Spectral analysis of normalized sample covariance matrices with large dimension and small sample size

Chen, Binbin

2013

Chen, B. (2013). Spectral analysis of normalized sample covariance matrices with large dimension and small sample size. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/54960>

<https://doi.org/10.32657/10356/54960>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**Spectral Analysis of Normalized
Sample Covariance Matrices with
Large Dimension and Small Sample
Size**

Binbin Chen

**Division of Mathematical Sciences
School of Physical and Mathematical Sciences**

2013

Spectral Analysis of Normalized Sample Covariance Matrices with Large Dimension and Small Sample Size

Binbin Chen

Division of Mathematical Sciences
School of Physical and Mathematical Sciences

A thesis submitted to Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy in Mathematical Sciences

October 2013

Acknowledgements

My most sincere gratitude goes to my advisor and mentor, Guangming Pan, professor in School of Physical and Mathematical Sciences, Nanyang Technological University. I thank him for introducing me to the wonders of scientific research and guiding me through my whole academic progress. I thank him for his invaluable guidance, unfailing encouragement and patience over the last four years without any reservation. I thank him for his wide knowledge and innovative ideas which have provided a good basis for the present thesis.

I owe deep gratitude to Prof. Wang Zhou, Department of Statistics and Applied Probability, National University of Singapore, for his insight and precious suggestions of improving my another research project, *large dimensional empirical likelihood*. I am gratitude to Prof. Shih-Feng Huang, National University of Kaohsiung, for his effort and suggestions of introducing me the literatures of mean-variance optimization problem. I am also gratitude to Prof. Jiti Gao, Department of Econometrics and Business Statistics, Monash University, for introducing me the field of econometrics. Their brilliant wisdoms have constantly been sparkling and inspiring me in my graduate studies.

I wish to express my warm thanks to the Department of Physical and Mathematics Sciences of Nanyang Technological University for providing me the scholarships to pursue research studies and the opportunities to learn with outstanding academicians. I am thankful to many teachers who have given me dedicated instructions during my long lasting student life. I am especially thank Prof. Heng Lian, Prof. Zhen Pang and Prof. Liming Xiang for their tremendous help and guidance.

I wish to thank many amazing people as friends who make my four year

study at NTU a wonderful experience. I am especially thank Zhigang Bao, Jiang Hu, Xiao Han, Peng Lai, Zhouping Li, Qing Yang, Yanrong Yang and Bo Zhang. I enjoy every moment talking with them of research problems and discussing the literatures during the weekly seminars.

Most importantly, I wish to express my forever love, respect and gratitude to my parents, Changzhen Dong and Shugan Zhou, for breeding me and cultivating me with their significant support, care and love. Finally, I owe my loving thanks to my girlfriend Yuling Liu, the most important person in my life, for her consistent encouragement and wholehearted support. To them, I dedicate this thesis.

List of Works

Below is the list of work done during my PhD studies in NTU.

1. Chen, B. B. and Pan, G.M. (2012) Convergence of the largest eigenvalue of normalized sample covariance matrices when p and n both tend to infinity with their ratio converging to zero. *Bernoulli*, **18**, 1405-1420.
2. Chen, B. B. and Pan, G.M. (2013) CLT for linear spectral statistics of normalized sample covariance matrices with larger dimension and small sample size. Under revision.
3. Chen, B. B., Pan, G. M. and Zhou, W. (2013) Large Dimensional Empirical Likelihood. Preprint.
4. Chen, B. B., Huang, S. F. and Pan, G. M. (2013) High Dimensional Mean-Variance Optimization Through Factor Analysis. Preprint.

Contents

Acknowledgements	iii
List of Works	v
Abstract	1
1 Introduction	3
1.1 Large dimensional sample covariance matrix	5
1.2 Wigner matrix	10
1.3 Normalized sample covariance matrix	12
1.4 Thesis outline	13
1.5 Notations	13
2 Convergence of the maximum eigenvalue	15
2.1 Preliminary Results	16
2.2 Proof of Theorem 1	17
2.3 Proof of Theorem 2	21
3 Central limit theorem for the linear spectral statistics	33
3.1 Preliminary Results	35

3.2	Truncation and strategy for the proofs of Theorem 3 and Theorem 4	40
3.2.1	Truncation	41
3.2.2	Strategy of the proof	43
3.3	Convergence of $M_n^{(1)}(z)$	46
3.3.1	Simplification of $M_n^{(1)}(z)$	47
3.3.2	Application of central limit theorem of martingales .	50
3.3.3	The asymptotic expression of $\mathbb{Z}_k$	52
3.3.3.1	Decomposition of $\mathbb{Z}_k$	53
3.3.3.2	Conclusion of the asymptotic expansion of $\mathbb{Z}_k$	56
3.3.4	Verify Condition 2	60
3.3.5	Tightness of $M_n^{(1)}(z)$	60
3.4	Uniform convergence of $EM_n(z)$	62
3.4.1	Proof of Lemma 1	66
3.5	Proof of Proposition 2	69
3.6	Calculation of the mean and covariance	70
3.6.1	The Mean	70
3.6.2	The covariance	72
3.7	Calibration	75
3.7.1	Calibration of the mean correction term in (3.7) . . .	75
3.7.2	Simulations for checking the accuracy of CLT	77
3.8	Some Lemmas	78
4	Application of CLTs to Hypothesis test	91
4.1	Simulation Study	91
5	Conclusion and future research	99
5.1	Conclusion	99

5.2	Future research	100
5.2.1	With sample mean	101
5.2.2	With covariance	101
Bibliography		103

List of Figures

2.1	A example of W -graph	24
3.1	Normal Q-Q plot with the sample generating from $N(0, 1)$ (red) and normalized $exp(1)$ (green). The sample size are set to $n = 50, 100, 150, 200$ corresponds to the figures from left to right while the dimension $p = n^2$	88
3.2	Normal Q-Q plot with the sample generating from $N(0, 1)$ (red) and normalized $exp(1)$ (green). The sample size are set to $n = 50, 100, 150, 200$ corresponds to the figures from left to right while the dimension $p = n^{2.5}$	89

List of Tables

3.1	Empirical mean and standard deviation	87
4.1	Empirical sizes of CZZ test and the proposed identity test at 5% significance level for normal, gamma, bernoulli random vectors. .	95
4.2	Empirical power of CZZ test and the proposed identity test at 5% significance level for normal random vectors. Two types of population covariance matrices are considered. In the first case $\Sigma_1 = \text{diag}(2 \times \mathbf{1}_{[\nu p]}, \mathbf{1}_{p-[\nu p]})$ for $\nu = 0.08$ and $\nu = 0.25$ respec- tively. In the second case $\Sigma_2 = \text{diag}(A_1, \text{diag}(\mathbf{1}_{p-[v_2 p]}))$, where A_1 is a $[v_2 p] \times [v_2 p]$ tridiagonal symmetric matrix with diagonal elements equal to 1 and elements beside diagonal all equal to v_1 for $v_1 = 0.5, v_2 = 0.8$ and $v_1 = 0.5, v_2 = 0.4$ respectively.	96

- 4.3 Empirical power of CZZ test and the proposed identity test at 5% significance level for standard gamma random vectors. Two types of population covariance matrices are considered. In the first case $\Sigma_1 = \text{diag}(2 \times \mathbf{1}_{[\nu p]}, \mathbf{1}_{p-[\nu p]})$ for $\nu = 0.08$ and $\nu = 0.25$ respectively. In the second case $\Sigma_2 = \text{diag}(A_1, \text{diag}(\mathbf{1}_{p-[v_2 p]}))$, where A_1 is a $[v_2 p] \times [v_2 p]$ tridiagonal symmetric matrix with diagonal elements equal to 1 and elements beside diagonal all equal to v_1 for $v_1 = 0.5, v_2 = 0.8$ and $v_1 = 0.5, v_2 = 0.4$ respectively. . 97
- 4.4 Empirical power of CZZ test and the proposed identity test at 5% significance level for standard bernoulli random vectors. Two types of population covariance matrices are considered. In the first case $\Sigma_1 = \text{diag}(2 \times \mathbf{1}_{[\nu p]}, \mathbf{1}_{p-[\nu p]})$ for $\nu = 0.08$ and $\nu = 0.25$ respectively. In the second case $\Sigma_2 = \text{diag}(A_1, \text{diag}(\mathbf{1}_{p-[v_2 p]}))$, where A_1 is a $[v_2 p] \times [v_2 p]$ tridiagonal symmetric matrix with diagonal elements equal to 1 and elements beside diagonal all equal to v_1 for $v_1 = 0.5, v_2 = 0.8$ and $v_1 = 0.5, v_2 = 0.4$ respectively. . 98

Abstract

Sample covariance matrix, which is to give an idea about the statistical interdependence structure of the data, is a fundamental tool in multivariate statistical analysis. Due to rapid development and wide applications in statistics, wireless communication and econometric theory, significant effort has been made to understand the asymptotic behaviour of the eigenvalues of large dimensional sample covariance matrices where the sample size n and the number of variables p are both very large but their ratio roughly tends to a constant.

In contrast, this thesis studies the spectral properties of large dimensional sample covariance matrices where the dimension is much larger than the sample size. The pioneer work was done by Bai and Yin (1988) in this direction. Under the assumption $p/n \rightarrow \infty$, they showed that the empirical spectral distribution of the large normalized sample covariance matrix $\mathbf{B} := \frac{1}{\sqrt{np}}(\mathbf{X}^T \mathbf{X} - p\mathbf{I}_n)$ converges to the semicircle law almost surely, where $\mathbf{X}$ is a $p \times n$ random matrix with independent, identically distributed entries. This thesis extends such result in two aspects:

In the first part of this work, we prove that the largest eigenvalue of $\mathbf{B}$ almost surely tends to 2, which is the right end point of the support of the semicircle law. Indeed, after truncation and normalization of the entries of the matrix $\mathbf{B}$, we show that the convergence rate is $o(n^\ell)$ for any $\ell > 0$.

