots\beta_{2^{k-1}}^{2^{-k+1}(4\nu_{k-1}-5\nu_{k-2})}}{\nu_1!\nu_2!\cdots\nu_{k-1}!} \\ &\quad\times\frac{\beta_{2^k}^{2^{-k}(2^{k+2}-5\nu_{k-1})}\beta_{2^{k+1}}^{2^{-k-1}\nu-5/2}}{(2^k)!},\end{aligned} \tag{B.1.10}$$

where $\mu=\nu_1+2\nu_2+\cdots+(k-1)\nu_{k-1}+k2^k$.

Note that

$$\begin{aligned}&4\nu_1+(4\nu_2-5\nu_1)+\cdots+(4\nu_{k-1}-5\nu_{k-2})+(2^{k+2}-5\nu_{k-1}) \\ &=2^{k+2}-\nu_1-\cdots-\nu_{k-1}=2^{k+2}+2^k-\nu>0.\end{aligned}$$

Applying $\beta_{2^s}\le\beta_{2^{k+1}}^{2^{-k-1+s}}$, which is a consequence of Hölder's inequality, we obtain

$$\begin{aligned}&\beta_2^{2\nu_1}\beta_4^{4^{-1}(4\nu_2-5\nu_1)}\cdots\beta_{2^{k-1}}^{2^{-k+1}(4\nu_{k-1}-5\nu_{k-2})}\beta_{2^k}^{2^{-k}(2^{k+2}-5\nu_{k-1})} \\ &\le\beta_{2^{k+1}}^{2^{-k-1}(4\nu_1+(5\nu_2-5\nu_1)+\cdots+(4\nu_{k-1}-5\nu_{k-2})+(2^{k+2}-5\nu_{k-1}))}=\beta_{2^{k+1}}^{5/2-2^{-k-1}\nu}.\end{aligned}$$

From this and (B.1.10), we obtain

$$\frac{h_{n,1}^{\nu_1}}{\nu_1!}\cdots\frac{h_{n,k-1}^{\nu_{k-1}}}{\nu_{k-1}!}\frac{h_{n,k}^{2^k}\beta_\nu}{(2^k)!}\le\frac{n^{-\nu}2^{\mu}}{\nu_1!\nu_2!\cdots\nu_{k-1}!(2^k)!}.$$

Therefore, noting that $\nu \geq 2^k$, we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} Q_{n,k}(x)F(dx) \\
&\leq \sum_{\nu_1+\cdots+\nu_{k-1}+2^k=\nu} \frac{(n^{-1}2)^{\nu_1}\cdots(n^{-1}2^{k-1})^{\nu_{k-1}}(n^{-1}2^k)^{2^k}}{\nu_1!\cdots\nu_{k-1}!(2^k)!} \\
&\leq \sum_{\substack{\nu_1+\cdots+\nu_k=\nu \\ \nu\geq 2^k}} \frac{(n^{-1}2)^{\nu_1}\cdots(n^{-1}2^k)^{\nu_k}}{\nu_1!\cdots\nu_k!} \\
&= \sum_{\nu=2^k}^{\infty}(n^{-1}(2+\cdots+2^k))^{\nu}/\nu! \leq \sum_{\nu=2^k}^{\infty}(2e/n)^{\nu} \\
&= (2e/n)^{2^k}\frac{n}{n-2e}.
\end{aligned}$$

Substituting this into (B.1.9), we get

$$|f(t)-g(t)| \leq \frac{n}{n-2e}\sum_{k=1}^{\infty}(2e/n)^{2^k} \to 0, \quad \text{letting} \quad n \to \infty.$$

The lemma then follows.

Remark B.4. Generally, the condition (B.1.4) cannot be further relaxed, which will be seen in examples given below. However, for one-sided distributions, this condition can be weakened. This is given in the following corollary. For ease of statement, in the following corollary, we assume the distributions are of nonnegative random variables. It is easy to see that the following corollary is true for one-sided distributions if we change the moments β_k to their absolute moments.

Corollary B.5. *Let F and G be two distribution functions with $F(0_-) = G(0_-) = 0$, $\beta_k(F) = \beta_k(G) = \beta_k$, for all integers $k \geq 1$, and*

$$\sum_{k=1}^{\infty}\beta_k^{-1/2k} = \infty. \tag{B.1.11}$$

Then, $F = G$.

Proof. Define $\tilde{F}$ by $\tilde{F}(x) = 1 - \tilde{F}(-x) = \frac{1}{2}(1 + F(x^2))$ for all $x > 0$ and similarly define $\tilde{G}$. Then, we have

$$\beta_{2k-1}(\tilde{F}) = \beta_{2k-1}(\tilde{G}) = 0 \quad \text{and} \quad \beta_{2k}(\tilde{F}) = \beta_{2k}(\tilde{G}) = \beta_k.$$

Applying Carleman's lemma, we get $\tilde{F} = \tilde{G}$. Consequently, $F = G$. The proof is complete.

The following example shows that, for distributions of nonnegative random variables, condition (B.1.11) cannot be further weakened as for some $\alpha > 0$,

$$\sum_{k=1}^{\infty} k^{\alpha}\beta_k^{-1/2k} = \infty. \tag{B.1.12}$$

Example B.6. For each $\alpha > 0$, there are two different distributions F and G with $F(0) = G(0) = 0$ such that, for each positive integer k, $\beta_k(F) = \beta_k(G) = \beta_k$ and

$$\sum_{k=1}^{\infty} k^{\alpha}\beta_k^{-1/2k} = \infty. \tag{B.1.13}$$

The example can be constructed in the following way. Set $\delta = 1/(2+2\alpha) < 1/2$ and define the densities of F and G by

$$f(x) = \begin{cases} ce^{-x^{\delta}}, & \text{if } x > 0,\\ 0, & \text{otherwise,}\end{cases}$$

and

$$g(x) = \begin{cases} ce^{-x^{\delta}}(1+\sin(ax^{\delta})), & \text{if } x > 0,\\ 0, & \text{otherwise,}\end{cases}$$

where $a = \tan(\pi\delta)$ and $c^{-1} = \int_0^{\infty} e^{-x^{\delta}}dx$. It is obvious that all moments of both F and G are finite. We begin our proof by showing that, for each k, $\beta_k(F) = \beta_k(G) = \beta_k$. To this end, it suffices to show that

$$\int_0^{\infty} x^k \exp(-x^{\delta})\sin(ax^{\delta})dx = 0. \tag{B.1.14}$$

Note that the integral on the left-hand side of the equality above is the negative imaginary part of the integral in the first line below:

$$\begin{aligned}&\int_0^{\infty} x^k \exp(-(1+ia)x^{\delta})dx\\ &= \delta^{-1}\int_0^{\infty} x^{(k+1)/\delta-1}\exp(-(1+ia)x)dx\\ &= \delta^{-1}(1+ia)^{(k+1)/\delta}\Gamma((k+1)/\delta).\end{aligned}$$

Note that $1 + ia = \exp(i\pi\delta)/\cos(\pi\delta)$, which implies that $(1 + ia)^{(k+1)/\delta}$ is real and hence the imaginary part of $\int_0^{\infty} x^k \exp(-(1+ia)x^{\delta})dx$ is zero. The proof of (B.1.14) is complete.

Note that

$$\beta_k = c\int_0^{\infty} x^k \exp(-x^{\delta})dx = c\delta^{-1}\Gamma\Big(\frac{k+1}{\delta}\Big).$$

Thus, by applying Stiring's formula,

$$k^{\alpha}\beta_k^{-1/2k} \sim k^{\alpha}\left(\frac{e\delta}{k}\right)^{1/2\delta} = \frac{1}{k}(e\delta)^{1/\delta},$$

which implies that $\sum k^{\alpha}\beta_k^{-1/2k} = \infty$.

Example B.7. For each $\alpha > 0$, there are two different distributions F and G such that, for each positive integer k, $\beta_k(F) = \beta_k(G) = \beta_k$ and

$$\sum_{k=1}^{\infty} k^{\alpha}\beta_{2k}^{-1/2k} = \infty. \tag{B.1.15}$$

In fact, construct $\widehat{F}$ and $\widehat{G}$ according to Example 1.4.1 with $\beta_k(\widehat{F}) = \beta_k(\widehat{G}) = \beta_{2k}$. For all $x > 0$, define $F(x) = 1 - F(-x) = \frac{1}{2}(1 + \widehat{F}(x^2))$ and $G(x) = 1 - G(-x) = \frac{1}{2}(1 + \widehat{G}(x^2))$. Then, F and G are the solutions.

B.2 Stieltjes Transform

Stieltjes transforms (also called Cauchy transforms in the literature) of functions of bounded variation are another important tool in RMT. If $G(x)$ is a function of bounded variation on the real line, then its Stieltjes transform is defined by

$$s_G(z) = \int \frac{1}{\lambda - z} dG(\lambda), \qquad z \in D,$$

where $z \in D \equiv \{z \in \mathbb{C} : \Im z > 0\}$.

B.2.1 Preliminary Properties

Theorem B.8. (Inversion formula). *For any continuity points $a < b$ of G, we have*

$$G\{[a, b]\} = \lim_{\varepsilon \to 0^+} \frac{1}{\pi} \int_a^b \Im\, s_G(x + i\varepsilon) dx.$$

If G is considered a finite signed measure, then Theorem B.8 shows a one-to-one correspondence between the finite signed measures and their Stieltjes transforms.

Proof. Note that

$$\frac{1}{\pi}\int_a^b \Im\, s_G(x + i\varepsilon) dx$$

$$= \frac{1}{\pi} \int_a^b \int \frac{\varepsilon dG(y)}{(x-y)^2+\varepsilon^2} dx$$
$$= \int \frac{1}{\pi} [\arctan(\varepsilon^{-1}(b-y)) - \arctan(\varepsilon^{-1}(a-y))] dG(y).$$

Letting $\varepsilon \to 0$ and applying the dominated convergence theorem, we find that the right-hand side tends to $G[a,b]$.

The importance of Stieltjes transforms also relies on the next theorem, which shows that to establish the convergence of ESD of a sequence of matrices, one needs only to show that convergence of their Stieltjes transforms and the LSD can be found by the limit Stieltjes transform.

Theorem B.9. *Assume that $\{G_n\}$ is a sequence of functions of bounded variation and $G_n(-\infty) = 0$ for all n. Then,*

$$\lim_{n\to\infty} s_{G_n}(z) = s(z) \;\; \forall z \in D \tag{B.2.1}$$

if and only if there is a function of bounded variation G with $G(-\infty) = 0$ and Stieltjes transform $s(z)$ and such that $G_n \to G$ vaguely.

Proof. If $G_n \to G$ vaguely, then (B.2.1) follows from the Helly-Bray theorem (see Loève [200]) since, for any fixed $z \in D$, both real and imaginary parts of $\frac{1}{x-z}$ are continuous and tending to 0 as $x \to \pm\infty$.

Conversely, suppose that (B.2.1) holds. For any subsequence of $\{G_n\}$, by Helly's selection theorem, we may select a further subsequence converging vaguely to a signed measure G. By (B.2.1) and the sufficiency part of the theorem, the Stieltjes transform of G is $s(z)$. Then, by Theorem B.8, the limit signed measure is unique. The proof of the theorem is complete.

Compared with the Fourier transform, an important advantage of Stieltjes transforms is that one can easily find the density function of a signed measure via its Stieltjes transform. We have the following theorem.

Theorem B.10. *Let G be a function of bounded variation and $x_0 \in \mathbb{R}$. Suppose that $\lim_{z\in D\to x_0} \Im\, s_G(z)$ exists. Call it $\Im\, s_G(x_0)$. Then G is differentiable at x_0, and its derivative is $\frac{1}{\pi}\Im\, s_G(x_0)$.*

Proof. Given $\varepsilon > 0$, let $\delta > 0$ be such that $|x - x_0| < \delta,\ 0 < y < \delta$ implies $\frac{1}{\pi}|\Im\, s_G(x+iy) - \Im\, s_G(x_0)| < \frac{\varepsilon}{2}$. Since all continuity points of G are dense in $\mathbb{R}$, there exist x_1, x_2 continuity points such that $x_1 < x_2$ and $|x_i - x_0| < \delta,\ i = 1, 2$. From Theorem B.8, we can choose y with $0 < y < \delta$ such that

$$\left| G(x_2) - G(x_1) - \frac{1}{\pi} \int_{x_1}^{x_2} \Im\, s_G(x+iy) dx \right| < \frac{\epsilon}{2}(x_2 - x_1).$$

For any $x \in [x_1, x_2]$, we have $|x - x_0| < \delta$. Thus

$$\begin{aligned}
&\left|\frac{G(x_2)-G(x_1)}{x_2-x_1}-\frac{1}{\pi}\Im\, s_G(x_0)\right| \\
&\le \frac{1}{x_2-x_1}\left|G(x_2)-G(x_1)-\frac{1}{\pi}\int_{x_1}^{x_2}\Im\, s_G(x+iy)dx\right| \\
&\quad+\frac{1}{x_2-x_1}\int_{x_1}^{x_2}\left|\frac{1}{\pi}\left(\Im\, s_G(x+iy)-\Im\, s_G(x_0)\right)\right|dx<\varepsilon.
\end{aligned}$$

Therefore, for all $\{x_n\}$ a sequence of continuity points of G with $x_n \to x_0$ as $n \to \infty$,

$$\lim_{n,m\to\infty}\frac{G(x_n)-G(x_m)}{x_n-x_m}=\frac{1}{\pi}\Im\, s_G(x_0).$$

This implies $\{G(x_n)\}$ is a Cauchy sequence. Thus, $\lim_{x\uparrow x_0} G(x) = \lim_{x\downarrow x_0} G(x)$, and therefore G is continuous at x_0.