In the second part of this work, we establish the central limit theorem for the linear spectral statistics of the eigenvalues of $\mathbf{B}$ under the existence of the fourth moment of underlying variables. Statistical applications covers the so-called “very large (or ultra) p and small n ” situations. We also explore the application of this result in testing whether a population covariance matrix is an identity matrix or not.

Chapter 1

Introduction

This thesis comes with new results that deal with a not yet touched but important class of random matrices, namely large dimensional sample covariance matrices where both dimensions go to infinity but the dimension is much larger than the sample size. Sample covariance matrix, which is to give an idea about the statistical interdependence structure of the data, is a fundamental tool in multivariate statistical analysis. In the classical settings, the sample covariance matrix is an unbiased and efficient estimator of the population covariance matrix when the dimension is fixed and sample size goes to infinity. Studying the eigenvalues of the sample covariance matrix has been one of the most important topic in the statistical applications as lots of statistics can be expressed by the functionals of the eigenvalues, e.g. *PCA*, *trace*, *determinant*, *etc.* Most of the relevant questions are concerned with the eigenvalues of the sample covariance matrix has been investigated in the classical books, e.g., Anderson (2003), Muirhead (1982).

The last few decades have seen explosive growth in data analysis, due to the rapid development of modern information technology. We are now in a setting where many very important data analysis problems are large

dimensional, with number of variables p and the sample size n being both very large. For example, in micro-array experiments, the number of genes can be tens of thousands or hundreds of thousands while there are only hundreds of samples. Such kind of data also arises in genetic, proteomic, functional magnetic resonance imaging studies and so on (see Chen, Zhang and Zhong 2010, Donoho 2010 and Fan and Fan 2008). In these applications, classical statistical methods and results based on fixed dimension and large sample size are often doubtful.

Let p denote the dimension and n denote the sample size. Instead of considering the classical framework “fixed p , large n ”, statisticians prefer a better theoretical framework for modern datasets, namely “large p , large n ”. In other words, one should consider that both p and n go to infinity, perhaps with certain restrictions, e.g. p is proportional to n . Sample covariance matrix with the dimension p going to infinity is usually called *large dimensional sample covariance matrix*, which has been well known as one of the classical models of the so-called *large dimensional random matrix*. Due to rapid development and wide applications in statistics, wireless communication and econometric theory, significant effort has been made to understand the asymptotic behaviours of the spectral of the large sample covariance matrices.

Spectral analysis of large dimensional random matrix was originated from the Wigner matrix model in the area of nuclear physics during the 1950’s. In the following, we start with large dimensional sample covariance matrix to introduce the important concepts that are investigated in the large dimensional random matrices.

1.1 Large dimensional sample covariance matrix

After the pioneering work of Wigner's semicircle law for a Gaussian (or Wigner) matrix by Wigner (1955, 1958), the study of the spectral analysis of large dimensional sample covariance matrix has attracted considerable interest in the literatures of random matrix theory. Let $\mathbf{s}_1, \dots, \mathbf{s}_n$ be independent, identically distributed (i.i.d.) p -dimensional random vectors. The elements of each $\mathbf{s}_k$ are i.i.d. with mean zero and variance 1. In the literature of random matrix theory, the sample covariance matrix is defined as

$$\mathbf{S} := \frac{1}{n} \sum_{k=1}^n \mathbf{s}_k \mathbf{s}_k^T = \frac{1}{n} \mathbf{X} \mathbf{X}^T,$$

where $\mathbf{X} = (\mathbf{s}_1, \dots, \mathbf{s}_n)$ is a $p \times n$ data matrix.

Large amount of the literature has been devoted to the theoretical framework “large p , large n ”, with the restriction that their ratio goes to a finite constant c , e.g. $p/n \rightarrow c$. The pioneering work in finding the limit of the empirical spectral distribution (ESD), which is named the limiting spectral distribution (LSD), of $\mathbf{S}$ is due to Marcenko and Pastur (1967).

Definition 1 (Empirical spectral distribution). *Let $\mathbf{M}_n$ be a $n \times n$ Hermitian matrix with real eigenvalues $\lambda_1(\mathbf{M}_n), \dots, \lambda_n(\mathbf{M}_n)$. Then the empirical spectral distribution of $\mathbf{M}_n$ is defined by*

$$F^{\mathbf{M}_n} = \frac{1}{n} \sum_{j=1}^n I(\lambda_j(\mathbf{M}_n) \leq x),$$

where $I(\cdot)$ is the indicator function.

Definition 2 (Limiting spectral distribution). *If for a class of $n \times n$ random matrices $\mathbf{M}_n$, it is proven with probability one (or in probability) that the*

empirical spectral distribution $F^{\mathbf{M}_n}$ converges weakly to some distribution function F as n tends to infinity, then F is said to be the limiting spectral distribution of the matrices $\mathbf{M}_n$ in the strong sense (or in the weak sense).

Let $F^{\mathbf{S}}$ be the ESD of $\mathbf{S}$. Assume $p/n \rightarrow c \in (0, \infty)$. For the matrix $\mathbf{S}$, Marcenko and Pastur (1967) proved that $F^{\mathbf{S}}$ converges weakly to the famous MP law with density

$$F'_c(x) = \begin{cases} \frac{1}{2\pi cx} \sqrt{(b-x)(x-a)}, & \text{if } a \leq x \leq b, \\ 0, & \text{otherwise,} \end{cases} \quad (1.1)$$

and a point mass $1 - 1/c$ at the origin if $c > 1$, where $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$.

Most of the known results are on the general sample covariance matrix taking the form $\mathbf{Q} = \frac{1}{n} \mathbf{T}^{1/2} \mathbf{X} \mathbf{X}^T \mathbf{T}^{1/2}$, where $\mathbf{X}$ is a $p \times n$ matrix consisting i.i.d. components with mean zero and variance 1, $\mathbf{T}$ is $p \times p$ nonnegative definite matrix and $\mathbf{X}, \mathbf{T}$ are independent. This type of random matrix represents a large class of matrices which are important in multivariate analysis. When $\mathbf{T} = \mathbf{I}_p$, $\mathbf{Q}$ becomes the sample covariance matrix $\mathbf{S}$. When $\mathbf{T}$ is non-random, $\mathbf{Q}$ is the sample covariance matrix of the data $\mathbf{T}^{1/2} \mathbf{s}_k, k = 1, \dots, n$ where $\mathbf{s}_k$ is the k -th column of $\mathbf{X}$.

The convergence of ESD of the general sample covariance matrix $\mathbf{Q}$ was established by Yin and Krishnaiah (1983) and Yin (1986) by using moment method. A better understanding of the spectral properties of the matrix $\mathbf{Q}$ is the proof of convergence of its ESD by using Stieltjes transform method. This work was done by Silverstein (1995), to which Silverstein and Bai (1995) is an important related work. Assuming that:

- i) The data matrix $\mathbf{X}$ is a $p \times n$ matrix consisting i.i.d. components with mean zero and variance 1 and bounded fourth moments,

- ii) With probability one, the ESD of $\mathbf{T}$, $F^{\mathbf{T}}$, converges weakly to a non-random distribution H ,
- iii) $p/n \rightarrow c > 0$ as $n \rightarrow \infty$.

It is proved by Silverstein (1995) that with probability one, $F^{\mathbf{Q}}$ converges in distribution to $F^{c,H}$ whose stieltjes transform, denoted by $m_{F^{c,H}}(z)$, is the unique solution to

$$m = \int \frac{dH(\lambda)}{\lambda(1 - c - czm) - z}. \quad (1.2)$$

Definition 3 (Stieltjes transform). *Let $G(x)$ be any distribution function. Then the Stieltjes transform of $G(x)$ is defined as*

$$m_G(z) = \int \frac{1}{x - z} dG(x), \quad z \in \mathbb{C}^+ = \{z \in \mathbb{C}, \text{Im} z > 0\},$$

where $\text{Im}(\cdot)$ denotes the imaginary part of a complex number.

The importance of ESD is due to the fact that many important statistics in multivariate analysis can be written as functionals of the ESD. For example the *log determinant statistic*: $\frac{1}{p} \log \det(\mathbf{S}) = \frac{1}{p} \sum_{k=1}^p \log \lambda_k(\mathbf{S}) = \int_0^\infty \log x dF^{\mathbf{S}}(x)$. Since $F^{\mathbf{S}}$ tends to MP law with probability one, it is reasonable to use $\int_0^\infty \log x dF_c(x)$ to approximate $\frac{1}{p} \log \det(\mathbf{S})$ by Helly-Bray theorem. To apply the Helly-Bray theorem, one has to make sure that the extreme eigenvalues of $\mathbf{S}$ remain in certain bounded intervals. This leads to another important topic after the LSD was found: the asymptotic behavior of the extreme eigenvalues.

There are extensive literatures for this problem. The first work is due to Geman (1980) in this direction. He proved that the largest eigenvalue of a sample covariance matrix tends to $(1 + \sqrt{c})^2$ when $p/n \rightarrow c \in (0, \infty)$ under a restriction on the growth rate of the moments of the underlying variables.

After that, the study of the extreme eigenvalues has been developed in various aspects. In Yin, Bai and Krishnaiah (1988), the assumption of Geman's work was reduced to the existence of the fourth moment of the underlying distribution. It was proved that the existence of the fourth moment of the underlying distribution is also the necessary condition to ensure the almost surely convergence of the maximum eigenvalue of $\mathbf{S}$ by Bai, Silverstein and Yin (1988).