Therefore, by choosing the sequence $\{x_1, x_0, x_2, x_0, \ldots\}$, we have

$$\lim_{n\to\infty}\frac{G(x_n)-G(x_0)}{x_n-x_0}=\frac{1}{\pi}\Im\, s_G(x_0). \tag{B.2.2}$$

To complete the proof of the theorem, we need to extend (B.2.2) to any sequence $\{x_n \to x_0\}$, where x_n may not necessarily be continuity points of G. To this end, let $\{x_n\}$ be a real sequence with $x_n \neq x_0$ and $x_n \to x_0$. For each n, we define $x_{nb},\ x_{nu}$ as follows. If there is a sequence y_{nm} of continuity points of G such that $y_{nm} \to x_n$ and $G(y_{n,m}) \to G(x_n)$ as $m \to \infty$, then we may choose y_{n,m_n} such that

$$\left|\frac{G(x_n)-G(x_0)}{x_n-x_0}-\frac{G(y_{n,m_n})-G(x_0)}{y_{n,m_n}-x_0}\right|<\frac{1}{n},$$

and then we define $x_{nb} = x_{nu} = y_{n,m_n}$. Otherwise, by the property of bounded variation, G should satisfy either $G(x_n-) < G(x_n) < G(x_n+)$ or $G(x_n-) > G(x_n) > G(x_n+)$. In the first case, we may choose continuity points x_{nb} and x_{nu} such that

$$x_n-\frac{1}{n}<x_{nb}<x_n<x_{nu}<x_n+\frac{1}{n}$$

and

$$\frac{G(x_{nb})-G(x_0)}{x_{nb}-x_0}<\frac{G(x_n)-G(x_0)}{x_n-x_0}<\frac{G(x_{nu})-G(x_0)}{x_{nu}-x_0}.$$

In the second case, we may choose continuity points x_{nb} and x_{nu} such that

$$x_n-\frac{1}{n}<x_{nu}<x_n<x_{nb}<x_n+\frac{1}{n}$$

and

$$\frac{G(x_{nb})-G(x_0)}{x_{nb}-x_0}<\frac{G(x_n)-G(x_0)}{x_n-x_0}<\frac{G(x_{nu})-G(x_0)}{x_{nu}-x_0}.$$

In all cases, we have $x_{nb} \to x_0$, $x_{nu} \to x_0$, x_{nb} and x_{nu} are continuity points of G and

$$\frac{G(x_{nb}) - G(x_0)}{x_{nb} - x_0} - \frac{1}{n} < \frac{G(x_n) - G(x_0)}{x_n - x_0} < \frac{G(x_{nu}) - G(x_0)}{x_{nu} - x_0} + \frac{1}{n}.$$

Letting $n \to \infty$ and applying (B.2.2) to the sequences x_{nb}, and x_{nu}, the inequality above proves that (B.2.2) is also true for the general sequence x_n and hence the proof of this theorem is complete.

In applications of Stieltjes transforms, its imaginary part will be used in most cases. However, we sometimes need to estimate its real part in terms of its imaginary part. We present the following result.

Theorem B.11. *For any distribution function F, its Stieltjes transform $s(z)$ satisfies*

$$|\Re(s(z))| \le v^{-1/2}\sqrt{\Im(s(z))}.$$

Proof. We have

$$\begin{aligned} |\Re(s(z))| &= \left| \int \frac{(x-u)dF(x)}{(x-u)^2 + v^2} \right| \\ &\le \int \frac{dF(x)}{\sqrt{(x-u)^2 + v^2}} \\ &\le \left(\int \frac{dF(x)}{(x-u)^2 + v^2} \right)^{1/2}. \end{aligned}$$

Then, the theorem follows from the observation that

$$\Im(s(z)) = v \int \frac{dF(x)}{(x-u)^2 + v^2}.$$

B.2.2 Inequalities of Distance between Distributions in Terms of Their Stieltjes Transforms

The following theorems create a methodology for establishing convergence rates of the ESD of RMs.

Theorem B.12. *Let F be a distribution function and let G be a function of bounded variation satisfying $\int |F(x) - G(x)| \mathrm{d}x < \infty$. Denote their Stieltjes transforms by $f(z)$ and $g(z)$, respectively. Then, we have*

$$\begin{aligned} ||F - G|| &:= \sup_x |F(x) - G(x)| \\ &\le \frac{1}{\pi(2\gamma - 1)} \left[\int_{-\infty}^{\infty} |f(z) - g(z)| du \right. \end{aligned}$$

$$+\frac{1}{v}\sup_x \int_{|y|\le 2va} |G(x+y)-G(x)|dy\Big], \tag{B.2.3}$$

where $z = u + iv,\ v > 0$, *and* a *and* γ *are constants related to each other by*

$$\gamma = \frac{1}{\pi}\int_{|u|<a} \frac{1}{u^2+1}du > \frac{1}{2}. \tag{B.2.4}$$

Proof. Write $\Delta = \sup_x |F(x) - G(x)|$. Without loss of generality, we can assume that $\Delta > 0$. Then, there is a sequence $\{x_n\}$ such that $F(x_n) - G(x_n) \longrightarrow \Delta$ or $-\Delta$.

We shall first consider the case where $F(x_n) - G(x_n) \longrightarrow \Delta$. For each x, we have

$$\begin{aligned}
&\frac{1}{\pi}\int_{-\infty}^{\infty} |f(z)-g(z)|du \\
&\ge \frac{1}{\pi}\int_{-\infty}^{x} \Im(f(z)-g(z))du \\
&= \frac{1}{\pi}\int_{-\infty}^{x}\left[\int_{-\infty}^{\infty} \frac{vd(F(y)-G(y))}{(y-u)^2+v^2}\right]du \\
&= \frac{1}{\pi}\int_{-\infty}^{x}\left[\int_{-\infty}^{\infty} \frac{2v(y-u)(F(y)-G(y))dy}{((y-u)^2+v^2)^2}\right]du \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty}(F(y)-G(y))\left[\int_{-\infty}^{x} \frac{2v(y-u)du}{((y-u)^2+v^2)^2}\right]dy \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{(F(x-vy)-G(x-vy))dy}{y^2+1}.
\end{aligned} \tag{B.2.5}$$

Here, the second equality follows from integration by parts, while the third follows from Fubini's theorem due to the integrability of $|F(y)-G(y)|$. Since F is nondecreasing, we have

$$\begin{aligned}
&\frac{1}{\pi}\int_{|y|<a} \frac{(F(x-vy)-G(x-vy))dy}{y^2+1} \\
&\ge \gamma(F(x-va)-G(x-va)) - \frac{1}{\pi}\int_{|y|<a} |G(x-vy)-G(x-va)|dy \\
&\ge \gamma(F(x-va)-G(x-va)) - \frac{1}{\pi v}\sup_x \int_{|y|<2va} |G(x+y)-G(x)|dy.
\end{aligned} \tag{B.2.6}$$

Take $x = x_n + va$. Then, (B.2.5) and (B.2.6) imply that

$$\frac{1}{\pi}\int_{-\infty}^{\infty} |f(z)-g(z)|du$$

$$\begin{aligned}
&\geq \gamma(F(x_n) - G(x_n)) \\
&\qquad -\frac{1}{\pi v} \sup_x \int_{|y|<2va} |G(x+y) - G(x)| dy - (1-\gamma)\Delta \\
&\longrightarrow (2\gamma - 1)\Delta - \frac{1}{\pi v} \sup_x \int_{|y|<2va} |G(x+y) - G(x)| dy,
\end{aligned}$$

which implies (B.2.3).

Now we consider the case where $F(x_n) - G(x_n) \longrightarrow -\Delta$. Similarly, we have, for each x,

$$\begin{aligned}
&\frac{1}{\pi} \int_{-\infty}^{\infty} |f(z) - g(z)| du \\
&\geq \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{(G(x - vy) - F(x - vy)) dy}{y^2 + 1} \\
&\geq \gamma(G(x + va) - F(x + va)) \\
&\quad -\frac{1}{\pi v} \sup_x \int_{|y|<2va} |G(x+y) - G(x)| dy - (1-\gamma)\Delta. \qquad \text{(B.2.7)}
\end{aligned}$$

By taking $x = x_n - va$, we have

$$\begin{aligned}
&\frac{1}{\pi} \int_{-\infty}^{\infty} |f(z) - g(z)| du \\
&\geq \gamma(G(x_n) - F(x_n)) \\
&\qquad -\frac{1}{\pi v} \sup_x \int_{|y|<2va} |G(x+y) - G(x)| dy - (1-\gamma)\Delta \\
&\longrightarrow (2\gamma - 1)\Delta - \frac{1}{\pi v} \sup_x \int_{|y|<2va} |G(x+y) - G(x)| dy,
\end{aligned}$$

which implies (B.2.3) for the latter case. This completes the proof of Theorem B.12.

Remark B.13. In the proof of Theorem B.12, one may find that the following version is stronger than Theorem B.12:

$$\begin{aligned}
||F - G|| \leq \frac{1}{\pi(2\gamma - 1)} \Bigg[&\int_{-\infty}^{\infty} |\Im(f(z) - g(z))| du \\
&+ \frac{1}{v} \sup_x \int_{|y| \leq 2va} |G(x+y) - G(x)| dy \Bigg]. \qquad \text{(B.2.8)}
\end{aligned}$$

However, in applying the inequalities, we did not find any significant superiority of (B.2.8) over (B.2.3).

Sometimes the functions F and G may have light tails, or both may even have bounded support. In such cases, we may establish a bound for $\|F - G\|$

by means of the integral of the absolute difference of their Stieltjes transforms on only a finite interval. We have the following theorem.

Theorem B.14. *Under the assumptions of Theorem B.12, we have*

$$\|F-G\| \le \frac{1}{\pi(1-\kappa)(2\gamma-1)}\left[\int_{-A}^{A}|f(z)-g(z)|du \right.$$
$$+2\pi v^{-1}\int_{|x|>B}|F(x)-G(x)|dx$$
$$\left. + v^{-1}\sup_x\int_{|y|\le 2va}|G(x+y)-G(x)|dy\right], \qquad \text{(B.2.9)}$$

where A and B are positive constants such that $A>B$ and

$$\kappa = \frac{4B}{\pi(A-B)(2\gamma-1)} < 1. \qquad \text{(B.2.10)}$$

The following corollary is immediate.

Corollary B.15. *In addition to the assumptions of Theorem B.12, assume further that, for some constant $B>0$, $F([-B,B])=1$ and $|G|((-\infty,-B)) = |G|((B,\infty)) = 0$, where $|G|((a,b))$ denotes the total variation of the signed measure G on the interval (a,b). Then, we have*

$$||F-G|| \le \frac{1}{\pi(1-\kappa)(2\gamma-1)}\left[\int_{-A}^{A}|f(z)-g(z)|du \right.$$
$$\left. +v^{-1}\sup_x\int_{|y|\le 2va}|G(x+y)-G(x)|dy\right],$$

where A, B, and κ are defined in (B.2.10).

Remark B.16. The benefit of using Theorem B.14 and Corollary B.15 is that we need only estimate the difference of Stieltjes transforms of the two distributions of interest on a finite interval. When Theorem B.14 is applied to establish the convergence rate of the spectral distribution of a sample covariance matrix in Chapter 3, it is crucial to the proof of Theorem 8.10 that A is independent of the sample size n. It should also be noted that the integral limit A in Girko's [122] inequality should tend to infinity with a rate of A^{-1} faster than the convergence rate to be established. Therefore, our Theorem B.14 and Corollary B.15 are much easier to use than Girko's inequality.

Proof of Theorem B.14. Using the notation given in the proof of Theorem B.12, we have

$$\int_A^\infty |f(z)-g(z)|du$$

$$= \int_A^\infty \left| \int_{-\infty}^\infty \frac{(F(x)-G(x))dx}{(x-z)^2} \right| du$$
$$\leq \int_A^\infty \left| \int_{-B}^B \frac{(F(x)-G(x))dx}{(x-z)^2} \right| du + \int_{-\infty}^\infty \left| \int_{|x|>B} \frac{(F(x)-G(x))dx}{(x-z)^2} \right| du$$
$$\leq 2B\Delta \int_A^\infty \frac{du}{(u-B)^2+v^2} + \pi v^{-1} \int_{|x|>B} |F(x)-G(x)|dx$$
$$\leq 2B\Delta/(A-B) + \pi v^{-1} \int_{|x|>B} |F(x)-G(x)|dx. \tag{B.2.11}$$

By symmetry, we get the same bound for $\int_{-\infty}^{-A} |f(z)-g(z)|du$. Substituting the inequality above into (B.2.3), we obtain (B.2.9) and the proof is complete.

B.2.3 Lemmas Concerning Levy Distance

Lemma B.17. *Let $L(F,G)$ be the Levy distance between the distributions F and G. Then we have*

$$L^2(F,G) \leq \int |F(x)-G(x)|dx. \tag{B.2.12}$$

Proof. Without loss of generality, assume that $L(F,G) > 0$. For any $r \in (0, L(F,G))$, there exists an x such that

$$F(x-r) - r > G(x) \quad (\text{or} \quad F(x+r) + r < G(x)).$$

Then the square between the points $(x-r, F(x-r)-r)$, $(x, F(x-r)-r)$, $(x-r, F(x-r))$, and $(x, F(x-r))$ (or $(x, F(x+r))$, $(x+r, F(x+r))$, $(x, F(x+r)+r)$, and $(x+r, F(x+r)+r)$ for the latter case) is located between F and G (see Fig. B.1). Then (B.2.12) follows from the fact that the right-hand side of (B.2.12) equals the area of the region between F and G. The proof is complete.

Lemma B.18. *If G satisfies $\sup_x |G(x+y)-G(x)| \leq D|y|^\alpha$ for all y, then*

$$L(F,G) \leq \|F-G\| \leq (D+1)L^\alpha(F,G), \quad \text{for all } F. \tag{B.2.13}$$

Proof. The inequality on the left-hand side is actually true for all distributions F and G. It can follow easily from the argument in the proof of Lemma B.17.

To prove the right-hand-side inequality, let us consider the case where, for some x,

$$F(x) > G(x) + \rho,$$

where $\rho \in (0, \|F-G\|)$. Since G satisfies the Lipschitz condition, we have

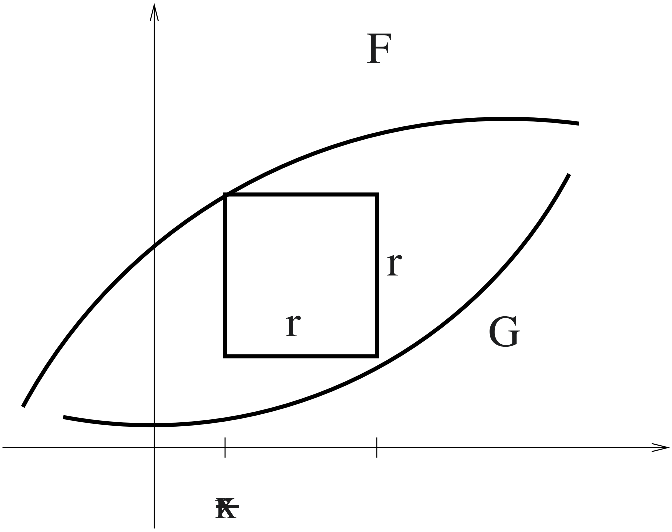

Fig. B.1 Levy distance.

$$G(x + (\rho/(D+1))^{1/\alpha}) + (\rho/(D+1))^{1/\alpha} \leq G(x) + \rho < F(x)$$

(see Fig. B.2), which implies that

$$L(F, G) \geq (\rho/(D+1))^{1/\alpha}.$$

Then, the right-hand-side inequality of (B.2.13) follows by making $\rho \to \|F - G\|$. The proof of the inequality for the other case (i.e., $G(x) > F(x) + \rho$) can be similarly proved.

Lemma B.19. *Let F_1, F_2 be distribution functions and let G satisfy* $\sup_x |G(x+u) - G(x)| \leq g(u)$*, for all u, where g is an increasing and continuous function such that $g(0) = 0$. Then*

$$\|F_1 - G\| \leq 3 \max\{\|F_2 - G\|, L(F_1, F_2), g(L(F_1, F_2))\}. \tag{B.2.14}$$

Proof. Let $0 < \rho < \|F_1 - G\|$, and assume that $\|F_2 - G\| < \rho/3$. Then, we may find an x_0 such that $F_1(x_0) - G(x_0) > \rho$ (or $F_1(x_0) - G(x_0)) < -\rho$ alternatively). Let $\eta > 0$ be such that $g(\eta) = \rho/3$. By the condition on G, for any $x \in [x_0, x_0 + \eta]$ (or $[x_0 - \eta, x_0]$ for the alternate case), we have $F_2(x) \leq G(x) + \frac{1}{3}\rho \leq G(x_0) + \frac{2}{3}\rho$ and $F_1(x) \geq F_1(x_0) > G(x_0) + \rho$. This shows that the rectangle $\{x_0 < x < x_0 + \eta, G(x_0) + \frac{2}{3}\rho < y < G(x_0) + \rho\}$ is located between F_1 and F_2 (see Fig. B.3). That means

$$L(F_1, F_2) \geq \min\left(\eta, \frac{1}{3}\rho\right).$$

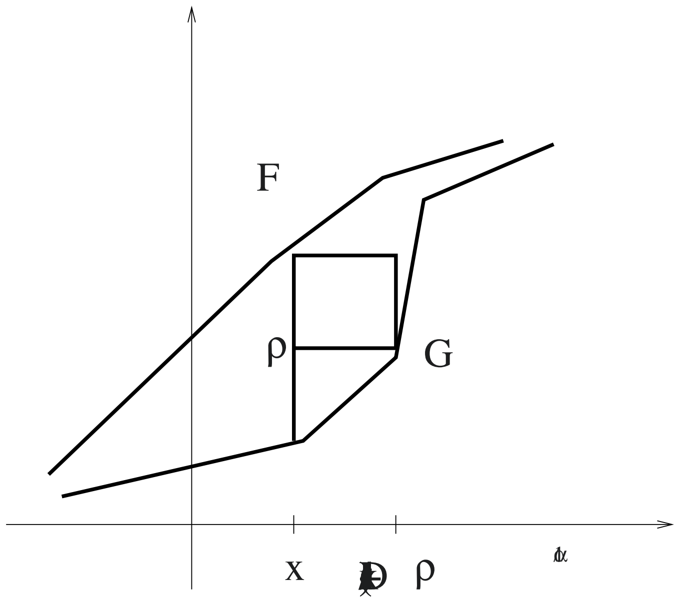

Fig. B.2 Relationship between Levy and Kolmogorov distances.

If $\eta < \frac{1}{3}\rho$, then $\eta \le L(F_1, F_2)$, which implies that

$$\frac{1}{3}\rho = g(\eta) \le g(L(F_1, F_2)).$$

Combining the three cases above, we conclude that

$$\rho \le 3\max\{\|F_2 - G\|,\ L(F_1, F_2),\ g(L(F_1, F_2))\}.$$

The lemma follows by letting $\rho \to \|F_1 - G\|$. The proof is complete.

B.3 Some Lemmas about Integrals of Stieltjes Transforms

Lemma B.20. *Suppose that $\phi(x)$ is a bounded probability density supported on a finite interval $[A, B]$. Then,*

$$\int_{-\infty}^{\infty} |s(z)|^2 du < 2\pi^2 M_\phi,$$

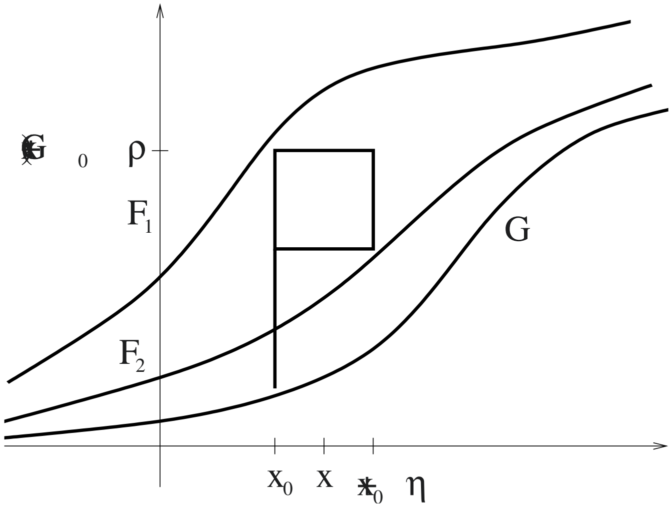

Fig. B.3 Further relationship between Levy and Kolmogorov distances.

where $s(z)$ is the Stieltjes transform of ϕ, M_ϕ is the upper bound of ϕ, and, in the integral, u is the real part of z.

Proof. We have

$$\begin{aligned} I &:= \int_{-\infty}^{\infty} |s(z)|^2 \mathrm{d}u \\ &= \int_{-\infty}^{\infty} \int_A^B \int_A^B \frac{\phi(x)\phi(y)\mathrm{d}x\mathrm{d}y}{(x-z)(y-\bar{z})} \mathrm{d}u \\ &= \int_A^B \int_A^B \phi(x)\phi(y)\mathrm{d}x\mathrm{d}y \int_{-\infty}^{\infty} \frac{1}{(x-z)(y-\bar{z})} \mathrm{d}u \ \ \text{(by Fubini)} \\ &= \int_A^B \int_A^B \frac{2\pi i}{y-x+2vi} \phi(x)\phi(y)\mathrm{d}x\mathrm{d}y \ \ \text{(residue theorem)}. \end{aligned}$$

Note that

$$\begin{aligned} &\int_A^B \int_A^B \Re\left(\frac{1}{y-x+2vi}\right) \phi(x)\phi(y)\mathrm{d}x\mathrm{d}y \\ &= \int_A^B \int_A^B \left(\frac{y-x}{(y-x)^2+4v^2}\right) \phi(x)\phi(y)\mathrm{d}x\mathrm{d}y = 0 \ \text{ by symmetry.} \end{aligned}$$

We finally obtain

$$I = -2\pi \int_A^B \int_A^B \Im\left(\frac{1}{y-x+2vi}\right) \phi(x)\phi(y)\mathrm{d}x\mathrm{d}y$$

$$
\begin{aligned}
&= 4\pi v \int_A^B \int_A^B \left(\frac{1}{(y-x)^2+4v^2}\right)\phi(x)\phi(y)\mathrm{d}x\mathrm{d}y \\
&\leq 4\pi v M_\phi \int_{-\infty}^{\infty}\int_A^B \phi(y)\left(\frac{1}{w^2+4v^2}\right)\mathrm{d}w\mathrm{d}y \quad \text{(making } w=x-y\text{)} \\
&= 2\pi^2 M_\phi.
\end{aligned}
$$

The proof is complete.

Corollary B.21. *When ϕ is the density of the semicircular law, we have*

$$
\int |s(z)|^2 du \leq 2\pi. \tag{B.3.1}
$$

Lemma B.22. *Let G be a function of bounded variation satisfying $\|G\| =: \sup_x |G(x)| < \infty$. Let $g(z)$ denote its Stieltjes transform. When $z = u + iv$ with $v > 0$, we have*

$$
I := \sup_u |g(z)| \leq \pi v^{-1}\|G\|. \tag{B.3.2}
$$

Proof. Using integration by parts, we have

$$
\begin{aligned}
|g(z) &= \left|\int \frac{G(x)}{(x-z)^2}dx\right| \\
&\leq \|G\| \int \frac{1}{(x-u)^2+v^2}dx \\
&= \pi v^{-1}\|G\|,
\end{aligned}
$$

which proves the lemma.

Lemma B.23. *Let G be a function of bounded variation satisfying $V(G) =: \int |G(du)| < \infty$. Let $g(z)$ denote its Stieltjes transform. When $z = u+iv$ with $v > 0$, we have*

$$
\int |g(z)|^2 \mathrm{d}u \leq 2\pi v^{-1} V(G)\|G\|. \tag{B.3.3}
$$

Proof. Following the same lines as in the proof of Lemma B.20, we may obtain

$$
\begin{aligned}
I &= 4\pi v \iint \left(\frac{1}{(u-x)^2+4v^2}\right) G(dx)G(du) \\
&= 8\pi v \int \left[\int \frac{(u-x)G(x)dx}{((u-x)^2+4v^2)^2}\right] G(du) \quad \text{(integration by parts)} \\
&\leq 2\pi v^{-1} V(G)\|G\|.
\end{aligned}
$$

Remark B.24. The two lemmas above give estimations for the difference of Stieltjes transforms of two distributions.

B.4 A Lemma on the Strong Law of Large Numbers

The Marcinkiewicz-Zygmund strong law of large numbers was first proved in [202], which gives necessary and sufficient conditions for the partial sample means from a single array of iid random variables with a rate of $n^{-(1-\alpha)}$, where $\alpha > \frac{1}{2}$. The following lemma is a generalization of this result to the case of multiple arrays of iid random variables.

Lemma B.25. *Let $\{X_{ij}, i, j = 1, 2, \cdots\}$ be a double array of iid complex random variables and let $\alpha > \frac{1}{2}$, $\beta \geq 0$, and $M > 0$ be constants. Then, as $n \to \infty$,*

$$\max_{j \leq Mn^\beta} \left| n^{-\alpha} \sum_{i=1}^{n} (X_{ij} - c) \right| \to 0 \quad \text{a.s.} \tag{B.4.1}$$

if and only if the following hold:

(i) $\mathrm{E}|X_{11}|^{(1+\beta)/\alpha} < \infty$;

(ii) $c = \begin{cases} \mathrm{E}(X_{11}), & \text{if } \alpha \leq 1, \\ \text{any number}, & \text{if } \alpha > 1. \end{cases}$

Furthermore, if $\mathrm{E}|X_{11}|^{(1+\beta)/\alpha} = \infty$, *then*

$$\limsup \max_{j \leq Mn^\beta} \left| n^{-\alpha} \sum_{i=1}^{n} (X_{ij} - c) \right| = \infty \quad \text{a.s.}$$

Proof of sufficiency. Without loss of generality, assume that $c = \mathrm{E}(X_{11}) = 0$ for the case $\alpha \leq 1$.

Define $X_{ijk} = X_{ij} I(|X_{ij}| \leq 2^{k\alpha})$. Then, by condition (i),

$$\begin{aligned}
&P\left(\max_{j \leq Mn^\beta} \left| n^{-\alpha} \sum_{i=1}^{n} X_{ij} \right| \neq \max_{j \leq Mn^\beta} \left| n^{-\alpha} \sum_{i=1}^{n} X_{ijk} \right|, \text{i.o.} \right) \\
&\leq \sum_{k=K}^{\infty} \mathrm{P} \left(\bigcup_{n=2^k}^{2^{k+1}-1} \bigcup_{i=1}^{n} \bigcup_{j \leq Mn^\beta} \{|X_{ij}| \geq n^\alpha\} \right) \\
&\leq \sum_{k=K}^{\infty} \mathrm{P} \left(\bigcup_{n=2^k}^{2^{k+1}-1} \bigcup_{i=1}^{2^{k+1}} \bigcup_{j \leq M2^{(k+1)\beta}} \{|X_{ij}| \geq 2^{k\alpha}\} \right) \\
&\leq \sum_{k=K}^{\infty} \mathrm{P} \left(\bigcup_{i=1}^{2^{k+1}} \bigcup_{j \leq M2^{(k+1)\beta}} \{|X_{ij}| \geq 2^{k\alpha}\} \right) \\
&\leq \sum_{k=K}^{\infty} M2^{(k+1)(\beta+1)} \mathrm{P}\left(|X_{11}| \geq 2^{k\alpha}\right)
\end{aligned}$$

$$= \sum_{k=K}^{\infty} M2^{(k+1)(\beta+1)} \sum_{\ell=k}^{\infty} \mathrm{P}\left(2^{\ell\alpha} \le |X_{11}| < 2^{(\ell+1)\alpha}\right)$$
$$= \sum_{\ell=K}^{\infty} \mathrm{P}\left(2^{\ell\alpha} \le |X_{11}| < 2^{(\ell+1)\alpha}\right) \sum_{k=K}^{\ell} M2^{(k+1)(\beta+1)}$$
$$\le \sum_{\ell=K}^{\infty} M2^{(\ell+2)(\beta+1)} \mathrm{P}\left(2^{\ell\alpha} \le |X_{11}| < 2^{(\ell+1)\alpha}\right)$$
$$\le M2^{\beta+1}\mathrm{E}|X_{11}|^{(\beta+1)/\alpha} I(|X_{11}| \ge 2^{K\alpha}) \to 0$$

as $K \to \infty$.