A more difficult problem concerns the smallest eigenvalue of large dimensional sample covariance matrix. Results about the smallest eigenvalue of Wishart matrix were obtained by Yin, Bai and Krishnaiah (1983), Silverstein (1984, 1985). In the most recent work by Bai and Yin (1993), it is proved that $\lambda_{\min}(\mathbf{S}) \xrightarrow{a.s.} (1 - \sqrt{c})^2$ under the existence of the fourth moment of the underlying distribution.

Another important topic concerns the central limit theorem for the linear spectral statistics of the random matrices. The linear spectral statistic (LSS) is the linear functional of the eigenvalues of the random matrix, which can represent many interesting statistics in multivariate analysis.

Definition 4 (Linear spectral statistic). *Let $\mathbf{M}_n$ be a $n \times n$ Hermitian matrix with real eigenvalues $\lambda_1(\mathbf{M}_n), \dots, \lambda_n(\mathbf{M}_n)$. For any function f , the linear spectral statistic of $\mathbf{M}_n$ with respect to f is defined as*

$$\hat{\theta}_n = \frac{1}{n} \sum_{k=1}^n f(\lambda_k(\mathbf{M}_n)).$$

The linear spectral statistic $\hat{\theta}_n$ can also be written in the form of empirical spectral distribution of $\mathbf{M}_n$ that

$$\hat{\theta}_n = \int_{-\infty}^{\infty} f(x) dF^{\mathbf{M}_n}(x).$$

Central limit theorem for linear spectral statistics of the general sample covariance matrices $\mathbf{Q}$ was first studied by Jonsson (1982) when the underlying variables are Gaussian. Such results were developed for a set of analytic functions in Bai and Silverstein (2004) by using a new method, namely central limit theorem for martingales. In Bai and Silverstein (2004), the underlying variables x_{ij} 's could be either real or complex with i) $Ex_{ij}^4 = 3$ if x_{ij} 's are real; ii) $Ex_{ij}^2 = 0$ and $E|x_{ij}|^4 = 2$ if x_{ij} 's are complex.

Let F^{c_n, H_n} be the distribution function whose stieljes transform is the unique solution to (1.2) with c, H being replaced by c_n, H_n . Define $X_n(f) := n \int f(x) d[F^{\mathbf{Q}}(x) - F^{c_n, H_n}(x)]$, $f \in \mathcal{A}$ where $\mathcal{A}$ is a contour of the complex plane enclosing the interval

$$\left[\liminf_{n \rightarrow \infty} \lambda_{\min}(\mathbf{T}) I(c \in (0, 1))(1 - \sqrt{c})^2, \liminf_{n \rightarrow \infty} \lambda_{\max}(\mathbf{T})(1 + \sqrt{c})^2 \right].$$

For any fixed constant k and $f_j \in \mathcal{A}, j = 1, \dots, k$, the central limit theorem for the finite random vector

$$(X_n(f_1), X_n(f_2), \dots, X_n(f_k))$$

is established.

Statistical applications have been explored by using the CLT for LSS of the general sample covariance matrix $\mathbf{Q}$. In Bai et al. (2009), the authors observed that the likelihood ratio procedures for testing one the covariance matrices from Gaussian populations fail when the dimension is large compared to the sample size. To cope with high dimensional effects, they propose necessary corrections for these likelihood ratio tests by using the CLT of the LSS of the sample covariance matrices.

Recent development of CLT for LSS of $\mathbf{Q}$ is due to Pan and Zhou (2008). They generalized the results in Bai and Silverstein (2004) by releasing the

conditions that $Ex_{ij}^4 = 3$ for real and $E|x_{ij}|^4 = 2$ for complex. Applications in the study of testing the independence of the high dimensional data were studied by Pan, Gao and Yang (2013). They incorporated the central limit theorem for LSS of $\mathbf{Q}$ in Pan and Zhou (2008) to establish the asymptotic normality of the proposed test statistics.

1.2 Wigner matrix

Spectral analysis of large dimensional random matrices was first studied by Wigner in the area of nuclear physics during the 1950's. At that time, theoretical analysis of low-lying excited states of complex nuclear achieved great success, but analyzing the highly excited states is marked by great challenges. Usually, the energy levels of nuclear can not be directly observable, but can be characterized by their statistical distributions. Wigner (1955) initially modeled the energy levels of nuclear by random matrices and produced the idea that the energy levels could be characterized by the eigenvalues of a matrix of observation.

The random matrix Wigner investigated is a $n \times n$ real symmetric matrix $\mathbf{W}_n = (w_{ij})$ whose entries on and above the diagonal are independent Gaussian random variables with mean 0 and variance 1 for off-diagonal entries and 2 for diagonal entries. The order n is very large due to the complication of the complex nuclear. To describe the energy level, the eigenvalue statistics, the so-called “empirical spectral distribution”, is considered. For matrix $\mathbf{W}_n$, the pioneering work was done by Wigner (1958). He proved that the expected empirical spectral distribution of $\mathbf{W}_n/\sqrt{n}$ converges weakly

to the famous semicircle law whose density is given by

$$F'_{sc}(x) = \begin{cases} \frac{1}{2\pi} \sqrt{4 - x^2}, & \text{if } |x| \leq 2, \\ 0, & \text{if } |x| > 2. \end{cases} \quad (1.3)$$

Since then, research on the limit spectral distribution of large dimensional random matrix has attracted considerable interest. In Grenander (1963) and Arnold (1967, 1971), Wigner's work was generalized in the sense of "almost surely". In the most recent review work, it was shown that two general assumptions can be used to define the Wigner matrix *i*) the entries above or on the diagonal of $\mathbf{W}_n$ are independent but may not be necessarily identically distributed; *ii*) the entries w_{ij} satisfy the Lindeberg type condition that for any $\eta > 0$, as $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i,j} E|w_{ij}|^2 I(|w_{ij}| \geq \eta\sqrt{n}) = 0.$$

It was shown in Bai (1999) that the ESD of this generalized Wigner matrix $\frac{1}{\sqrt{n}}\mathbf{W}_n$ also tends to semicircle law almost surely.

The asymptotic behaviors of the extreme eigenvalues of $\frac{1}{\sqrt{n}}\mathbf{W}_n$ were solved by Bai and Yin (1988). In this work, the sufficient and necessary condition for the largest eigenvalue to converge almost surely to a finite constant was given. Central limit theorems for the linear spectral statistics of the Wigner matrix were considered in Bai and Yao (2005). The work of Bai and Yao (2005) used similar arguments developed in the significant work Bai and Silverstein (2004) on the sample covariance matrices. Mainly, let F_n^w be the empirical spectral distribution of $\mathbf{W}_n$, F_{sc} be the semicircle law and $\mathcal{B}$ be the set of analytic functions on a open set enclosing the support of the semicircle law. Define the so-called empirical spectral process $Y_n(f)$ as

$$Y_n(f) = n \int f(x)[F_n^w(x) - F_{sc}(x)], f \in \mathcal{B}.$$

The asymptotic normality of the finite random vector $(Y_n(f_1), \dots, Y_n(f_k))$ for $f_l \in \mathcal{B}, l = 1, \dots, k$ and a fixed k is established.

1.3 Normalized sample covariance matrix

In contrast with asymptotic behaviours of $\mathbf{S}$ in the case of $p/n \rightarrow c$, we turn our attention to the setting $p/n \rightarrow \infty$, that is the dimension is much larger than the sample size. It is conceivable that the first largest n eigenvalues of $\mathbf{S}$ will tend to 1 and the rest $p - n$ eigenvalues are zeros. Hence the ESD of $\mathbf{S}$ will tend to a degenerate distribution with total mass concentrating at 0. It seems that studying the original sample covariance matrix will not be proper.

As zero eigenvalues of $\mathbf{S}$ are of no interest, we turn to investigate the spectral of $\frac{1}{n}\mathbf{X}^T\mathbf{X}$ whose eigenvalues are the same as the nonzero eigenvalues of $\mathbf{S}$. When n is fixed, as $p \rightarrow \infty$, it is easily to show that $\frac{1}{p}\mathbf{X}^T\mathbf{X} \rightarrow \mathbf{I}_n$, where $\mathbf{I}_n$ is the identity matrix with order n . Central limit theorem implies that $\sqrt{p}(\frac{1}{p}\mathbf{X}^T\mathbf{X} - \mathbf{I}_n)$ will tend to a Wigner matrix. Then if we further normalize the matrix by $\sqrt{n}$ and let $n \rightarrow \infty$, we can imagine that the ESD of the normalized matrix $\sqrt{\frac{p}{n}}(\frac{1}{p}\mathbf{X}^T\mathbf{X} - \mathbf{I}_n)$ will tend to semicircle law, which is the limit of ESD of Wigner matrix. Indeed, Bai and Yin (1988) proved that the ESD of this normalized sample covariance matrix converges to the semicircle law with probability one by using moment method. From this point of view, the matrix $\sqrt{\frac{p}{n}}(\frac{1}{p}\mathbf{X}^T\mathbf{X} - \mathbf{I}_n)$ also establishes the connection between Wigner matrix and sample covariance matrix.

In the following, we shall name the matrix $\mathbf{B} := \sqrt{\frac{p}{n}}(\frac{1}{p}\mathbf{X}^T\mathbf{X} - \mathbf{I}_p) = \frac{1}{\sqrt{np}}(\mathbf{X}^T\mathbf{X} - p\mathbf{I}_n)$ the *normalized sample covariance matrix*. This thesis aims at two targets: 1) Prove that the largest eigenvalue of $\mathbf{B}$ almost surely

converges to 2, which is the right end point of support of the semicircle law. Regarding the smallest eigenvalue of $\mathbf{B}$, the result can be obtained similarly by investigating the largest eigenvalue of $-\mathbf{B}$. 2) Study the central limit theorem for the linear spectral statistics of $\mathbf{B}$ as well as its application in hypothesis testing.

1.4 Thesis outline

The main content of the thesis is organized as follows.