This proves that the convergence (B.4.1) is equivalent to

$$\max_{j \le Mn^\beta} \left| n^{-\alpha} \sum_{i=1}^{n} X_{ijk} \right| \to 0 \ \text{ a.s., as } \ 2^k < n \le 2^{k+1} \to \infty. \tag{B.4.2}$$

Note that

$$\max_{j \le Mn^\beta} \left| n^{-\alpha} \sum_{i=1}^{n} \mathrm{E}(X_{ijk}) \right| = n^{-\alpha+1} |\mathrm{E}X_{11k}|$$
$$\le \begin{cases} n^{-\alpha+1} 2^{-k(\beta+1)} \mathrm{E}|X_{11}|^{(\beta+1)/\alpha} I(|X_{11}| > 2^{k\alpha}), & \text{if } \alpha \le 1 \\ n^{-(\alpha-1)/2} + n^{-\alpha+1} 2^{k(\alpha-1)} \mathrm{E}|X_{11}|^{(\beta+1)/\alpha} I(|X_{11}| \ge n^{(\alpha-1)/2}), & \text{if } \alpha > 1 \end{cases}$$
$$\to 0.$$

Therefore, the proof of (B.4.2) further reduces to showing that

$$\max_{j \le Mn^\beta} \left| n^{-\alpha} \sum_{i=1}^{n} (X_{ijk} - \mathrm{E}X_{11k}) \right| \to 0 \ \text{ a.s., as } \ 2^k < n \le 2^{k+1} \to \infty. \tag{B.4.3}$$

For any $\varepsilon > 0$, choose an integer $m = [(\beta+1)/(2\alpha-1)] + 1$. We have

$$\mathrm{P}\left(\max_{j \le Mn^\beta} \left| n^{-\alpha} \sum_{i=1}^{n} (X_{ijk} - \mathrm{E}X_{11k}) \right| \ge \varepsilon, \text{i.o.} \right)$$
$$\le \lim_{N\to\infty} \sum_{k=N}^{\infty} \mathrm{P}\left(\max_{2^k < n \le 2^{k+1}} \max_{j \le M2^{(k+1)\beta}} \left| \sum_{i=1}^{n} (X_{ijk} - \mathrm{E}X_{11k}) \right| \ge \varepsilon 2^{k\alpha} \right)$$
$$\le \lim_{N\to\infty} \sum_{k=N}^{\infty} M2^{(k+1)\beta} \mathrm{P}\left(\max_{n \le 2^{k+1}} \left| \sum_{i=1}^{n} (X_{i1k} - \mathrm{E}X_{11k}) \right| \ge \varepsilon 2^{k\alpha} \right)$$
$$\le M\varepsilon^{-2m} \lim_{N\to\infty} \sum_{k=N}^{\infty} 2^{(k+1)(\beta-2m\alpha)} \mathrm{E}\left| \sum_{i=1}^{2^{k+1}} (X_{i1k} - \mathrm{E}X_{11k}) \right|^{2m}. \tag{B.4.4}$$

Here, the first two inequalities are trivial, while the third follows from an extension of Kolmogorov's inequality that can be found in any textbook on martingales.[1]

By multinomial expansion,

$$\begin{aligned} & \mathrm{E}\left|\sum_{i=1}^{n}(X_{i1k}-\mathrm{E}X_{11k})\right|^{2m} \\ & \leq \sum_{\ell=1}^{m} \sum_{\substack{i_1+\cdots+i_\ell=2m \\ i_1,\cdots,i_\ell\geq 2}} \frac{(2m)!n^\ell}{i_1!\cdots i_\ell!\ell!}\mathrm{E}|X_{11k}-\mathrm{E}X_{11k}|^{i_1}\cdots\mathrm{E}|X_{11k}-\mathrm{E}X_{11k}|^{i_\ell}. \end{aligned}$$

Choose an integer ν such that $0 \leq \nu - 1 \leq (\beta+1)/\alpha < \nu$. For each $i \geq 2$, we have

$$|\mathrm{E}X_{11k}^i| \leq \begin{cases} C\mathrm{E}|X_{11k}|^\nu 2^{(i-\nu)\alpha k}, & \text{if } i \geq \nu, \\ C2^{(i-(\beta+1)/\alpha)\alpha k} \leq C2^{(i\alpha-1)k}, & \text{if } i > (\beta+1)/\alpha, \\ C \leq C2^{(i\alpha-1)k}, & \text{if } i \leq (\beta+1)/\alpha. \end{cases}$$

Then, for some constant C,

$$\begin{aligned} & \mathrm{E}|X_{11k}-\mathrm{E}X_{11k}|^{i_1}\cdots\mathrm{E}|X_{11k}-\mathrm{E}X_{11k}|^{i_\ell} \\ & \leq \begin{cases} C\mathrm{E}|X_{11k}-\mathrm{E}X_{11k}|^\nu 2^{(2m-\nu)\alpha k-(\ell-1)k}, & \text{if } \max\{i_1,\cdots,i_\ell\} \geq \nu, \\ C, & \text{otherwise.} \end{cases} \end{aligned}$$

Hence,

$$\mathrm{E}\left|\sum_{i=1}^{n}(X_{i1k}-\mathrm{E}X_{11k})\right|^{2m} \leq C\mathrm{E}|X_{11k}-\mathrm{E}X_{11k}|^\nu 2^{(2m-\nu)\alpha k+k} + C2^{km}.$$

Substituting this with $n = 2^{k+1}$ into (B.4.4), we have

$$\sum_{k=N}^{\infty} 2^{(k+1)(\beta-2m\alpha+m)} \to 0 \text{ since } \beta - 2m\alpha + m < -1,$$

and

$$\begin{aligned} & \sum_{k=N}^{\infty} 2^{(k+1)(\beta-2m\alpha)}\mathrm{E}|X_{11k}-\mathrm{E}X_{11k}|^\nu 2^{(2m-\nu)\alpha k+k} \\ & \leq C\sum_{k=N}^{\infty} 2^{k(\beta+1-\nu\alpha)}\mathrm{E}|X_{11k}|^\nu \end{aligned}$$

[1] The inequality states that, for any martingale sequence $\{s_n\}$, any constant $\varepsilon > 0$, and $p > 1$, we have $\mathrm{P}(\max_{i\leq n}|s_n| \geq \varepsilon) \leq \varepsilon^{-p}\mathrm{E}|s_n|^p$.

$$
\begin{aligned}
&= C \sum_{k=N}^{\infty} 2^{k(\beta+1-\nu\alpha)} \sum_{\ell=1}^{k} [\mathrm{E}|X_{11}|^{\nu} I(2^{\alpha(\ell-1)} < |X_{11}| \leq 2^{\alpha\ell}) + \mathrm{E}|X_{11}|^{\nu} I(|X_{11}| \leq 1)] \\
&= C \sum_{\ell=1}^{\infty} \mathrm{E}|X_{11}|^{\nu} I(2^{\alpha(\ell-1)} < |X_{11}| \leq 2^{\alpha\ell}) \sum_{k=\ell\vee N} 2^{k(\beta+1-\nu\alpha)} + \sum_{k=N}^{\infty} 2^{k(\beta+1-\nu\alpha)} \\
&\leq C \sum_{\ell=1}^{\infty} 2^{(\ell\vee N)(\beta+1-\nu\alpha)} \mathrm{E}|X_{11}|^{\nu} I(2^{\alpha(\ell-1)} < |X_{11}| \leq 2^{\alpha\ell}) + \sum_{k=N}^{\infty} 2^{k(\beta+1-\nu\alpha)} \\
&\leq C \sum_{\ell=1}^{\infty} 2^{(\ell\vee N)(\beta+1-\nu\alpha)+\ell(\nu\alpha-\beta-1)} \mathrm{E}|X_{11}|^{(\beta+1)/\alpha} I(2^{\alpha(\ell-1)} < |X_{11}| \leq 2^{\alpha\ell}) \\
&\quad + \sum_{k=N}^{\infty} 2^{k(\beta+1-\nu\alpha)} \\
&\leq C\mathrm{E}|X_{11}|^{(\beta+1)/\alpha} I(|X_{11}| \geq 2^{\alpha N/2}) + C\mathrm{E}|X_{11}|^{(\beta+1)/\alpha} 2^{N(\beta+1-\nu\alpha)/2} \\
&\quad + \sum_{k=N}^{\infty} 2^{k(\beta+1-\nu\alpha)} \to 0
\end{aligned}
$$

as $N \to \infty$. Consequently, these, together with (B.4.4), imply that

$$\max_{j \leq Mn^{\beta}} \left| n^{-\alpha} \sum_{i=1}^{n} (X_{ijk} - \mathrm{E}X_{11k}) \right| \to 0, \text{ a.s.}$$

The proof of the sufficiency of the lemma is complete.

Proof of necessity. From (B.4.1), one can easily derive

$$\max_{j \leq M(n-1)^{\beta}} n^{-\alpha} |X_{nj}| \to 0 \quad \text{a.s.},$$

which, together with the Borel-Cantelli lemma, implies that

$$\sum_{n} \mathrm{P}\left(\max_{j \leq M(n-1)^{\beta}} |X_{1j}| \geq n^{\alpha} \right) < \infty.$$

By the convergence theorem for an infinite product, the inequality above is equivalent to the convergence of the product

$$\prod_{n=1}^{\infty} \mathrm{P}\left(\max_{j \leq M(n-1)^{\beta}} |X_{1j}| < n^{\alpha} \right) = \prod_{n=1}^{\infty} \mathrm{P}\left(|X_{11}| < n^{\alpha} \right)^{[M(n-1)^{\beta}]}.$$

Again, using the same theorem, the convergence above is equivalent to the convergence of the series

$$\sum_{n} (n-1)^{\beta} \mathrm{P}\left(|X_{11}| \geq n^{\alpha} \right) < \infty.$$

From this, it is routine to derive $\mathrm{E}|X_{11}|^{(\beta+1)/\alpha} < \infty$. Applying the sufficiency part, the second condition of the lemma follows.

(The divergence). Assume that $\mathrm{E}|X_{11}|^{(1+\beta)/\alpha} = \infty$. Then, for any $N > 0$, we have

$$\sum_{k=1} M 2^{(\beta+1)k} \mathrm{P}(|x_{11}| \geq N 2^{\alpha k}) = \infty.$$

Then, by the convergence theorem of infinite products, the equality above is equivalent to

$$\prod_{k=1}^{\infty} \left(\mathrm{P}(|x_{11}| < N 2^{\alpha k})\right)^{[M 2^{\beta+1)k}]} = 0$$

$$\Longleftrightarrow \sum_{k=1} \mathrm{P}\Big(\max_{2^k < n \leq 2^{K-1}, M2^{(k-1)\beta} < j \leq M2^{k\beta}} |x_{nj}| \geq N 2^{\alpha k}\Big) = \infty.$$

By the Borel-Cantelli lemma, we have

$$\mathrm{P}\left(\max_{2^k < n \leq 2^{K-1}, M2^{(k-1)\beta} < j \leq M2^{k\beta}} |x_{nj}| \geq N 2^{\alpha k}, \text{ i.o.}\right) = 1.$$

Therefore,

$$\limsup \max_{j \leq M n^{\beta}} \left| n^{-\alpha} \sum_{i=1}^{n} (X_{ij} - c) \right| \geq N/2 \quad \text{a.s.}$$

This proves the second conclusion of the lemma. The proof is complete.

B.5 A Lemma on Quadratic Forms

Lemma B.26. *Let* $\mathbf{A} = (a_{ij})$ *be an* $n \times n$ *nonrandom matrix and* $\mathbf{X} = (x_1, \cdots, x_n)'$ *be a random vector of independent entries. Assume that* $\mathrm{E}x_i = 0$, $\mathrm{E}|x_i|^2 = 1$, *and* $\mathrm{E}|x_j|^\ell \leq \nu_\ell$. *Then, for any* $p \geq 1$,

$$\mathrm{E}|\mathbf{X}^*\mathbf{A}\mathbf{X} - \mathrm{tr}\mathbf{A}|^p \leq C_p\Big(\big(\nu_4 \mathrm{tr}(\mathbf{A}\mathbf{A}^*)\big)^{p/2} + \nu_{2p}\mathrm{tr}(\mathbf{A}\mathbf{A}^*)^{p/2}\Big),$$

where C_p *is a constant depending on* p *only.*

In the proof of this lemma, we need a lemma from Dilworth [95].

Lemma B.27. *Let* $\{\mathscr{F}_k\}$ *be a sequence of increasing* σ*-fields and* $\{X_k\}$ *be a sequence of integrable random variables. Then, for any* $1 \leq q \leq p < \infty$, *we have*

$$\mathrm{E}\Bigg(\sum_{k=1}^{\infty} \big|\mathrm{E}(X_k|\mathscr{F}_k)\big|^q\Bigg)^{p/q} \leq \left(\frac{p}{q}\right)^{p/q} \mathrm{E}\Bigg(\sum_{k=1}^{\infty} |X_k|^q\Bigg)^{p/q}.$$

*Proof of Lemma B.*26. We use the expression

$$\mathbf{X}^*\mathbf{A}\mathbf{X} - \mathrm{tr}\mathbf{A} = \sum_{i=1}^{n} a_{ii}(|X_i|^2 - 1) + \sum_{i=1}^{n}\sum_{j=1}^{i-1}(a_{ji}\bar{X}_j X_i + a_{ij}X_j\bar{X}_i). \quad \text{(B.5.1)}$$

By Theorem A.13, it is seen that

$$\begin{aligned}
&\mathrm{E}|\mathbf{X}^*\mathbf{A}\mathbf{X} - \mathrm{tr}\mathbf{A}| \\
&\leq \sum_{i=1}^{n} |a_{ii}|\mathrm{E}\big||X_i|^2 - 1\big| + \left(\mathrm{E}\Big|\sum_{i=1}^{n}\sum_{j=1}^{i-1} a_{ji}\bar{X}_j X_i + a_{ij}X_j\bar{X}_i\Big|^2\right)^{1/2} \\
&\leq C\big[\mathrm{tr}(\mathbf{A}\mathbf{A}^*)^{1/2} + (\nu_4 \mathrm{tr}\mathbf{A}\mathbf{A}^*)^{1/2}\big],
\end{aligned}$$

which proves the lemma for the case $p = 1$. Note that here we have used the fact that $\nu_4 \geq 1$.

Now, assume $1 < p \leq 2$. By Lemma 2.12 and Theorem A.13, we have

$$\begin{aligned}
&\mathrm{E}\left|\sum_{i=1}^{n} a_{ii}(|X_i|^2 - 1)\right|^p \leq C\mathrm{E}\left(\sum_{i=1}^{n} |a_{ii}|^2\big||X_i|^2 - 1\big|^2\right)^{p/2} \\
&\leq C\sum_{i=1}^{n} |a_{ii}|^p \mathrm{E}||X_i|^2 - 1|^p \leq C\nu_{2p}\mathrm{tr}(\mathbf{A}\mathbf{A}^*)^{p/2}.
\end{aligned}$$

Furthermore, by the Hölder inequality,

$$\begin{aligned}
&\mathrm{E}\left|\sum_{i=1}^{n}\sum_{j=1}^{i-1}(a_{ij}\bar{X}_j X_i + a_{ji}X_j\bar{X}_i)\right|^p \\
&\leq C\left[\mathrm{E}\left|\sum_{i=1}^{n}\sum_{j=1}^{i-1}(a_{ij}\bar{X}_j X_i + a_{ji}X_j\bar{X}_i)\right|^2\right]^{p/2} \\
&\leq C(\nu_4 \mathrm{tr}\mathbf{A}\mathbf{A}^*)^{p/2}.
\end{aligned}$$

Combining the two inequalities above, we complete the proof of the lemma for the case $1 < p \leq 2$.

Now, we proceed with the proof of the lemma by induction on p. Assume that the lemma is true for $1 \leq p \leq 2^t$, and then consider the case $2^t < p \leq 2^{t+1}$ with $t \geq 1$. By Lemma 2.13 and Theorem A.13,

$$\begin{aligned}
&\mathrm{E}\left|\sum_{i=1}^{n} a_{ii}(|X_i|^2 - 1)\right|^p \\
&\leq C_p\left(\left(\sum_{i=1}^{n} |a_{ii}|^2 \mathrm{E}(|X_{ii}|^2 - 1)^2\right)^{p/2} + \sum_{i=1}^{n} |a_{ii}|^p \mathrm{E}\big||X_{ii}|^2 - 1\big|^p\right) \\
&\leq C\Big(\big(\nu_4 \mathrm{tr}\mathbf{A}\mathbf{A}^*\big)^{p/2} + \nu_{2p}\mathrm{tr}(\mathbf{A}\mathbf{A}^*)^{p/2}\Big).
\end{aligned}$$

For the same reason, with notation E_i for the conditional expectation given $\{X_1, \cdots, X_i\}$, we have

$$\begin{aligned}
&\mathrm{E}\left|\sum_{i=1}^{n}\sum_{j=1}^{i-1} a_{ij}\bar{X}_j X_i\right|^p \\
&\le C_p\left(\mathrm{E}\left(\sum_{i=1}^{n}\mathrm{E}_{i-1}\left|\sum_{j=1}^{i-1} a_{ij}\bar{X}_i X_j\right|^2\right)^{p/2} + \sum_{i=1}^{n}\mathrm{E}\left|\sum_{j=1}^{i-1} a_{ij}\bar{X}_i X_j\right|^p\right) \\
&\le C_p\left(\mathrm{E}\left(\sum_{i=1}^{n}\left|\sum_{j=1}^{i-1} a_{ij} X_j\right|^2\right)^{p/2} + \sum_{i=1}^{n}\nu_p\mathrm{E}\left|\sum_{j=1}^{i-1} a_{ij} X_j\right|^p\right) \\
&\le C_p\left(\mathrm{E}\left(\sum_{i=1}^{n}\left|\mathrm{E}_{i-1}\sum_{j=1}^{n} a_{ij} X_j\right|^2\right)^{p/2} + \sum_{i=1}^{n}\nu_p\left(\sum_{j=1}^{i-1}|a_{ij}|^2\right)^{p/2} + \nu_p^2\sum_{j=1}^{i-1}|a_{ij}|^p\right) \\
&\le C_p\left(\mathrm{E}\left(\sum_{i=1}^{n}\mathrm{E}_{i-1}\left|\sum_{j=1}^{n} a_{ij} X_j\right|^2\right)^{p/2} + \nu_p\sum_{i=1}^{n}\left((\mathbf{A}\mathbf{A}^*)_{ii}\right)^{p/2} + \nu_{2p}\mathrm{tr}(\mathbf{A}\mathbf{A}^*)^{p/2}\right).
\end{aligned} \tag{B.5.2}$$

Using Lemma B.27 with $q = 1$ applied to the first term and the induction hypothesis with $\mathbf{A}$ replaced by $\mathbf{A}^*\mathbf{A}$, we obtain

$$\begin{aligned}
&\mathrm{E}\left(\sum_{i=1}^{n}\mathrm{E}_{i-1}\left|\sum_{j=1}^{n} a_{ij} X_j\right|^2\right)^{p/2} \\
&\le C_p\mathrm{E}\left(\sum_{i=1}^{n}\left|\sum_{j=1}^{n} a_{ij} X_j\right|^2\right)^{p/2} = C_p\mathrm{E}(\mathbf{X}^*\mathbf{A}^*\mathbf{A}\mathbf{X})^{p/2} \\
&\le C_p\left((\mathrm{tr}\mathbf{A}^*\mathbf{A})^{p/2} + \mathrm{E}\left|\mathbf{X}^*\mathbf{A}^*\mathbf{A}\mathbf{X} - \mathrm{tr}\mathbf{A}^*\mathbf{A}\right|^{p/2}\right) \\
&\le C_p\left((\mathrm{tr}\mathbf{A}^*\mathbf{A})^{p/2} + [\nu_4\mathrm{tr}(\mathbf{A}^*\mathbf{A})^2]^{p/4} + \nu_p\mathrm{tr}(\mathbf{A}^*\mathbf{A})^{p/2}\right).
\end{aligned}$$

Note that $\mathrm{tr}(\mathbf{A}^*\mathbf{A})^2 \le (\mathrm{tr}\mathbf{A}^*\mathbf{A})^2$ and $\sum_{i=1}^{n}\left((\mathbf{A}\mathbf{A}^*)_{ii}\right)^{p/2} \le \mathrm{tr}(\mathbf{A}^*\mathbf{A})^{p/2}$ by Lemma A.13. Substituting the above into (B.5.2), the proof of the lemma is complete.

Relevant Literature

1. Agrell, E., Eriksson, T., Vardy, A., and Zeger, K. (2002). Closest point search in lattices. *IEEE Trans. Inf. Theory* **48**(8), 2201–2214.
2. Anderson, G. W. and Zeitouni, O. (2006). A CLT for a band matrix model. *Probab. Theory Relat. Fields* **134**, 283–338.
3. Albeverio, S., Pastur, L., and Shcherbina, M. (2001). On the $1/n$ expansion for some unitary invariant ensembles of random matrices. Dedicated to Joel L. Lebowitz. *Comm. Math. Phys.* **224**(1), 271–305.
4. Andersson, Per-Johan, Öberg, Andreas, and Guhr, Thomas (2005). Power mapping and noise reduction for financial correlations. *Acta Physica Polonica B* **36**(9), 2611–2619.
5. Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, third edition. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.
6. Angls d'Auriac, J. Ch. and Maillard, J. M. (2003). Random matrix theory in lattice statistical mechanics. Statphys-Taiwan-2002: Lattice models and complex systems (Taipei/Taichung). *Physica A* **321**(1–2), 325–333.
7. Arnold, L. (1971). On Wigner's semicircle law for the eigenvalues of random matrices. *Z. Wahrsch. Verw. Gebiete* **19**, 191–198.
8. Arnold, L. (1967). On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.* **20**, 262–268.
9. Aubrun, Guillaume (2005). A sharp small deviation inequality for the largest eigenvalue of a random matrix. In*Seminaire de Probability XXXVIII*, Lecture Notes in Mathematics 1857, M. Émery, M. Ledous, and M. Yor, editiors, 320–337. Springer, Berlin.
10. Bai, J. (2003). Inferential theory for factor models of large dimensions, *Econometrica* **71**, 135–173.
11. Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *J. Business Econ. Statist.* **25**(1), 52–60.
12. Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.
13. Bai, Z. D. (1999). Methodologies in spectral analysis of large dimensional random matrices: A review. *Statistica Sinica* **9**(3), 611–677.
14. Bai, Z. D. (1997). Circular law. *Ann. Probab.* **25**(1), 494–529.
15. Bai, Z. D. (1995). Spectral analysis of large dimensional random matrices. *J. Chinese Statist. Assoc.* **33**(3), 299–317.
16. Bai, Z. D. (1993a). Convergence rate of expected spectral distributions of large random matrices. Part I. Wigner matrices. *Ann. Probab.* **21**(2), 625–648.
17. Bai, Z. D. (1993b). Convergence rate of expected spectral distributions of large random matrices. Part II. Sample covariance matrices. *Ann. Probab.* **21**(2), 649–672.

18. Bai, Z. D., Jiang, Dandan, Yao, Jian-feng, and Zheng Shurong (2009). Corrections to LRT on large dimensional covariance matrix by RMT. To appear in *Ann. Statist.* arXiv:0902.0552[stat]
19. Bai, Z. D., Liang, W. Q., and Vervaat, W. (1987). Strong representation of weak convergence. *Technical report 186*. Center for Stochastic Processes, University of North Carolina.
20. Bai, Z. D., Liu, H. X., and Wong, W. K. (2009). Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory. To appear in *Math. Finance*.
21. Bai, Z. D., Miao, B. Q., and Jin, B. S. (2007). On limit theorem for the eigenvalue of product of two random matrices. *J. Multivariate Anal.* **98**, 76–101.
22. Bai, Z. D., Miao, B. Q., and Pan, G. M. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab.* **35**(4), 1532–1572.
23. Bai, Z. D., Miao, B. Q., and Tsay, J. (2002). Convergence rates of the spectral distributions of large Wigner matrices. *Int. Math. J.* **1**(1), 65–90.
24. Bai, Z. D., Miao, B. Q., and Tsay, J. (1999). Remarks on the convergence rate of the spectral distributions of Wigner matrices. *J. Theoret. Probab.* **12**(2), 301–311.
25. Bai, Z. D., Miao, B. Q., and Tsay, J. (1997). A note on the convergence rate of the spectral distributions of large dimensional random matrices. *Statist. Probab. Lett.* **34**(1), 95–102.
26. Bai, Z. D., Miao, B. Q., and Yao, J. F. (2003). Convergence rates of spectral distributions of large sample covariance matrices. *SIAM J. Matrix Anal. Appl.* **25**(1), 105–127.
27. Bai, Z. D. and Saranadasa, H. (1996). Effect of high dimension comparison of significance tests for a high dimensional two sample problem. *Statistica Sinica* **6**(2), 311–329.
28. Bai, Z. D. and Silverstein, J. W. (2007). On the signal-to-interference-ratio of CDMA systems in wireless communications. *Ann. Appl. Probab.* **17**(1), 81–101.
29. Bai, Z. D. and Silverstein, J. W. (2006). *Spectral Analysis of Large Dimensional Random Matrices*. First edition, Science Press, Beijing.
30. Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32**(1), 553–605.
31. Bai, Z. D. and Silverstein, J. W. (1999). Exact separation of eigenvalues of large dimensional sample covariance matrices. *Ann. Probab.* **27**(3), 1536–1555.
32. Bai, Z. D., and Silverstein, J. W., (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26**(1), 316–345.
33. Bai, Z. D., Silverstein, J. W., and Yin, Y. Q. (1988). A note on the largest eigenvalue of a large dimensional sample covariance m. *J. Multivariate Anal.* **26**, 166–168.
34. Bai, Z. D. and Yao, J. F. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Statist.* **44**(3), 447–474.
35. Bai, Z. D. and Yao, J. F. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11**(6), 1059–1092.
36. Bai, Z. D. and Yin Y. Q. (1993). Limit of the smallest eigenvalue of large dimensional covariance matrix. *Ann. Probab.* **21**(3), 1275–1294.
37. Bai, Z. D. and Yin, Y. Q. (1988a). A convergence to the semicircle law. *Ann. Probab.* **16**(2), 863–875.
38. Bai, Z. D. and Yin, Y. Q. (1988b). Necessary and sufficient conditions for the almost sure convergence of the largest eigenvalue of Wigner matrices. *Ann. Probab.* **16**(4), 1729–1741.
39. Bai, Z. D. and Yin, Y. Q. (1986). Limiting behavior of the norm of products of random matrices and two problems of Geman-Hwang. *Probab. Theory. Relat. Fields* **73**, 555–569.
40. Bai, Z. D., Yin, Y. Q., and Krishnaiah, P. R. (1987). On the limiting empirical distribution function of the eigenvalues of a multivariate F matrix. *Theory Probab. Appl.* **32**, 537–548.

41. Bai, Z. D., Yin, Y. Q., and Krishnaiah, P. R. (1986). On limiting spectral distribution of product of two random matrices when the underlying distribution is isotropic. *J. Multivariate Anal.* **19**, 189–200.
42. Bai, Z. D. and Zhang, L. X. (2006). Semicircle law for Hadamard products. *SIAM J. Matrix Anal. Appl.* **29** (2), 473–495.
43. Baik, J., Ben Arous, G., and Péché, S. (2005). Phase transition of the largest eigenvalue for non-null complex sample covariance matrices. *Ann. Probab.* **33**, 1643–1697.
44. Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97**(6), 1382–1408.
45. Barry, R. P. and Pace, R. K. (1999). Monte Carlo estimates of the log determinant of large sparse matrices. Linear algebra and statistics (Istanbul, 1997). *Linear Algebra Appl.* **289**(1–3), 41–54.
46. Baum, K. L., Thomas, T. A., Vook, F. W., and Nangia, V. (2002). Cyclic-prefix CDMA: an improved transmission method for broadband DS-CDMA cellular systems, *Proc. IEEE WCNC'2002* **1**, 183–188.
47. Beheshti, S. Isabelle, S. H., and Wornel, G. W. (1998). Joint intersymbol and multiple access interference suppression algorithms for CDMA systems. *Eur. Trans. Telecommun.* **9**, 403–418.
48. Bekakos, M. P. and Bartzi, A. A. (1999). Sparse matrix and network minimization schemes. In *Computational Methods and Neural Networks,* M. P. Bekakos and M. Sambandham, editors, 61–94. Dynamic, Atlanta, GA.
49. Bellman, Richard Ernest (1957). *Dynamic Programming.* Princeton University Press, Princeton, NJ.
50. Ben-Hur, A., Feinberg, J., Fishman, S., and Siegelmann, H. T. (2004). Random matrix theory for the analysis of the performance of an analog computer: a scaling theory. *Phys. Lett. A* **323**(3–4), 204–209.
51. Bernanke, B. S., Boivin, J., and Eliasz, P. (2004). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. NBER Working Paper 10220.
52. Biane, P. (1998). Free probability for probabilists, MSRI Preprint No. 1998–40.
53. Biely, C. and Thurner, S. (2006). Random matrix ensemble of time-lagged correlation matrices: derivation of eigenvalue spectra and analysis of financial time-series. *arXiv: physics/0609053.* **I**, 7.
54. Biglieri, E., Caire, G., and Taricco, G. (2001). Limiting performance of block-fading channels with multiple antennas. *IEEE Trans. Inf. Theory* **47**(4), 1273–1289.
55. Biglieri, E., Proakis, J., and Shamai, S. (1998). Fading channels: information-theoretic and communications aspects. *IEEE Trans. Inf. Theory* **44**(6), 2619–2692.
56. Billingsley, P. (1995). *Probability and Measure*, third edition. Wiley Series in Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York.
57. Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.
58. Black, F. (1972). Capital market equilibrium with restricted borrowing. *J. Business* **45**, 444–454.
59. Bleher, P. and Kuijlaars, A. B. J. (2004). Large n limit of Gaussian random matrices with external source. I. *Comm. Math. Phys.* **252**(1–3), 43–76.
60. Bliss, D. W., Forsthe, K. W., and Yegulalp, A. F. (2001). MIMO communication capacity using infinite dimension random matrix eigenvalue distributions. *Thirty-Fifth Asilomar Conf. Signals, Systems, Computers* **2**, 969–974.
61. Boley, D. and Goehring, T. (2000). LQ-Schur projection on large sparse matrix equations. Preconditioning techniques for large sparse matrix problems in industrial applications (Minneapolis, MN, 1999). *Numer. Linear Algebra Appl.* **7**(7–8), 491–503.
62. Botta, E. F. F. and Wubs, F. W. (1999). Matrix renumbering ILU: an effective algebraic multilevel ILU preconditioner for sparse matrices. Sparse and structured matrices and their applications (Coeur d'Alene, ID, 1996). *SIAM J. Matrix Anal. Appl.* **20**(4), 1007–1026.