- In Chapter 2, we proved that the largest eigenvalue of $\mathbf{B}$ converges to 2 almost surely when $p \rightarrow \infty, n \rightarrow \infty, p/n \rightarrow \infty$ by using moment method.
- In Chapter 3, we prove the CLT for the LSS of $\mathbf{B}$ defined by the eigenvalues under the assumption that the fourth moment exists. The explicit mean and covariance functions are obtained.
- In Chapter 4, we provide a statistical application of the CLT for the LSS of $\mathbf{B}$ which is to test $H_0 : \Sigma = \mathbf{I}$ vs $H_1 : \Sigma \neq \mathbf{I}$. The proposed test statistic is compared with Chen, Zhang and Zhong (2010) and shows its advantage of power when n is extremely small and p is large.

1.5 Notations

We summarize the notations used in this thesis. The Stieltjes transform $m(z)$ of the semicircle law F_{sc} is given by $m(z) = -\frac{1}{2}(z - \sqrt{z})$ which satisfies the equation $m^2 + zm + 1 = 0$. Here $\sqrt{z}$ (and in what follows) is the square root of z with positive imaginary part and $\text{Im}(z)$ denotes the imaginary part

of a complex number z . For any square matrix, $\lambda_{\max}(\mathbf{M})$, $\lambda_{\min}(\mathbf{M})$, $\lambda_k(\mathbf{M})$ denote the minimum eigenvalue, maximum eigenvalue and k -th largest eigenvalue of $\mathbf{M}$, respectively; $\text{tr}\mathbf{M}$ denotes the trace; $\overline{\mathbf{M}}$ represents the complex conjugate matrix of $\mathbf{M}$; The norm $\|\mathbf{M}\|$ represents the spectral norm of $\mathbf{M}$, e.g. $\|\mathbf{M}\| = \sqrt{\lambda_{\max}(\mathbf{M}\overline{\mathbf{M}})}$. The notation $o(1)$ stands for a term converging to zero; $o_{L_p}(1)$ is a term that converges to zero in L_p norm. $\xrightarrow{a.s.}$ means “converge almost surely to”; $\xrightarrow{i.p.}$ means “converge in probability to”; $\xrightarrow{d}$ means “converge in distribution to”. Throughout this paper, K denotes a constant which may take different values at different places. The identity matrix with order n is denoted by $\mathbf{I}_n$, but we may use just $\mathbf{I}$ for simplification.

Chapter 2

Convergence of the maximum eigenvalue

Empirical spectral distribution studies the bulk property of a random matrix, that is, the property of the full set eigenvalues. In this chapter, we focus on the extreme eigenvalues, which address the largest and the smallest eigenvalue of a random matrix.

In PCA, eigenvalue decomposition of a data covariance matrix is usually used to construct the principal components whose variances are the largest few eigenvalues. In the estimation of a sparse mean vector, the maximum eigenvalue of n i.i.d. Gaussian noise variables plays a key role. Similarly, in distinguishing a “signal subspace” of high variance from many noise variables, one expects the largest eigenvalue of sample covariance matrix to play a basic role.

Now we turn our attention to the maximum eigenvalue of normalized sample covariance matrix $\mathbf{B}$.

2.1 Preliminary Results

The main results are presented in the following theorems.

Theorem 1. *Let $\mathbf{X} = (X_{ij})_{p \times n}$ where $\{X_{ij} : i = 1, 2, \dots, p; j = 1, 2, \dots, n\}$ are i.i.d. real random variables with $EX_{11} = 0, EX_{11}^2 = 1$ and $EX_{11}^4 < \infty$.*

Suppose that $p = p(n) \rightarrow \infty$ and $p/n \rightarrow \infty$ as $n \rightarrow \infty$. Define

$$\mathbf{B} = (B_{ij})_{n \times n} = \frac{1}{\sqrt{np}}(\mathbf{X}^T \mathbf{X} - p\mathbf{I}_n).$$

Then as $n \rightarrow \infty$

$$\lambda_{\max}(\mathbf{B}) \rightarrow 2 \quad a.s..$$

Indeed, after truncation and normalization of the entries of the matrix $\mathbf{B}$, we may obtain a better result:

Theorem 2. *Let $p = p(n) \rightarrow \infty$ and $p/n \rightarrow \infty$ as $n \rightarrow \infty$. Define a $n \times n$ random matrix $\mathbf{B}$:*

$$\mathbf{B} = (B_{ij})_{n \times n} = \frac{1}{\sqrt{np}}(\mathbf{X}^T \mathbf{X} - p\mathbf{I}_n),$$

where $\mathbf{X} = (X_{ij})_{p \times n}$. Suppose that X_{ij} 's are i.i.d. real random variables and satisfy the following conditions

- 1) $EX_{11} = 0, EX_{11}^2 = 1, EX_{11}^4 < \infty$ and
- 2) $|X_{ij}| \leq \delta_n \sqrt[4]{np}$, where $\delta_n \downarrow 0$, but $\delta_n \sqrt[4]{np} \uparrow +\infty$, as $n \rightarrow \infty$.

Then, for any $\epsilon > 0, \ell > 0$

$$P(\lambda_{\max}(\mathbf{B}) \geq 2(1 + \epsilon)) = o(n^{-\ell}).$$

Remark 1. *In theorem 2, it is shown that the largest eigenvalue of $\mathbf{B}$ (after the entries have been truncated) not only converges to 2 with probability one, but also with a polynomial rate of arbitrary order ℓ .*

Remark 2. *Theorems 1-2 are stated for the real random matrix $\mathbf{X}$, but they also hold for the complex case under moment conditions $EX_{11} = 0$, $E|X_{11}|^2 = 1$ and $E|X_{11}|^4 < \infty$. The proofs are similar to those for the real case except some notation changes.*

2.2 Proof of Theorem 1

Throughout the paper, all limits are taken as $n \rightarrow \infty$.

It follows from Theorem in Bai and Yin (1988) that

$$\liminf_{n \rightarrow \infty} \lambda_{\max}(\mathbf{B}) \geq 2 \quad a.s.. \quad (2.1)$$

Thus, it suffices to show that

$$\limsup_{n \rightarrow \infty} \lambda_{\max}(\mathbf{B}) \leq 2 \quad a.s.. \quad (2.2)$$

Let $\hat{\mathbf{B}} = \frac{1}{2\sqrt{np}}(\hat{\mathbf{X}}^T \hat{\mathbf{X}} - p\mathbf{I}_n)$, where $\hat{\mathbf{X}} = (\hat{X}_{ij})_{p \times n}$ and $\hat{X}_{ij} = X_{ij}I(|X_{ij}| \leq \delta_n \sqrt[4]{np})$ where δ_n is chosen as the larger of δ_n constructed as in (2.3) and δ_n as in (2.5). On the one hand, since $EX_{11}^4 < \infty$ for any $\delta > 0$ we have

$$\lim_{n \rightarrow \infty} \delta^{-4} E|X_{11}|^4 I(|X_{11}| > \delta \sqrt[4]{np}) = 0.$$

Since the above is true for arbitrary positive δ there exists a sequence of positive δ_n such that

$$\lim_{n \rightarrow \infty} \delta_n = 0, \quad \lim_{n \rightarrow \infty} \delta_n^{-4} E|X_{11}|^4 I(|X_{11}| > \delta_n \sqrt[4]{np}) = 0, \quad \delta_n \sqrt[4]{np} \uparrow +\infty. \quad (2.3)$$

On the other hand, since $EX_{11}^4 < \infty$ for any $\varepsilon > 0$

$$\sum_{k=1}^{\infty} 2^k P\left(|X_{11}| > \varepsilon 2^{\frac{k}{4}}\right) < \infty.$$

In view of the arbitrariness of ε there is a sequence of positive number ε_k such that

$$\varepsilon_k \rightarrow 0, \text{ as } k \rightarrow \infty, \quad \sum_{k=1}^{\infty} 2^k P\left(|X_{11}| > \varepsilon_k 2^{\frac{k}{4}}\right) < \infty. \quad (2.4)$$

For each k , let n_k be the maximum n such that $p(n) \cdot n \leq 2^k$. For $n_{k-1} < n \leq n_k$, set

$$\delta_n = 2\varepsilon_k. \quad (2.5)$$

Let $Z_t = X_{ij}, t = (i-1)p + j$ and obviously $\{Z_t\}$ are *i.i.d.* We then conclude from (2.4) and (2.5) that

$$\begin{aligned} P(\mathbf{B} \neq \hat{\mathbf{B}}, i.o.) &\leq \lim_{K \rightarrow \infty} P\left(\bigcup_{k=K}^{\infty} \bigcup_{n_{k-1} < n \leq n_k} \bigcup_{i \leq p, j \leq n} \{|X_{ij}| > \delta_n \sqrt[4]{np}\}\right) \\ &\leq \lim_{K \rightarrow \infty} \sum_{k=K}^{\infty} P\left(\bigcup_{n_{k-1} < n \leq n_k} \bigcup_{t=1}^{2^k} \{|Z_t| > \varepsilon_k 2^k\}\right) \\ &= \lim_{K \rightarrow \infty} \sum_{k=K}^{\infty} P\left(\bigcup_{t=1}^{2^k} \{|Z_t| > \varepsilon_k 2^k\}\right) \\ &\leq \lim_{K \rightarrow \infty} \sum_{k=K}^{\infty} 2^k P\left(|Z_1| > \varepsilon_k 2^{\frac{k}{4}}\right) \\ &= 0 \quad a.s. \end{aligned}$$

It follows that $\lambda_{\max}(\mathbf{B}) - \lambda_{\max}(\hat{\mathbf{B}}) \rightarrow 0 \quad a.s.$ as $n \rightarrow \infty$.