63. Breitung, Jörg and Eickmeier, Sandra (2005). Dynamic factor models. Discussion Paper Series 1: Economics Studies. N038/2005. Deutsche Bundesbank.
64. Bryan, M. F. and Cecchetti, S. G. (1994). Measuring core inflation. In *Monetary Policy.* N.G. Mankiw, editor, 195–220. University of Chicago Press, Chicago.
65. Bryan, M. F. and Cecchetti, S. G. (1993). The consumer price index as a measure of inflation. *Federal Reserve Bank Cleveland Econ. Rev.* **29**, 15–24.
66. Burda, Zdzisław and Jurkiewicz, Jerzy (2004). Signal and noise in financial correlation matrices. *Physica A* **344**(1–2), 67–72.
67. Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probab.* **1**, 19–42.
68. Burns A. F. and Mitchell W. C. (1946). *Measuring Business Cycle.* National Bureau of Economic Research, New York.
69. Caire, C., Taricco, G., and Biglieri, E. (1999). Optimal power control over fading channels, *IEEE Trans. Inf. Theory,* **45**(5), 1468–1489.
70. Caire, G., Müller, R. R., and Tanaka, T. (2004). Iterative multiuser joint decoding: optimal power allocation and low-complexity implementation, *IEEE Trans. Inf. Theory* **50**(9), 1950–1973.
71. Camba-Méndez, G. and Kapetanios, G. (2004). Forecasting Euro Area inflation using dynamic factor measures of underlying inflation. Working paper series No. 402.
72. Campbell, J. Y., Lo, A. W., and Macknlay, A. C. (1997). *The Econometrics of Financial Markets.* Princeton University Press, Princeton, NJ.
73. Chamberlain, G. and Rothschild, M. (1983). Aribitrage factor structure and mean-variance analysis on large asset markets. *Econometrica* **51**, 1281–1324.
74. Caselle, M. and Magnea, U. (2004). Random matrix theory and symmetric spaces. *Phys. Rep.* **394**(2–3), 41–156.
75. Chan, A. M. and Wornell, G. W. (2002). Approaching the matched-filter bound using iterated-decision equalization with frequency-interleaved encoding. *Proc. IEEE Globecom'02,* **1**, 17–21.
76. Chan, A. M. and Wornell, G. W. (2001). A class of block-iterative equalizers for intersymbol interference channels: fixed channel results. *IEEE Trans. Commun.* **49**(11), 1966–1976.
77. Chatterjee, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Relat. Fields* **143**(1–2), 1–40.
78. Chuah, C. N., Tse, D., Kahn, J. M., and Valenzuels, R. A. (2002). Capacity scaling in MIMO wireless systems under correlated fading. *IEEE Trans. Inf. Theory* **48,** 637–650.
79. Chung, K. L. (1974). *A Course in Probability Theory*, second edition. Probability and Mathematical Statistics 21. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London.
80. Cioffi, J. M. (2003). *Digital Communications.* Stanford University Course Reader.
81. Cipollini, A. and Kapetanios, G. (2008). Forecasting financial crises and contagion in Asia using dynamic factor analysis. Working Paper, No. 538. Queen Mary University of London.
82. Conlon, T., Ruskin, H. J., and Crane, M. (2007). Random matrix theory and fund of funds portfolio optimisation. *Physica A* **382**(2), 565–576.
83. Cottatellucci, L. and Müller, R. (2007). CDMA systems with correlated spatial diversity: a generalized resource pooling result. *IEEE Trans. Inf. Theory* **53**(3), 1116–1136.
84. Cover, T. and Thomas, J. (1991). *Elements of Information Theory.* Wiley, New York.
85. Damen, M. O., Gamal, H. E., and Caire, G. (2003). On maximum-likelihood detection and the search for the closest lattice point. *IEEE Trans. Inf. Theory* **49**(10), 2389–2402.
86. Debbah, M., Hachem, W., Loubaton, P., and Courvill, M. (2003). MMSE analysis of certain large isometric random precoded systems. *IEEE Trans. Inf. Theory* **49**(5), 1293–1311.
87. Deift, P. (2003). Four lectures on random matrix theory. In *Asymptotic combinatorics with applications to mathematical physics* (St. Petersburg, 2001), Lecture Notes in Mathematics 1815, A. M. Vershik, editor, 21–52. Springer, Berlin.

88. Deift, P. A. (1999). *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach.* Courant Lecture Notes in Mathematics 3. New York University, Courant Institute of Mathematical Sciences, New York, American Mathematical Society, Providence, RI.
89. Del Corso, G. M. and Romani, F. (2001). Heuristic spectral techniques for the reduction of bandwidth and work-bound of sparse matrices. In memory of W. Gross. *Numer. Algorithms* **28**(1–4), 117–136.
90. Dembo, A., Guionnet, A., and Zeitouni, O. (2003). Moderate deviations for the spectral measure of certain random matrices. *Ann. Inst. H. Poincare Probab. Statist.* **39**(6), 1013–1042.
91. Dempster A. P. (1958). A high dimensional two sample significance test. *Ann. Math. Statist.* **29**, 995–1010.
92. des Cloizeaux, J. and Mehta, M. L. (1973). Asymptotic behavior of spacing distributions for the eigenvalues of random matrices. *J. Math. Phys.* **14**, 1648–1650.
93. Diaconis, P. and Evans, S. N. (2002). A different construction of Gaussian fields from Markov chains: Dirichlet covariances. *Ann. Inst. Henri Poincare B: Probab. Statist.* **38**(6), 863–878.
94. Diaconis, P. and Evans, S. N. (2001). Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.* **353**(7), 2615–2633.
95. Dilworth, S. J. (1993). Some probabilistic inequalities with applications to functional analysis. In *Banach Spaces: Contemporary Mathematics*, 144, B. L. Lin and W. B. Johnson, editors, 53–67. AMS, Providence, RI.
96. Dudley, R. M. (1985). An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions. In *Probability in Banach Spaces*, Lecture Notes in Mathematics 1153, A. Beck, R. Dudley, M. Hahn, J.Kuelbs, and M. Marcus, editors, 141–178. Springer, Berlin.
97. Dumitriu, I. and Edelman, A. (2002). Matrix models for beta ensembles. *J. Math. Phys.* **43**(11), 5830–5847.
98. Edelman, A. (1997). The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law. *J. Multivariate Anal.* **60**(2), 203–232.
99. Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**(1), 1–26.
100. Eldar, Y. C. and Chan, A. M. (2003). On the asymptotic performance of the decorrelator. *IEEE. Trans. Inf. Theory* **49**, 2309–2313.
101. El Karoui, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35**(2), 663–714.
102. Engle, R. F., and Gearger, C. W. J. (1987). Co-integration and error correction representaton, estimation and testing. *Econometrica* **55**(2), 251–276.
103. Fan, K. (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. Nat. Acad. Sci. U.S.A.* **37**, 760–766.
104. Fan, K. (1949). On a theorem of Weyl concerning eigenvalues of linear transformations. I. *Proc. Nat. Acad. Sci. U.S.A.* **35,** 652–655.
105. Form, M. and Lippi, M. (1999). Aggregation of linear dynamic microeconomic models. *J. Math. Ecn.* **31**, 131–158.
106. Forni, M., Giamond, D., Lippi, M., and Reichlin, L. (2004). Opening the black box: structural factor models versus structural VARS. CEPR Discussion Papers No. 4133.
107. Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model one-side estimation and forecasting. *J. Amer. Statist. Assoc.* **100**, 830–840.
108. Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2004). The generalized dynamic factor model: consistency and rates. *J. Economet.* **119**, 231–255.
109. Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2003). Do financial variables help forecasting inflation and read activity in the Euro area? *J. Monetary Econ.* **50**, 1243–1255.

110. Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2001). Coincident and leading indicators for the Euro area. *Econ. J.* **111**, 62–85.
111. Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic factor model: identification and estimation. *Rev. Econ. Statist.* **82**, 540–554.
112. Forni, M. and Reichlin, L. (1998). Let's get real: a factor analytical approach to disaggregated business cycle dynamic. *Rev. Econ. Studies* **65**, 453–473.
113. Foschini, G. J. (1996). Layered space-time architechture for wireless communications in a fading environment when using multi-element antennas. *Bell Labs Tech. J.* **1**(2), 41–59.
114. Foschini, G. J. and Gans, M. (1998). On limits of wireless communications in fading environment when using multiple antennas. *Wireless Personal Commun.* **6**(3), 315–335.
115. Gadzhiev, Ch. M. (2004). Determination of the upper bound for the spectral norm of a random matrix and its application to the diagnosis of the Kalman filter. (Russian). *Kibernet. Sistem. Anal.* **40**(2), 72–80.
116. Gamburd, A., Lafferty, J., and Rockmore, D. (2003). Eigenvalue spacings for quantized cat maps. Random matrix theory. *J. Phys. A* **36**(12), 3487–3499.
117. Geman, S. (1986). The spectral radius of large random matrices. *Ann. Probab.* **14**(4), 1318–1328.
118. Geman, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.* **8**, 252–261.
119. Gerlach, D. and Paulraj, A. (1994). Adaptive transmitting antenna arrays with feedback. *IEEE Signal Process. Lett.* **1**(10), 150–152.
120. Ginibre, J. (1965). Statistical ensembles of complex, quaternion and real matrices. *J. Math. Phys.* **6**, 440–449.
121. Girko, V. L. (1990). *Theory of Random Determinants.* Translated from the Russian. Mathematics and Its Applications (Soviet Series) 45. Kluwer Academic Publishers Group, Dordrecht.
122. Girko, V. L. (1989). Asymptotics of the distribution of the spectrum of random matrices. (Russian). *Uspekhi Mat. Nauk* **44**(4), 7–34, translation in *Russian Math. Surv.* **44**(4), 3–36.
123. Girko, V. L. (1984a). Circle law. *Theory Probab. Appl.* **4**, 694–706.
124. Girko, V. L. (1984b). On the circle law. *Theory Probab. Math. Statist.* **28**, 15–23.
125. Girko, V. L., Kirsch, W., and Kutzelnigg, A. (1994). A necessary and sufficient condition for the semicircular law. *Random Oper. Stoch. Equ.* **2**, 195–202.
126. Girshick, M. A. (1939). On the sampling theory of roots of determinantal equations. *Ann. Math. Statist.* **10**, 203–224.
127. Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Reading, MA.
128. Goldberg, G. and Neumann, M. (2003). Distribution of subdominant eigenvalues of matrices with random rows. *SIAM J. Matrix Anal. Appl.* **24**(3), 747–761.
129. Goldberger, A. S. (1972). Maximum-likelihood estimation of regression containing un-observable independent variables. *Int. Econ. Rev.* **13**(1), 1–15.
130. Goldsmith, A. J. and Varaiya, P. P. (1997). Capacity of fading channels with channel side information, *IEEE Trans Inf. Theory* **43**(6), 1986–1992.
131. Götze, F. and Tikhomirov, A. (2007). On the circular law. *arXiv:math/0702386v1.*
132. Götze, F., Tikhomirov, A., and Timushev, D. A. (2007). Rate of convergence to the semi-circle law for the deformed Gaussian unitary ensemble. *Cent. Eur. J. Math.* **5**(2), 305–334.
133. Götze, F. and Tikhomirov, A. (2004). Limit theorems for spectra of positive random matrices under dependence. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. Veroyatn. i Stat. (POMI)* **311**(7), 92–123.
134. Götze, F. and Tikhomirov, A. (2003) Rate of convergence to the semi-circular law. *Probab. Theory Relat. Fields* **127**, 228–276.
135. Grant A. J. and Alexander, P. D. (1998). Random sequence multisets for synchronous code-division multiple-access channels. *IEEE Trans. Inf. Theory,* **44**(7), 2832–2836.