From now on we write δ for δ_n to simplify notation. Moreover, set $\tilde{\mathbf{B}} = \frac{1}{\sqrt{np}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} - p\mathbf{I}_n)$, where $\tilde{\mathbf{X}} = (\tilde{X}_{ij})_{p \times n}$ and $\tilde{X}_{ij} = \frac{\hat{X}_{ij} - E\hat{X}_{11}}{\sigma}$. Here $\sigma^2 = E(\hat{X}_{11} - E\hat{X}_{11})^2$ and $\sigma^2 \rightarrow 1$ as $n \rightarrow \infty$.

We obtain via (2.3)

$$|E\hat{X}_{11}| \leq \frac{E|X_{11}|^4 I(|X_{11}| > \delta \sqrt[4]{np})}{\delta^3 (np)^{3/4}} \leq \frac{K}{(np)^{3/4}}, \quad (2.6)$$

and

$$\begin{aligned} |\sigma^2 - 1| &\leq KE|X_{11}|^2 I(|X_{11}| > \delta \sqrt[4]{np}) \\ &\leq \frac{E|X_{11}|^4 I(|X_{11}| > \delta \sqrt[4]{np})}{\delta^2 \sqrt{np}} = o\left(\frac{1}{\sqrt{np}}\right). \end{aligned} \quad (2.7)$$

We conclude from the Rayleigh-Ritz theorem that

$$\begin{aligned} &|\lambda_{\max}(\tilde{\mathbf{B}}) - \lambda_{\max}(\hat{\mathbf{B}})| \\ &\leq \frac{1}{\sqrt{np}} \left| \sup_{\|\mathbf{z}\|=1} \left(\sum_{i \neq j} z_i z_j \sum_{k=1}^p \hat{X}_{ki} \hat{X}_{kj} + \sum_{i=1}^n z_i^2 \sum_{k=1}^p (\hat{X}_{ki}^2 - 1) \right) \right. \\ &\quad \left. - \sup_{\|\mathbf{z}\|=1} \left(\sum_{i \neq j} z_i z_j \sum_{k=1}^p \tilde{X}_{ki} \tilde{X}_{kj} + \sum_{i=1}^n z_i^2 \sum_{k=1}^p (\tilde{X}_{ki}^2 - 1) \right) \right| \\ &\leq \frac{1}{\sqrt{np}} \left| 1 - \frac{1}{\sigma^2} \right| \sup_{\|\mathbf{z}\|=1} \left| \sum_{i \neq j} z_i z_j \frac{1}{\sqrt{np}} \sum_{k=1}^p \hat{X}_{ki} \hat{X}_{kj} + \sum_{i=1}^n z_i^2 \sum_{k=1}^p (\hat{X}_{ki}^2 - 1) \right| \\ &\quad + \frac{1}{\sqrt{np}} \frac{2|EX_{11}|}{\sigma^2} \sup_{\|\mathbf{z}\|=1} \left| \sum_{i=1}^n \sum_{j=1}^n z_i z_j \sum_{k=1}^p \hat{X}_{ki} \right| \\ &\quad + \frac{1}{\sqrt{np}} \frac{p|EX_{11}|^2}{\sigma^2} \sup_{\|\mathbf{z}\|=1} \left| \sum_{i=1}^n \sum_{j=1}^n z_i z_j \right| + \frac{p}{2\sqrt{np}} \left| 1 - \frac{1}{\sigma^2} \right| \\ &= A_1 + A_2 + A_3 + A_4. \end{aligned}$$

By (2.7) and the strong law of large numbers, we have

$$\begin{aligned} A_1 &= \frac{|\sigma^2 - 1|}{\sqrt{np}\sigma^2} \sup_{\|\mathbf{z}\|=1} \left| \sum_{k=1}^p \left(\left(\sum_{i=1}^n z_i \hat{X}_{ki} \right)^2 - \sum_{i=1}^n z_i^2 \hat{X}_{ki}^2 \right) + \sum_{i=1}^n z_i^2 \sum_{k=1}^p (\hat{X}_{ki}^2 - 1) \right| \\ &\leq \frac{|\sigma^2 - 1|\sqrt{np}}{\sigma^2} \cdot \frac{1}{np} \left(2 \left| \sum_{k=1}^p \sum_{i=1}^n \hat{X}_{ki}^2 \right| + \left| \sum_{i=1}^n \sum_{k=1}^p (\hat{X}_{ki}^2 - 1) \right| \right) \\ &\leq \frac{|\sigma^2 - 1|\sqrt{np}}{\sigma^2} \cdot \frac{1}{np} \left(3 \left| \sum_{k=1}^p \sum_{i=1}^n \hat{X}_{ki}^2 \right| + np \right) \\ &\rightarrow 0 \quad a.s.. \end{aligned}$$

Similarly, (2.6), Hölder's inequality and the strong law of large numbers

yield

$$\begin{aligned}
A_2 &\leq \frac{1}{\sqrt{np}} \cdot \frac{2|E\hat{X}_{11}|}{\sigma^2} \sup_{\|\mathbf{z}\|=1} \left| \sum_{j=1}^n z_j \right| \left| \sum_{i=1}^n z_i \sum_{k=1}^p \hat{X}_{ki} \right| \\
&\leq \frac{1}{\sqrt{np}} \cdot \frac{K}{\sigma^2(np)^{3/4}} \cdot \sqrt{n} \cdot \left(\sum_{i=1}^n \left(\sum_{k=1}^p \hat{X}_{ki} \right)^2 \right)^{1/2} \\
&\leq \frac{1}{\sqrt{np}} \cdot \frac{K}{\sigma^2(np)^{3/4}} \cdot \sqrt{n} \cdot \left(p \sum_{i=1}^n \sum_{k=1}^p \hat{X}_{ki}^2 \right)^{1/2} \\
&\leq \frac{K}{\sigma^2(np)^{1/4}} \left| \frac{1}{np} \sum_{i=1}^n \sum_{k=1}^p \hat{X}_{ki}^2 \right|^{1/2} \\
&\leq \frac{K}{\sigma^2(np)^{1/4}} \left| \frac{1}{np} \sum_{i=1}^n \sum_{k=1}^p X_{ki}^2 \right|^{1/2} \rightarrow 0 \quad a.s..
\end{aligned}$$

It is straightforward to conclude from (2.6) and (2.7) that

$$A_3 \rightarrow 0 \quad a.s. \quad A_4 \rightarrow 0 \quad a.s.$$

Thus, we have $\lambda_{\max}(\hat{\mathbf{B}}) - \lambda_{\max}(\tilde{\mathbf{B}}) \rightarrow 0 \quad a.s..$ By the above results, to prove (2.2), it is sufficient to show that $\limsup_{n \rightarrow \infty} \lambda_{\max}(\tilde{\mathbf{B}}) \leq 2 \quad a.s..$ To this end, we note that the matrix $\tilde{\mathbf{B}}$ satisfies all the assumptions in Theorem 2. Therefore we obtain (2.2) by Theorem 2 (whose argument is given in the next section). Together with (2.1), we finishes the proof of Theorem 1.

2.3 Proof of Theorem 2

Suppose that $\mathbf{z} = (z_1, \dots, z_n)$ is a unit vector. By the Rayleigh-Ritz theorem we then have

$$\begin{aligned}\lambda_{\max}(\mathbf{B}) &= \max_{\|\mathbf{z}\|=1} \left(\sum_{i,j} z_i z_j B_{ij} \right) \\ &= \max_{\|\mathbf{z}\|=1} \left(\sum_{i \neq j} z_i z_j B_{ij} + \sum_{i=1}^p z_i^2 B_{ii} \right) \\ &\leq \lambda_{\max}(\mathbf{G}) + \max_{i \leq n} |B_{ii}|,\end{aligned}$$

where $\mathbf{G} = (G_{ij})_{n \times n}$ with

$$G_{ij} = \begin{cases} 0, & \text{if } i = j, \\ \frac{1}{\sqrt{np}} \sum_{k=1}^p X_{ki} X_{kj}, & \text{if } i \neq j. \end{cases}$$

To prove Theorem 2, it is sufficient to prove, for any $\epsilon > 0, \ell > 0$

$$P(\lambda_{\max}(\mathbf{G}) > 2(1 + \epsilon)) = o(n^{-\ell}) \quad (2.8)$$

and

$$P\left(\max_{i \leq n} \frac{1}{\sqrt{np}} \left| \sum_{k=1}^p (X_{ki}^2 - 1) \right| > \epsilon\right) = o(n^{-\ell}). \quad (2.9)$$

We first prove (2.9). To simplify notation, let $Y_i = X_{i1}^2 - 1$ and $K_1 = E|Y_1|^2$. Then $EY_i = 0$. Choose an appropriate sequence $h = h_n$ such that it satisfies, as $n \rightarrow \infty$

$$\begin{cases} h/\log n \rightarrow \infty \\ \delta^2 h/\log n \rightarrow 0 \\ \frac{\delta^4 n}{K_1} \geq \sqrt{n}. \end{cases} \quad (2.10)$$