136. Grenander, U. (1963). *Probabilities on Algebraic Structures.* John Wiley & Sons, Inc., New York-London; Almqvist & Wiksells, Stockholm-Göteborg-Uppsala.
137. Grenander, U. and Silverstein, J. W. (1977). Spectral analysis of networks with random topologies. *SIAM J. Appl. Math.* **32**(2), 499–519.
138. Grönqvist, J., Guhr, T., and Kohler, H. (2004). The k-point random matrix kernels obtained from one-point supermatrix models. *J. Phys. A* **37**(6), 2331–2344.
139. Guhr, T. and Kälber, B. (2003). A new method to estimate the noise in financial correlation matrices. *J. Phys. A.* **36**, 3009–3032.
140. Guionnet, A. (2004). Large deviations and stochastic calculus for large random matrices. *Probab. Surv.* **1**, 72–172 (electronic).
141. Guo, D. and Verdú, S. (2005). Randomly spread CDMA: asymptotics via statistical physics. *IEEE Trans. Inf. Theory* **51**(6), 1982–2010.
142. Guo, D., Verdú, S., and Rasmussen, L. K. (2002). Asymptotic normality of linear multiuser receiver outputs, *IEEE Trans. Inf. Theory* **48**(12), 3080–3095.
143. Gupta, A. (2002). Improved symbolic and numerical factorization algorithms for unsymmetric sparse matrices. *SIAM J. Matrix Anal. Appl.* **24**(2), 529–552.
144. Khorunzhy, A. and Rodgers, G. J. (1997). Eigenvalue distribution of large dilute random matrices. *J. Math. Phys.* **38**, 3300–3320.
145. Halmos, P. R. (1950). *Measure Theory.* D. Van Nostrand Company, Inc., New York.
146. Hanly, S. and Tse, D. (2001). Resource pooling and effective bandwidths in CDMA networks with multiuser receivers and spatial diversity. *IEEE Trans. Inf. Theory* **47**, 1328–1351.
147. Hara, S. and Prasad, R. (1997). Overview of multicarrier CDMA. *IEEE Commun. Mag.* **35**(12), 126–133.
148. Haykin, S. (2005). Cognitive radio: brain-empowered wireless communications, *IEEE J. Selected Areas Commun.* **23**(2), 201–220.
149. Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Am. Statist. Assoc.* **58**, 13–30.
150. Hoeffding, W. (1940). Masstabinvariante Korrelations-theorie. *Schr. Math. Inst. Univ. Berlin* **5**, 181–233.
151. Honig, M. L. and Ratasuk, R. (2003). Large-system performance of iterative multiuser decision-feedback detection. *IEEE Trans. Commun.* **51**(8), 1368–1377.
152. Honig, M. L. and Xiao, W. M. (2001). Performance of reduced-rank linear interference suppression, *IEEE Trans. Inf. Theory* **47**(5), 1928–1946.
153. Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis.* Cambridge University Press, Cambridge.
154. Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis.* Cambridge University Press, Cambridge.
155. Hsu, P. L. (1939). On the distribution of roots of certain determinantal equations. *Ann. Eugenics* **9**, 250–258.
156. Huber, Peter J. (1975). Robustness and designs. A survey of statistical design and linear models. *Proceeding of an International Symposium, Colorado State University, Ft. Collins Colorado*, J.N. Srivastava, editor, 287–301. North-Holland, Amsterdam.
157. Huber, Peter J. (1973). The 1972 Wald Memorial Lecture. Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1**, 799–821.
158. Huber, Peter J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73–101.
159. Hwang, C. R. (1986). A brief survey on the spectral radius and the spectral distribution of large dimensional random matrices with iid entries. In *Random Matrices and Their Applications: Contemporary Mathematics* 50, J. E. Cohen, H. Kesten, and C. M. Newman, editors, 145–152. AMS, Providence, RI.
160. Ingersol, J. (1984). Some results in the theory of arbitrage pricing. *J. Finance* **39**, 1021–1039.
161. Issler, J. V. and Issler, F. V. (2001). Common cycles and the importance of transitory shocks to macroeconomic aggregates. *J. Monetary Econ.* **47**, 449–475.

162. Janik, R. A. (2002). New multicritical random matrix ensembles. *Nuclear Phys. B* **635**(3), 492–504.
163. Jiang, Tiefeng (2004). The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā* **66**(1), 35–48.
164. Johansson, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* **209**, 437–476.
165. Johansson, K. (1998). On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.* **91**(1), 151–204.
166. Johnstone, I. M. (2008). Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy-Widom limits and rates of convergence. *Ann. Statist.* **36**(6), 2638–2716.
167. Johnstone, I. M. (2007). High-dimensional statistical inference and random matrices. *Proc. Int. Congress Math.* **I**, 307–333.
168. Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29**(2), 295–327.
169. Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12**, 1–38.
170. Kamath, M. A. and Hughes, B. L. (2005). The asymptotic capacity of multiple-antenna Rayleigh-fading channels. *IEEE Trans. Inf. Theory* **51**(12), 4325–4333.
171. Kapetanios, G. (2004). A note on modelling core inflation for the UK using a new dynamic factor estimation method and a large disaggregated price index dataset. *Econ. Lett.* **85**, 63–69.
172. Kapetanios, G. and Marcellino, M. (2006). A parametric estimation method for dynamic factor models of large dimensions. CEPR Discussion Paper No. 5620. Available at SSRN: http://ssrn.com/abstract=913392.
173. Keating, J. P. (2002). Random matrices and the Riemann zeta-function. In *Highlights of Mathematical Physics* (London), A. Fokas, J. Halliwell, T. Kibble, and B. Zegarlinski, editors, 153–163. AMS, Providence, RI.
174. Keating, J. P. and Mezzadri, F. (2004). Random matrix theory and entanglement in quantum spin chains. *Comm. Math. Phys.* **252**(1–3), 543–579.
175. Keeling, K. B. and Pavur, R. J. (2004). A comparison of methods for approximating the mean eigenvalues of a random matrix. *Comm. Statist. Simulation Comput.* **33**(4), 945–961.
176. Khorunzhy, A. and Rodgers, G. J. (1998). On the Wigner law in dilute random matrices. *Rept. Math. Phys.* **42**, 297–319.
177. Khorunzhy, A. and Rodgers, G. J. (1997). Eigenvalue distribution of large dilute random matrices. *J. Math. Phys.* **38**, 3300–3320.
178. Krzanowski, W. J. (1984). Sensitivity of principal components. *J.R. Statist. Soc. B* **46**(3), 558–563.
179. Kuhn, K. W. and Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neymann, editor, 481–492. University of California Press, Berkeley and Los Angeles.
180. Lahiri, I. K., and Moore, G. H., eds. (1991). Leading economic indicators. In *New Approches and Forecasting Records*, 63–89. Cambridge University Press, Combridge.
181. Laloux, L., Cizeau, P., Bouchaud, J. P., and Potters, M. (1999). Noise dressing of financial correlation matrices. *Phys. Rev. Lett.* **83**, 1467–1470.
182. Laloux, L., Cizeau, P., Bouchaud, J. P., and Potters., M. (1999). Random matrix theory applied to portfolio optimization in Japanese stock market. *Risk* **12**(3), 69.
183. Laloux, L., Cizeau, P., Potters, M., and Bouchaud, J. (2000). Random matrix theory and financial correlations. *Int. J. Theor. Appl. Finance* **3**, 391–397.
184. Latała, R. (2005). Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.* **133**(5), 1273–1282.
185. Le Car, G. and Delannay, R. (2003). The distributions of the determinant of fixed-trace ensembles of real-symmetric and of Hermitian random matrices. *J. Phys. A* **36**(38), 9885–9898.

186. Letac, G. and Massam, H. (2001). The normal quasi-Wishart distribution. In *Algebraic methods in statistics and probability: Contemporary Mathematics* 287, M. A. .G. Viana and D. P. Richards, editors, 231–239. AMS, Providence, RI
187. Li, L., Tulino, A. M., and Verdú, S. (2004). Design of reduced-rank MMSE multiuser detectors using random matrix methods. *IEEE Trans. Inf. Theory* **50**(6), 986–1008.
188. Li, P., Paul, D., Narasimhan, R., and Cioffi, J. (2006). On the distribution of SINR for the MMSE MIMO receiver and performance analysis. *IEEE Trans. Inf. Theory* **52**(1), 271–286.
189. Liang, Y. C. (2006). Asymptotic performance of BI-GDFE for large isometric and random precoded systems. *Proceedings of IEEE VTC-Spring,* Stockholm, Sweden.
190. Liang, Y. C., Chen, H. H., Mitolla, J., Mahonen III, P., Kohno, R., and Reed, J. H. (2008). Guest Editorial, Cognitive Radio: Theory and Application, *IEEE J. Selected Areas Commun.* **26**(1), 1–4.
191. Liang, Y. C., Cheu, E. Y., Bai, L., and Pan, G. (2008). On the relationship between MMSE-SIC and BI-GDFE receivers for large multiple-input multiple-output (MIMO) channels, *IEEE Trans. Signal. Proc.*, **56**(8), 3627–3637.
192. Liang, Y. C. and Chin, F. (2001). Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems with antenna array. *IEEE J. Selected Areas Commun.* **19**(2), 222–232.
193. Liang, Y. C., Chin, F., and Liu, K. J. R. (2001). Downlink beamforming for DSCDMA mobile radio with multimedia services. *IEEE Trans. Commun.* **49**(7), 1288–1298.
194. Liang, Y. C., Pan, G. M., and Bai, Z. D. (2007). Asymptotic performance analysis of MMSE receivers for large MIMO systems using random matrix theory. *IEEE Trans. Inf. Theory* **53**(11), 4173–4190.
195. Liang, Y. C., Sun, S., and Ho, C. K. (2006). Block-iterative generalized decision feedback equalizers (BI-GDFE) for large MIMO systems: algorithm design and asymptotic performance analysis. *IEEE Trans. Signal Process.* **54**(6), 2035–2048.
196. Liang, Y. C., Zhang, R., and Cioffi, J. M. (2006). Subchannel grouping and statistical water-filling for vector block fading channels. *IEEE Trans. Commun.* **54**(6), 1131–1142.
197. Lin, Y. (2001). Graph extensions and some optimization problems in sparse matrix computations. *Adv. Math.* **30**(1), 9–21.
198. Lindvall, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0,\infty)$. *J. Appl. Probab.* **10**, 109–121.
199. Litvak, A. E., Pajor, A., Rudelson, M., and Tomczak-Jaegermann, N. (2005). Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.* **195**, 491–523.
200. Loève, M. (1977). *Probability Theory,* fourth edition. Springer-Verlag, New York.
201. Marčenko, V. A. and Pastur, L. A. (1967). Distribution for some sets of random matrices. *Math. USSR-Sb.* **1**, 457–483.
202. Marcinkiewicz, J. and Zygmund, A. (1937). Sur les fonctions indpéndantes. *Fund. Math.* **29**, 60–90.
203. Markowitz, H. (1991). *Portfolio Selection: Efficient Diversification of Investments*, second edition. Blackwell Publishing, Oxford.
204. Markowitz, H. (1956). The optimization of a quadratic function subject to linear constraints. *Naval Res. Logistics Quarterly* **3**, 111–133.
205. Markowitz, H. (1952). Portpolio selection. *J. Finance* **7**(1), 77–91.
206. Martegna, R. N. (1999). Hierarchical structure in financial market. *Eur. Phys. J. B* **11**, 193–197.
207. Marti, R., Laguna, M., Glover, F., and Campos, V. (2001). Reducing the bandwidth of a sparse matrix with tabu search. Financial modelling. *Eur. J. Oper. Res.* **135**(2), 450–459.
208. Masaro, J. and Wong, Ch. S. (2003). Wishart distributions associated with matrix quadratic forms. *J. Multivariate Anal.* **85**(1), 1–9.
209. Matthew, C. H. (2007). *Essays in Econometrics and Random Matrix Theory.* Ph.D. Thesis, MIT.

210. Mayya, K. B. K. and Amritkar, R. E. (2006). Analysis of delay correlation matrices. *oar: arXiv.org. cond-mat/0601279*.
211. McKenzie, B. J. and Bell, T. (2001). Compression of sparse matrices by blocked Rice coding. *IEEE Trans. Inf. Theory* **47**(3), 1223–1230.
212. Mehta, M. L. (2004). *Random Matrices*, third edition. Elsevier/Academic Press, Amsterdam.
213. Mehta, M. L. (1960). On the statistical properties of the level-spacings in nuclear spectra. *Nuclear Phys.* **18**, 395–419.
214. Mitola, J. and Maguire, G. Q. (1999). Cognitive radios: making software radios more personal. *IEEE Personal Commun.* **6**(4), 13–18.
215. Müller, R. R. (1998). *Power and Bandwidth Efficiency of Multiuser Systems with Random Spreading.* Ph.D. Thesis, Universtät Erlanger-Nürnberg, Erlangen, Germany.
216. Naulin, J. M. (2002). A contribution of sparse matrices tools to matrix population model analysis. Deterministic and stochastic modeling of biointeraction. *Math. Biosci.* **177/178**, 25–38.
217. Ndawana, M. L. and Kravtsov, V. E. (2003). Energy level statistics of a critical random matrix ensemble. Random matrix theory. *J. Phys. A* **36**(12), 3639–3645.
218. Neuß, N. (2002). A new sparse-matrix storage method for adaptively solving large systems of reaction-diffusion-transport equations. *Computing* **68**(1), 19–36.
219. von Neumann, J. (1937). Some matrix-inequalities and metrization of matrix-space. *Tomsk. Univ. Rev.* **1**, 286–300.
220. von Neumann, J. and Morgenstern, O. (1953). *Theory of Games and Economic Behavior*, third edition. Princeton University Press, Princeton, NJ.
221. Nica, A. and Speicher, R. (2006) Lectures on the combinatorics of free probability. *London Mathematical Society Lecture Note Series* 335. Cambridge University Press, Cambridge.
222. Onatski, A. (2008). Determining the number of factors from empirical distribution of eigenvalue. Columbia University, Department of Economics, in its series of Discussion Papers with number 0405–19.
223. Ormerod, P. (2004). Extracting information from noisy time series data. (Paper given at the CASS Business School conference December 2004, attended by leading econometricians Robert Engle and David Hendry.) Volterra Consulting Ltd. (Pormerod@volterra.co.uk).
224. Oyman, O., Nabar, R. U., Bolcskei, H., and Paulraj, A. J. (2002). Tight lower bounds on the ergodic capacity of Rayleigh fading MIMO channels. In *Proceedings of IEEE Globecom 2002*, Taipei, 1172–1176.
225. Pafka, Sz., Potters, M., and Kondor, I. (2004). Exponential weighting and random-matrix-theory-based filtering of financial covariance matrices for portfolio optimization. Science & Finance (CFM) working paper archive 500050, Science & Finance, Capital Fund Management.
226. Pan, G. M., Guo, M., and Liang, Y. C. (2007). Asymptotic performance of reduced-rank linear receivers with principal component filter, *IEEE Trans. Inf. Theory,* **53**(3), 1148–1151.
227. Pan, G. M. and Zhou, W. (2007). Circular law, extreme singular values and potential theory. *arXiv:0705.3773v2*.
228. Papp, G., Pafka, Sz., Nowak, M. A., and Kondor, I. (2005). Random matrix filtering in portfolio optimization. *Acta Physica Polonica B* **36**(9), 2757–2765.
229. Pastur, L. A. (1973). Spectra of random self-adjoint operators. *Uspekhi Mat. Nauk,* **28**(1), 4–63, (*Russian Math. Surv.* **28**(1), 1–67).
230. Pastur, L. A. (1972). On the spectrum of random matrices. *Teoret. Mat. Fiz.* **10**, 102–112, (*Teoret. Mat. Phys.* **10**, 67–74).
231. Paul, D. (2007). Asymptotics of sample eigenstruture for a large dimensional spiked covariance model. *Statistica Sinica* **17**(4), 1617–1642.
232. Paulraj, A., Nabar, R., and Gore, D. (2003). *Introduction to Space-Time Wireless Communications.* Cambridge University Press, Combridge.

233. Petz, D. and Rffy, J. (2004). On asymptotics of large Haar distributed unitary matrices. *Period. Math. Hungar.* **49**(1), 103–117.
234. Plerou, V., Gopikrishnan, P., Rosenow, B., Amasel, L. A. N., and Stanley, H. E. (2000). A random matrix theory approach to financial cross-correlations. *Physica A* **287**, 374–382.
235. Plerou, V., Gopikrishnan, P., Rosenow, B., Amasel, L. A. N., and Stanley, H. E. (1999). Universal and non-universal properties of cross-correlations in financial time series. *Phys. Rev. Lett.* **83**, 1471–1474.
236. Potters, M., Bouchaud, J. P., and Laloux, L. (2005). Financial application of random matrix theory: old laces and new pieces. *Acta Physica Polonica B* **36**(9), 2767–2784.
237. Rao, M. B. and Rao, C. R. (1998). *Matrix Algebra and Its Applications to Statistics and Econometrics.* World Scientific, Singapore.
238. Rashid-Farrokhi, F., Liu, K. J. R., and Tassiulas, L. (1998). Transmit beamforming and power control for cellular wireless systems. *IEEE J. Selected Areas Commun.* **16**(8), 1437–1449.
239. Rashid-Farrokhi, F., Tassiulas, L., and Liu, K. J. R. (1998). Joint optimal power control and beamforming in wireless networks using antenna arrays. *IEEE Trans. Commun.* **46**(10), 1313–1324.
240. Ratnarajah, T., Vaillancourt, R., and Alvo, M. (2004/2005). Eigenvalues and condition numbers of complex random matrices. *SIAM J. Matrix Anal. Appl.* **26**(2), 441–456.
241. Robinson, P. M. (1974). Identification, estimation and large sample theory for regressions containing unobservable variables. *Int. Econ. Rev.* **15**(3), 680–692.
242. Rojkova, V. and Kantardzic, M. (2007). Analysis of inter-domain traffic correlations: random matrix theory approach. *arXiv.org:cs/0706.2520.*
243. Rojkova, V. and Kantardzic, M. (2007). Delayed correlation in inter-domain network traffic. *arXiv:0707.10838 v1.* [cs.NI], 7.
244. Ross, S. (1976). The arbitrage theory of capital asset pricing. *J. Econ. Theory* **13**, 341–360.
245. Rudelson, M. (2007). Invertibility of random matrices: norm of the inverse. *Ann. Math.* To appear.
246. Rudelson, M. and Vershynin, R. (2008). The least singular value of a random square matrix is $O(n^{-1/2})$. *C. R. Math. Acad. Sci. Paris* **346**(15–16), 893–896.
247. Rudelson, M. and Vershynin, R. (2008). The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.* **218**(2), 600–633.
248. Sandra, E. and Christina, Z. (2006). How good are dynamic factor models at forecasting output and inflation. Ameta-analytic approach. Discussion Paper Series 1: Economics Studies No. 42, Deutsche Bundesbank.
249. Saranadasa, H. (1993). Asymptotic expansion of the misclassification probabilities of D-and A-criteria for discrimination from two high dimensional populations using the theory of large dimensional random matrices. *J. Multivariate Anal.* **46**, 154–174.
250. Savage, L. J. (1954). *The Foundations of Statistics.* Wiley, New York.
251. Schur, I. (1911). Bemerkungen zur theoríe der beschränkten bilinearformen mit unendlich vielen veränderlichen. *J. Reine Angew. Math.* **140**, 1–28.
252. Semerjian, G. and Cugliandolo, L. F. (2002). Sparse random matrices: the eigenvalue spectrum revisited. *J. Phys. A.* **35**(23), 4837–4851.
253. Sharifi, S., Crane, M., Shamaie, A., and Ruskin, H. (2004). Random matrix theory for portfolio optimization: a stability approach. *Physica A* **335**(3–4), 629–643.
254. Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *J. Finance* **19**, 425–442.
255. Shohat, J. A. and Tamarkin, J. D. (1950). *The Problem of Moments.* Mathematical Surveys No. 1. American Mathematical Society, Providence, RI.
256. Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.* **5**, 331–339.
257. Silverstein, J. W. (1990). Weak convergence of random functions defined by the eigenvectors of sample covariance matrices. *Ann. Probab.* **18**, 1174–1194.

258. Silverstein, J. W. (1989a). On the eigenvectors of large dimensional sample covariance matrices. *J. Multivariate Anal.* **30**, 1–16.
259. Silverstein, J. W. (1989b). On the weak limit of the largest eigenvalue of a large dimensional sample covariance matrix. *J. Multivariate Anal.* **30**, 307–311.
260. Silverstein, J. W. (1985a). The limiting eigenvalue distribution of a multivariate F matrix. *SIAM J. Math. Anal.* **16**(3), 641–646.
261. Silverstein, J. W. (1985b). The smallest eigenvalue of a large dimensional Wishart Matrix. *Ann. Probab.* **13**(4), 1364–1368.
262. Silverstein, J. W. (1984a). Comments on a result of Yin, Bai and Krishnaiah for large dimensional multivariate F matrices. *J. Multivariate Anal.* **15**, 408–409.
263. Silverstein, J. W. (1984b). Some limit theorems on the eigenvectors of large dimensional sample covariance matrices. *J. Multivariate Anal.* **15**, 295–324.
264. Silverstein, J. W. (1981). Describing the behavior of eigenvectors of random matrices using sequences of measures on orthogonal groups. *SIAM J. Math. Anal.* **12**, 174–281.
265. Silverstein, J. W. (1979). On the randomness of eigenvectors generated from networks with random topologies *SIAM J. Appl. Math.* **37**, 235–245.
266. Silverstein, W. J. and Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.* **54**, 175–192.
267. Silverstein, W. J. and Choi, S. I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.* **54**, 295–309.
268. Silverstein, J. W. and Combettes, P. L. (1992). Signal detection via spectral theory of large dimensional random matrices. *IEEE Trans. Signal Process.* **40**, 2100–2104.
269. Sinai, Ya. and Soshnikov, A. (1998). Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bol. Soc. Brasil. Mat. (N. S.)* **29**(1), 1–24.
270. Skorohod, A. V. (1956). Limit theorems for stochastic processes. (Russian). *Teor. Veroyatnost. i Primenen.* **1**, 289–319.
271. Stariolo, D. A., Curado, E. M. F., and Tamarit, F. A. (1996). Distributions of eigenvalues of ensembles of asymmetrically diluted Hopfield matrices. *J. Phys. A* **29**, 4733–4739.
272. Stock, J. H. and Watson, M.W. (2002). Macroeconomic forecasting using diffusion indexes. *J. Business Econ. Statist.* **20**, 147–162.
273. Tao, T. and Vu, V. (2008). Random matrices: the circular law. *Commun. Contemp. Math.* **10**(2), 261–307.
274. Telatar, E. (1999). Capacity of multi-antenna Gaussian channels. *Eur. Trans. Telecommun.* **10**(6), 585–595.
275. Titchmarsh, E. C. (1939). *The Theory of Functions*, second edition. Oxford University Press, London.
276. Tracy, C. A. and Widom, H. (2002). Distribution functions for largest eigenvalues and their applications. *Proceedings of the International Congress of Mathematicians* **1**, 587–596. Higher Education Press, Beijing, 2002.
277. Tracy, C. A. and Widom, H. (1998). Correlation functions, cluster functions, and spacing distributions for random matrices. *J. Statist. Phys.* **92**, 809–835.
278. Tracy, C. A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177**(3), 727–754.
279. Tracy, C. A. and Widom, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159**, 151–174.
280. Trichard, L. G., Evans, J. S., and Collings, I. B. (2002). Large system analysis of linear multistage parallel interference cancellation. *IEEE Trans. Commun.* **50**(11), 1778–1786.
281. Tse, D. and Hanly, S. (1999). Linear multiuser receivers: effective interference, effective bandwidth and user capacity. *IEEE Trans. Inf. Theory* **47**(2), 641–657.
282. Tse, D. and Zeitouni, O. (2000). Linear multiuser receivers in random environments. *IEEE Trans. Inf. Theory* **46**(1), 171–188.
283. Tulino, A. and Verdu, S. (2005). Asymptotic outage capacity of multiantenna channels. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, Vols. 1–5 - Speech Processsing Book Series: International Conference on Acoustics Speech and Signal Processing (ICASSP)*, V825–V828. IEEE, New York.

284. Tulino, A. M. and Verdú, S. (2004). *Random Matrix Theory and Wireless Communications*, now Publishers Inc., Hanover.
285. Vassilevski, P. S. (2002). Sparse matrix element topology with application to AMG(e) and preconditioning. Preconditioned robust iterative solution methods, PRISM '01 (Nijmegen). *Numer. Linear Algebra Appl.* **9**(6–7), 429–444.
286. Verdú, S. and Shamai, S. (1999). Spectral efficiency of CDMA with random spreading. *IEEE Trans. Inf. Theory* **45**(3), 622–640.
287. Voiculescu, D. (1995). Free probability theory: random matrices and von Neumann algebras. In *Proceedings of the ICM 1994*, S. D. Chatterji, editor, 227–241. Birkhäuser, Boston.
288. Voiculescu, D. (1991). Limit laws for random matrices and free products. *Invent. Math.* **104**, 201–220.
289. Voiculescu, D. (1986). Addition of certain non-commuting random variables. *J. Funct. Anal.* **66**, 323–346.
290. Wachter, K. W. (1980). The limiting empirical measure of multiple discriminant ratios. *Ann. Statist.* **8**, 937–957.
291. Wachter, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann Probab.* **6**(1), 1–18.
292. Wang, K. and Zhang, J. (2003). MSP: a class of parallel multistep successive sparse approximate inverse preconditioning strategies. *SIAM J. Sci. Comput.* **24**(4), 1141–1156.
293. Wang, X. and Poor, H. V. (1999). Iterative (turbo) soft interference cancellation and decoding for coded CDMA. *IEEE Trans. Commun.* **47**(7), 1046–1061.
294. Wiegmann, P. and Zabrodin, A. (2003). Large scale correlations in normal non-Hermitian matrix ensembles. Random matrix theory. *J. Phys. A* **36**(12), 3411–3424.
295. Wigner, E. P. (1958). On the distributions of the roots of certain symmetric matrices. *Ann. Math.* **67**, 325–327.
296. Wigner, E. P. (1955). Characteristic vectors bordered matrices with infinite dimensions. *Ann. Math.* **62**, 548–564.
297. Witte, N. S. (2004). Gap probabilities for double intervals in Hermitian random matrix ensembles as τ-functions-spectrum singularity case. *Lett. Math. Phys.* **68**(3), 139–149.
298. Wolfe, P. (1959). The simplex method for quadratic programming. *Econometrica* **27**(3), 382–398.
299. Xia, M. Y., Chan, C. H., Li, S. Q., Zhang, B., and Tsang, L. (2003). An efficient algorithm for electromagnetic scattering from rough surfaces using a single integral equation and multilevel sparse-matrix canonical-grid method. *IEEE Trans. Antennas Propagation* **51**(6), 1142–1149.
300. Yin, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20**, 50–68.
301. 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. *Probab. Theory Relat. Fields* **78**, 509–521.
302. Yin, Y. Q., Bai, Z. D., and Krishnaiah, P. R. (1983). Limiting behavior of the eigenvalues of a multivariate F matrix. *J. Multivariate Anal.* **13**, 508–516.
303. Yin, Y. Q. and Krishnaiah, P. R. (1985). Limit theorem for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic. *Theory Probab. Appl.* **30**, 861–867.
304. Yin, Y. Q. and Krishnaiah, P. R. (1983). A limit theorem for the eigenvalues of product of two random matrices. *J. Multivariate Anal.* **13**, 489–507.
305. Zabrodin, A. (2003). New applications of non-Hermitian random matrices. *Ann. Henri Poincare* **4**(2), S851–S861.
306. Zellner, A. (1970). Estimation of regression relationships containing unobservable independent variables. *Int. Econ. Rev.* **11**(3), 441–454.

307. Zeng, Y. H. and Liang, Y. C. (2007). Maximum-minimum eigenvalue detection for cognitive radio. In *Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communication* IEEE, New York. 1–5.
308. Zeng, Y. H. and Liang, Y. C. (2007). Eigenvalue based spectrum sensing algorithms for cognitive radio, submitted to *IEEE Trans. Commun.* arXiv:0804.2960v1.
309. Zhang, J., Chong, E., and Tse, D. (2001). Output MAI distributions of linear MMSE multiuser receivers in DS-CDMA systems. *IEEE Trans. Inf. Theory* **47**(3), 1028–1144.
310. Zheng, Shurong (2009). Central Limit Theorem for Linear Spectral Statistics of Large Dimensional F-Matrix. To appear in *Ann. Inst. Henri Poincaré Probab. Statist.*

Index

The aim of the book is to introduce basic concepts, main results, and widely applied mathematical tools in the spectral analysis of large dimensional random matrices. The core of the book focuses on results established under moment conditions on random variables using probabilistic methods, and is thus easily applicable to statistics and other areas of science. The book introduces fundamental results, most of them investigated by the authors, such as the semicircular law of Wigner matrices, the Marčenko-Pastur law, the limiting spectral distribution of the multivariate F-matrix, limits of extreme eigenvalues, spectrum separation theorems, convergence rates of empirical distributions, central limit theorems of linear spectral statistics, and the partial solution of the famous circular law. While deriving the main results, the book simultaneously emphasizes the ideas and methodologies of the fundamental mathematical tools, among them being: truncation techniques, matrix identities, moment convergence theorems, and the Stieltjes transform. Its treatment is especially fitting to the needs of mathematics and statistics graduate students and beginning researchers, having a basic knowledge of matrix theory and an understanding of probability theory at the graduate level, who desire to learn the concepts and tools in solving problems in this area. It can also serve as a detailed handbook on results of large dimensional random matrices for practical users. This second edition includes two additional chapters, one on the authors' results on the limiting behavior of eigenvectors of sample covariance matrices, another on applications to wireless communications and finance. While attempting to bring this edition up-to-date on recent work, it also provides summaries of other areas which are typically considered part of the general field of random matrix theory.

Zhidong Bai is a professor of the School of Mathematics and Statistics at Northeast Normal University and Department of Statistics and Applied Probability at National University of Singapore. He is a Fellow of the Third World Academy of Sciences and a Fellow of the Institute of Mathematical Statistics.

Jack W. Silverstein is a professor in the Department of Mathematics at North Carolina State University. He is a Fellow of the Institute of Mathematical Statistics.