We then have

$$\begin{aligned}
P(\max_{j \leq n} \frac{1}{\sqrt{np}} |\sum_{i=1}^p (X_{ij}^2 - 1)| > \epsilon) &\leq n \cdot P\left(|\sum_{i=1}^p (X_{i1}^2 - 1)| > \epsilon \sqrt{np}\right) \\
&\leq \epsilon^{-h} n (\sqrt{np})^{-h} E|\sum_{i=1}^p Y_i|^h \\
&\leq \epsilon^{-h} n (\sqrt{np})^{-h} \sum_{m=1}^{h/2} \sum_{1 \leq i_1 < i_2 < \dots < i_m \leq n} \sum_{\substack{j_1+j_2+\dots+j_m=h \\ j_1 \geq 2, \dots, j_m \geq 2}} \frac{h!}{j_1! j_2! \dots j_m!} E|Y_{i_1}|^{j_1} E|Y_{i_2}|^{j_2} \dots E|Y_{i_m}|^{j_m} \\
&\leq \epsilon^{-h} n (\sqrt{np})^{-h} \sum_{m=1}^{h/2} \sum_{\substack{j_1+j_2+\dots+j_m=h \\ j_1 \geq 2, \dots, j_m \geq 2}} \frac{p!}{m!(p-m)!} \frac{h!}{j_1! j_2! \dots j_m!} E|Y_1|^{j_1} E|Y_1|^{j_2} \dots E|Y_1|^{j_m} \\
&\leq \epsilon^{-h} n (\sqrt{np})^{-h} \sum_{m=1}^{h/2} \sum_{\substack{j_1+j_2+\dots+j_m=h \\ j_1 \geq 2, \dots, j_m \geq 2}} p^m \frac{h!}{j_1! j_2! \dots j_m!} K_1^m (\delta^2 \sqrt{np})^{h-2m} \\
&\leq \epsilon^{-h} n \sum_{m=1}^{h/2} m^h \left(\frac{\delta^4 n}{K_1}\right)^{-m} \delta^{2h} \leq \epsilon^{-h} n \frac{h}{2} \cdot \left(\frac{\delta^2 h}{\log(\delta^4 n/K_1)}\right)^h \\
&\leq \left(\left(\frac{nh}{2}\right)^{1/h} \cdot \frac{2\delta^2 h}{\log n} \cdot \epsilon^{-1}\right)^h \leq \left(\frac{\xi}{\epsilon}\right)^h = o(n^{-\ell}),
\end{aligned}$$

where ξ is a constant satisfying $0 < \xi < \epsilon$. Below are some interpretations of the above inequalities:

- a) The fifth inequality is because, $\frac{p!}{m!(p-m)!} < n^m$, $|Y_1| < \delta^2 \sqrt{np}$.
- b) We use the fact $\sum_{\substack{j_1+j_2+\dots+j_m=h \\ j_1 \geq 2, \dots, j_m \geq 2}} \frac{h!}{j_1! j_2! \dots j_m!} < m^h$ in the sixth inequality.
- c) The seventh inequality uses the elementary inequality

$$a^{-t} t^b \leq \left(\frac{b}{\log a}\right)^b, \quad \text{for all } a > 1, b > 0, t \geq 1 \quad \text{and} \quad \frac{b}{\log a} > 1.$$

- d) The last two inequalities are due to (2.10).

- e) With the facts that $\frac{\xi}{\epsilon} < 1$, $h/\log n \rightarrow \infty$, the last equality is true.

Thus (2.9) follows.

Next consider (2.8). For any $\varepsilon > 0$, we have

$$\begin{aligned} P(\lambda_{\max}(\mathbf{G}) \geq 2(1 + \varepsilon)) &\leq \frac{E\lambda_{\max}^k(\mathbf{G})}{2^k(1 + \varepsilon)^k} \leq \frac{Etr(\mathbf{G}^k)}{2^k(1 + \varepsilon)^k} \\ &\leq \frac{1}{2^k(1 + \varepsilon)^k} \cdot \frac{1}{(\sqrt{np})^k} \sum E(X_{i_1 j_1} X_{i_1 j_2} X_{i_2 j_2} X_{i_2 j_3} \dots X_{i_k j_k} X_{i_k j_1}), \end{aligned}$$

where $k = k_n$ satisfies, as $n \rightarrow \infty$

$$\left\{ \begin{array}{l} k/\log n \rightarrow \infty \\ \delta^{1/3}k/\log n \rightarrow 0 \quad , \\ \frac{\delta^2 \sqrt[4]{n}}{k^3} \geq 1 \end{array} \right.$$

and the summation is taken with respect to $i_1, i_2, \dots, i_k$ running over all integers in $\{1, 2, \dots, p\}$ and $j_1, j_2, \dots, j_k$ running over all integers in $\{1, 2, \dots, n\}$ subject to the condition that $j_1 \neq j_2, j_2 \neq j_3, \dots, j_k \neq j_1$.

In order to get an up bound for $|\sum EX_{i_1 j_1} X_{i_1 j_2} \dots X_{i_k j_k} X_{i_k j_1}|$, we need to construct a graph for given $i_1, \dots, i_k$ and $j_1, \dots, j_k$, as in Geman (1980), Yin, Bai and Krishnaiah (1988) and Bai and Yin (1993). We follow the presentation in Bai and Yin (1993) and Yin, Bai and Krishnaiah (1988) to introduce some fundamental concepts associated with the graph.

For the sequence $(i_1, i_2, \dots, i_k)$ from $\{1, 2, \dots, p\}$ and the sequence $(j_1, j_2, \dots, j_k)$ from $\{1, 2, \dots, n\}$, we define a directed graph as follows. Plot two parallel real lines, referred to as *I-line* and *J-line*, respectively. Draw $\{i_1, i_2, \dots, i_k\}$ on the *I-line*, called *I-vertices* and draw $\{j_1, j_2, \dots, j_k\}$ on the *J-line*, known as *J-vertices*. The vertices of the graph consist of the *I-vertices* and *J-vertices*. The edges of the graph are $\{e_1, e_2, \dots, e_{2k}\}$, where for $a = 1, \dots, k$, $e_{2a-1} = j_a i_a$ are called the row edges and $e_{2a} = i_a j_{a+1}$ are called column edges with the convention that $j_{2k+1} = j_1$. For each row edge e_{2a-1} , the vertices j_a and i_a are called the ends of the edge $j_a i_a$ and moreover j_a and i_a are, re-

spectively, the initial and the terminal of the edge $j_a i_a$. Each row edge e_{2a} starts from the vertex i_b and ends with the vertex j_{b+1} .

Two vertices are said to coincide if they are both in the *I-line* or both in the *J-line* and they are identical. That is $i_a = i_b$ or $j_a = j_b$. Readers are also reminded that the vertices i_a and j_b are not coincident even if they have the same value because they are in different lines. We say that two edges are coincident if two edges have the same set of ends.

The graph constructed above is said to be a *W-graph* if each edge in the graph coincides with at least one other edge. Below is an example of a *W-graph*:

Figure 2.1: A example of W-graph

Two graphs are said to be isomorphic if one becomes another by an appropriate permutation on $\{1, 2, \dots, p\}$ of *I-vertices* and an appropriate permutation on $\{1, 2, \dots, n\}$ of *J-vertices*. A *W-graph* is called a canonical graph if $i_a \leq \max\{i_1, i_2, \dots, i_{a-1}\} + 1$ and $j_a \leq \max\{j_1, j_2, \dots, j_{a-1}\} + 1$ with $i_1 = j_1 = 1$, where $a = 1, 2, \dots, k$.

In the canonical graph, if $i_a = \max\{i_1, i_2, \dots, i_{a-1}\} + 1$, then the edge $j_a i_a$ is called a row innovation and if $j_{a+1} = \max\{j_1, j_2, \dots, j_a\} + 1$, then the edge $i_a j_{a+1}$ is called a column innovation. Apparently, a row innovation and a

column innovation, respectively, lead to a new I-vertex and a new J-vertex except the first row innovation $j_1 i_1$ leading to a new J-vertex j_1 and a new I-vertex i_1 .

We now classify all edges into three types, T_1 , T_3 and T_4 . Let T_1 denote the set of all innovations including row innovations and column innovations. We further distinguish the row innovations as follows. An edge $j_a i_a$ is called a T_{11} edge if it is a row innovation and the edge $i_a j_{a+1}$ is a column innovation; An edge $j_b i_b$ is referred to as a T_{12} edge if it is a row innovation but $i_b j_{b+1}$ is not a column innovation. An edge e_j is said to be a T_3 edge if there is an innovation edge $e_i, i < j$ so that e_j is the first one to coincide with e_i . An edge is called a T_4 edge if it does not belong to a T_1 edge or T_3 edge. The first appearance of a T_4 edge is referred to as a T_2 edge. There are two kinds of T_2 edges: (a) the first appearance of an edge that coincides with a T_3 edge, denoted by T_{21} edge; (b) the first appearance of an edge that is not an innovation, denoted by T_{22} edge.

We say that an edge e_i is single up to the edge $e_j, j \geq i$ if it does not coincide with any other edges among $e_1, \dots, e_j$ except itself. A T_3 edge e_i is said to be regular if there are more than one innovations with a vertex equal to the initial vertex of e_i and single up to e_{i-1} , among the edges $\{e_1, \dots, e_{i-1}\}$. All other T_3 edges are called irregular T_3 edges.

Corresponding to the above classification of the edges, we introduce the following notation and list some useful facts.

1. Denote by l the total number of innovations.
2. Let r be the number of the column innovations. Moreover, let c denote the row innovations. We then have $r + c = l$.
3. Define r_1 to be the number of the T_{11} edges. Then $r_1 \leq r$ by the

definition of a T_{11} edge. Also, the number of the T_{12} edges is $l - r - r_1$

4. Let t be the number of the T_2 edges. Note that the number of the T_3 edges is the same as the number of the innovations and there are a total of $2k$ edges in the graph. It follows that the number of the T_4 edges is $2k - 2l$. On the other hand, each T_2 edge is also a T_4 edge. Therefore $t \leq 2k - 2l$.
5. Define μ to be the number of T_{21} edges. Obviously, $\mu \leq t$. The number of T_{22} edge is then $t - \mu$. Since each T_{21} edge coincides with one innovation, we let $n_i, i = 1, 2, \dots, \mu$ denote the number of T_4 edges which coincide with the i -th such innovation, $n_i \geq 0$.
6. Let μ_1 be the number of T_{21} edges which do not coincide with the other T_4 edges. That is $\mu_1 = \#\{i : n_i = 1, i = 1, 2, \dots, \mu\}$, where $\#\{\cdot\}$ denotes the cardinality of the set $\{\cdot\}$.
7. Let $m_j, j = 1, 2, \dots, t - \mu$ denote the number of T_4 edges which coincide with and include the j -th T_{22} edge. Note that $m_j \geq 2$.

We now claim that

$$\begin{aligned}
Etr(\mathbf{G}^k) &\leq (\sqrt{np})^{-k} \sum E(X_{i_1 j_1} X_{i_1 j_2} \dots X_{i_k j_k} X_{i_k j_1}) \\
&= (\sqrt{np})^{-k} \sum' \sum'' \sum''' \sum_{*} E(X_{i_1 j_1} X_{i_1 j_2} \dots X_{i_k j_k} X_{i_k j_1}) \\
&\leq (\sqrt{np})^{-k} \sum_{l=1}^k \sum_{r=1}^l \sum_{r_1=0}^r \sum_{t=0}^{2k-2l} \sum_{\mu=0}^t \sum_{\mu_1=0}^{\mu} \sum_{*} \binom{k}{r} \binom{r}{r_1} \binom{k-r_1}{l-r-r_1} \\
&\quad \times \binom{2k-l}{l} k^{3t} (t+1)^{6k-6l} (\delta \sqrt[4]{np})^{2k-2l-2t+\mu_1} n^{r+1} p^{l-r},
\end{aligned} \tag{2.11}$$

where the summation $\sum'$ is with respect to different arrangements of three types of edges at the $2k$ different positions, the summation $\sum''$ over different

canonical graphs with a given arrangement of the three types of edges for $2k$ positions, the third summation $\sum'''$ with respect to all isomorphic graphs for a given canonical graph and the last notation $\sum_*$ denotes the constraint that $j_1 \neq j_2, j_2 \neq j_3, \dots, j_k \neq j_1$.

Now, we explain why the above estimate is true:

- (i). The factor $(\sqrt{np})^{-k}$ is obvious.
- (ii). If the graph is not a *W-graph*, which means there is a single edge in the graph, then the mean of the product of X_{ij} corresponding to this graph is zero (since $EX_{11} = 0$). Thus we have $l \leq k$. Moreover, the facts that $r \leq l$, $r_1 \leq r$, $t \leq 2k - 2l$, $\mu \leq t$ and $\mu_1 \leq \mu$ are easily obtained from the fact 1 to the fact 7 listed before.
- (iii). There are at most $\binom{k}{r}$ ways to choose r edges out of the k column edges to be the r column innovations. Subsequently, we consider how to select the row innovations. Observe that the definition of T_{11} edges, there are $\binom{r}{r_1}$ ways to select r_1 column innovations out of the total r column innovations so that the edge before each such r_1 column innovations is a T_{11} edge, row innovation. Moreover there are at most $\binom{k-r_1}{l-r-r_1}$ ways to choose $l-r-r_1$ edges out of the remaining $k-r_1$ row edges to be the $l-r-r_1$ T_{12} edges, the remaining row innovations.
- (iv). Given the position of the l innovations, there are at most $\binom{2k-l}{l}$ ways to select l edges out of the $2k-l$ edges to be T_3 edges. And the rest positions are for the T_4 edges. Therefore the first summation $\sum'$ is bounded by $\sum_{l=1}^k \sum_{r=1}^l \sum_{r_1=0}^r \binom{k}{r} \binom{r}{r_1} \binom{k-r_1}{l-r-r_1} \binom{2k-l}{l}$.
- (v). By definition, each innovation (or each irregular T_3 edges) is uniquely determined by the subgraph prior to the innovation (or the irregular

T_3). Moreover by Lemma 3.2 in Yin, Bai and Krishnaiah (1988) for each regular T_3 edge, there are at most $t + 1$ innovations so that the regular T_3 edge coincides with one of them and by Lemma 3.3 in Yin, Bai and Krishnaiah (1988) there are at most $2t$ regular T_3 edges. Therefore there are at most $(t + 1)^{2t} \leq (t + 1)^{2(2k-2l)}$ ways to draw the regular T_3 edges.

- (vi). Once the positions of the innovations and the T_3 edges are fixed there are at most $\binom{(r+1)c}{t} \leq \binom{k^2}{t} \leq k^{2t}$ ways to arrange the t T_2 edges, as there are $r + 1$ J -vertices and c I -vertices. After t positions of T_2 edges are determined there are at most t^{2k-2l} ways to distribute $2k - 2l$ T_4 edges among the t positions. So there are at most $k^{2t} \cdot t^{2k-2l}$ ways to arrange T_4 edges. It follows that $\sum''$ is bounded by $\sum_{t=0}^{2k-2l} (t + 1)^{2(2k-2l)} k^{2t} \cdot t^{2k-2l}$.
- (vii). The third summation $\sum'''$ is bounded by $p^c n^{r+1}$ because the number of graphs in the isomorphic class for a given graph is $p(p - 1) \cdots (p - c + 1)n(n - 1) \cdots (n - r)$.

- (viii). Recalling the definitions of $l, r, t, \mu, \mu_1, n_i, m_i$ we have

$$\begin{aligned} & EX_{i_1 j_1} X_{i_1 j_2} \cdots X_{i_k j_k} X_{i_k j_1} \\ &= (EX_{11}^2)^{l-\mu} \left(\prod_{i=1}^{\mu} EX_{11}^{n_i+2} \right) \left(\prod_{i=1}^{t-\mu} EX_{11}^{m_i} \right), \end{aligned} \quad (2.12)$$

where $\sum_{i=1}^{\mu} n_i + \sum_{i=1}^{t-\mu} m_i = 2k - 2l$. Without loss of generality, we suppose $n_1 = n_2 = \dots = n_{\mu_1} = 1$ and $n_{\mu_1+1}, \dots, n_{\mu} \geq 2$ for convenience.

It is easy to check that

$$E|X_{11}^s| \leq \begin{cases} M(\delta \sqrt[4]{np})^{s-4}, & \text{if } s \geq 4, M = \max\{EX_{11}^4, |EX_{11}^3|\} \\ (\delta \sqrt[4]{np})^{s-2}, & \text{if } s \geq 2. \end{cases}$$

Thus, (2.12) becomes

$$\begin{aligned}
& |EX_{i_1 j_1} X_{i_1 j_2} \dots X_{i_k j_k} X_{i_k j_1}| \\
& \leq \sum_{\mu=0}^t \sum_{\mu_1=0}^{\mu} |EX_{11}^3|^{\mu_1} |EX_{11}^4|^{t-\mu_1} (\delta \sqrt[4]{np})^{\sum_{i=\mu_1+1}^{\mu} n_i - 2(\mu - \mu_1)} (\delta \sqrt[4]{np})^{\sum_{i=1}^{t-\mu} m_i - 2(t-\mu)} \\
& \leq \sum_{\mu=0}^t \sum_{\mu_1=0}^{\mu} M^t (\delta \sqrt[4]{np})^{2k-2l-2t+\mu_1} \\
& \leq \sum_{\mu=0}^t \sum_{\mu_1=0}^{\mu} k^t (\delta \sqrt[4]{np})^{2k-2l-2t+\mu_1}, \quad \text{when } k \text{ is large enough.}
\end{aligned} \tag{2.13}$$

The above points regarding the T_2 edges are discussed for $t > 0$, but they are still valid when $t = 0$ with the convention that $0^0 = 1$ in the term t^{2k-2l} , because in this case there are only T_1 edges and T_3 edges in the graph and thus $l = k$.

Consider the constraint $\sum_{*}$ now. Note that for each T_{12} edge, say $j_a i_a$, it is a row innovation, but the next column edge $i_a j_{a+1}$ is not a column innovation. Since $j_{a+1} \neq j_a$, the edge $i_a j_{a+1}$ cannot coincide with the edge $j_a i_a$. Moreover, it also doesn't coincide with any edges before the edge $j_a i_a$ since i_a is a new vertex. So $i_a j_{a+1}$ must be a T_{22} edge. Thus, the number of the T_{12} edges can not exceed the number of the T_{22} edges. This implies $l - r - r_1 \leq t - \mu$. Moreover, note that $\mu_1 \leq \mu$. We then have

$$\begin{aligned}
& n^{-k/2} p^{-k/2} p^{l-r} n^{r+1} (np)^{k/2-l/2-t/2+\mu_1/4} \\
& = (p/n)^{l/2} \cdot p^{-r-t/2+\mu_1/4} n^{r+1-t/2+\mu_1/4} \\
& \leq \left(\sqrt{\frac{n}{p}}\right)^{r-r_1} \cdot n^{-t/2} n.
\end{aligned} \tag{2.14}$$

We thus conclude from (2.11) and (2.14) that

$$\begin{aligned}
Etr(\mathbf{G}^k) & \leq \sum_{l=1}^k \sum_{r=1}^l \sum_{r_1=0}^r \sum_{t=0}^{2k-2l} \sum_{\mu=0}^t \sum_{\mu_1=0}^{\mu} \binom{k}{r} \binom{r}{r_1} \binom{k-r_1}{l-r-r_1} \binom{2k-l}{l} \\
& \quad \times \left(\sqrt{\frac{n}{p}}\right)^{r-r_1} n^{-t/2} n k^{3t} (t+1)^{6k-6l} \delta^{2k-2l-2t+\mu_1}.
\end{aligned} \tag{2.15}$$

Moreover we claim that

$$\begin{aligned}
& n \left[\binom{k}{r} \right] \left[\binom{r}{r_1} \left(\sqrt{\frac{n}{p}} \right)^{r-r_1} \right] \left[\binom{k-r_1}{l-r-r_1} \delta^{l-r-r_1} \right] \\
& \quad \times \left[\binom{2k-l}{l} \left(\frac{\sqrt{n}\delta^3}{k^3} \right)^{-t} (t+1)^{6k-6l} \delta^{2k-2l} \right] \\
& \quad \times \delta^{-(l-r-r_1)+3t-(2k-2l)} \cdot \delta^{2k-2l-2t+\mu_1} \\
& \leq 2^k n^2 \left(1 + \sqrt{\frac{n}{p}} \right)^k (1+\delta)^k \left(1 + \frac{24^3 k^3 \delta}{\log^3 n} \right)^{2k}.
\end{aligned} \tag{2.16}$$

Indeed, the above claim is based on the following five facts.

- 1) $\binom{k}{r} \leq \sum_{r=0}^k \binom{k}{r} = 2^k$.
- 2) $\binom{r}{r_1} \left(\sqrt{\frac{n}{p}} \right)^{r-r_1} = \binom{r}{r-r_1} \left(\sqrt{\frac{n}{p}} \right)^{r-r_1} \leq \sum_{s=0}^r \binom{r}{s} \left(\sqrt{\frac{n}{p}} \right)^s = \left(1 + \sqrt{\frac{n}{p}} \right)^r \leq \left(1 + \sqrt{\frac{n}{p}} \right)^k$.
- 3) $\binom{k-r_1}{l-r-r_1} \delta^{l-r-r_1} \leq \sum_{s=0}^{k-r_1} \binom{k-r_1}{s} \delta^s = (1+\delta)^{k-r_1} \leq (1+\delta)^k$.
- 4) By the fact that $\binom{2k-l}{l} \leq \binom{2k}{2l}$, and the inequality $a^{-t}(t+1)^b \leq a \left(\frac{b}{\log a} \right)^b$, for $a > 1$, $b > 0$, $t > 0$ and $\frac{\delta^2 \sqrt{n}}{k^3} \geq \sqrt[4]{n}$, we have

$$\begin{aligned}
& \binom{2k-l}{l} \left(\frac{\sqrt{n}\delta^3}{k^3} \right)^{-t} (t+1)^{6k-6l} \delta^{2k-2l} \\
& \leq \binom{2k}{2l} \frac{\sqrt{n}\delta^3}{k^3} \left(\frac{6k-6l}{\log(\sqrt{n}\delta^3/k^3)} \right)^{6k-6l} \delta^{2k-2l} \\
& \leq n \binom{2k}{2l} \left(\frac{24k}{\log n} \right)^{6k-6l} \delta^{2k-2l} \\
& \leq n \binom{2k}{2l} \left(\frac{24^3 k^3 \delta}{\log^3 n} \right)^{2k-2l} \\
& \leq n \sum_{s=0}^{2k} \binom{2k}{s} \left(\frac{24^3 k^3 \delta}{\log^3 n} \right)^{2k-s} \\
& = n \left(1 + \frac{24^3 k^3 \delta}{\log^3 n} \right)^{2k}.
\end{aligned}$$

5) When n is large enough, $\delta^{-(l-r-r_1)+3t-(2k-2l)} \cdot \delta^{2k-2l-2t+\mu_1} = \delta^{t-(l-r-r_1)}$.

$\delta^{\mu_1} \leq 1$, since $\delta \rightarrow 0$ and $l-r-r_1 \leq t-\mu$.

Summarizing (2.15) and (2.16) we obtain that

$$\begin{aligned}
 Etr(\mathbf{G}^k) &\leq \sum_{l=1}^k \sum_{r=1}^l \sum_{r_1=0}^r \sum_{t=0}^{2k-2l} \sum_{\mu=0}^t \sum_{\mu_1=0}^{\mu} 2^k n^2 \left(1 + \sqrt{\frac{n}{p}}\right)^k (1+\delta)^k \left(1 + \frac{24^3 l^3 \delta}{\log^3 n}\right)^{2k} \\
 &\leq 2^k \cdot 8k^6 n^2 \left(1 + \sqrt{\frac{n}{p}}\right)^k (1+\delta)^k \left(1 + \frac{24^3 l^3 \delta}{\log^3 n}\right)^{2k} \\
 &= \left(2(8k^6)^{1/k} n^{2/k} (1 + \sqrt{\frac{n}{p}}) (1+\delta) \left(1 + \frac{24^3 k^3 \delta}{\log^3 n}\right)^2\right)^k \\
 &\leq (2\eta)^k,
 \end{aligned}$$

where η is a constant satisfying $1 < \eta < 1 + \epsilon$. Here the last inequality uses the facts below:

- i. $(n^2)^{1/k} \rightarrow 1$, because $k/\log n \rightarrow \infty$,
- ii. $(8k^6)^{1/k} \rightarrow 1$, because $k \rightarrow \infty$,
- iii. $\left(1 + \sqrt{\frac{n}{p}}\right) \rightarrow 1$, because $n/p \rightarrow 0$,
- iv. $(1 + \delta) \rightarrow 1$, because $\delta \rightarrow 0$,
- v. $\frac{24^3 \cdot k^3 \delta}{\log^3 n} \rightarrow 0$, because $\frac{\delta^{1/3} k}{\log n} \rightarrow 0$.

It follows that

$$P(\lambda_{max}(\mathbf{G}) > 2(1 + \epsilon)) \leq \left(\frac{\eta}{1 + \epsilon}\right)^k = o(n^{-\ell})$$

since $k/\log n \rightarrow \infty$ and $\frac{\eta}{1+\epsilon} < 1$. The proof is complete.

Chapter 3

Central limit theorem for the linear spectral statistics

As mentioned in the introduction, many important statistics in multivariate analysis as well as in wireless communication can be written as functional of ESD of some random matrices. Consider the linear spectral statistic $\hat{\theta} = \int f(x) dF^{\mathbf{S}}(x) = \frac{1}{n} \sum_{k=1}^p f(\lambda_k(\mathbf{S}))$, which is associated with the sample covariance matrix $\mathbf{S}$. The statistic $\hat{\theta}$ can be viewed as a estimator of $\theta = \int f(x) dF_c(x)$, where $F_c(x)$ is the LSD of $\mathbf{S}$. For testing hypothesis about θ , the limiting distribution of $t(n) \int f(x) d(F^{\mathbf{S}}(x) - F_c(x))$ is necessary to construct the critical region, where $t(n)$ is a suitable normalizer.

Another example is the statistic $I_n(\rho) = \frac{1}{p} \log \left(\det(\mathbf{S} + \rho \mathbf{I}_p) \right) = \frac{1}{p} \sum_{k=1}^p \log \left(\lambda_k(\mathbf{S}) + \rho \right)$, where $\rho > 0$ is a given parameter. This statistic known as the mutual information for multiple antenna radio channels is very popular in wireless communication. Understanding its fluctuations and in particular being able to approximate its standard deviation is of major interest for various purposes such as the computation of the so-called outage probability.

Gaussian fluctuations in random matrices are investigated by different

Chapter 3. Central limit theorem for the linear spectral statistics

authors, starting with Costin and Lebowitz (1995). Johansson (1998) considered an extended random ensembles whose entries follow a specific class of densities and established the CLT of the linear spectral statistics (LSS). Recently, the CLT for LSS of sample covariance matrices is studied by Bai and Silverstein (2004) and of Wigner matrices is studied by Bai and Yao (2005).

In this chapter, we turn our attention to the normalized sample covariance matrix

$$\mathbf{B} = \frac{1}{\sqrt{np}}(\mathbf{X}^T \mathbf{X} - p\mathbf{I}) \quad (3.1)$$

under the framework “ $p \rightarrow \infty, n \rightarrow \infty, p/n \rightarrow \infty$ ”. As the matrix $\mathbf{B}$ is the same as $\mathbf{A}$ if we interchange the role of p and n , the ESD $F^{\mathbf{B}}$ converges to the semicircle law F_{sc} almost surely whose density is given in (1.3). From the result of Chapter 2, we also have $\lambda_{\max}(\mathbf{B}) \xrightarrow{a.s.} 2$.

The main contribution of this chapter is summarized as follows:

- The central limit theorem of linear spectral statistic of the eigenvalues of this normalized sample covariance matrices is established for the case $p/n \rightarrow 0$. Such an asymptotic theory complements the results of Bai and Silverstein (2004) for the case $p/n \rightarrow c \in (0, \infty)$ and Bai and Yao (2005) for Wigner matrix.
- Noticing the close relationship between $\mathbf{B}$ and the Wigner matrix, we can conjecture the CLT for LSS of $\mathbf{B}$ from that of Wigner matrix in Bai and Yao (2005) for $n^3/p \rightarrow 0$. This paper provide a rigorous proof of this conjecture. What’s more, we find a correction term in (3.2) which can not be guessed by the result of Bai and Yao (2005).
- The results in this chapter have many potential applications covering the so-call “very large (or ultra) p and small n ” situation. One exam-

ple provided in the next chapter is to test $H_0 : \Sigma = \mathbf{I}$ vs $H_1 : \Sigma \neq \mathbf{I}$. The proposed test statistic is compared with Chen, Zhang and Zhong (2010) and shows its advantage of power when n is extremely small and p is large.

This chapter is organized as follows. The preliminary results are formulated in Section 3.1. Section 3.2 gives the strategy of proving Theorem 3 and two intermediate results, Proposition 1 and 2, and truncation steps of the underlying random variables are given as well. Sections 3.3-3.4 are devoted to the proof of Proposition 1. We present the proof of Proposition 2 in Section 3.5. Section 3.6 derives mean and covariance in Theorem 3. Section 3.7 provides a calibration of the mean correction term in (3.7) and simulation to check the accuracy of the calibrated CLTs in Theorem 3. Some useful lemmas are presented in the last section.

3.1 Preliminary Results

In order to study CLT of the linear functions of eigenvalues of $\mathbf{B}$, let $\mathcal{S}$ denote any open region on the real plane including $[-2, 2]$, which is the support of $F(x)$, and $\mathcal{M}$ be the set of functions analytic on $\mathcal{S}$. For any $f \in \mathcal{M}$, define

$$Q_n(f) \triangleq n \int_{-\infty}^{+\infty} f(x) d(F^{\mathbf{B}}(x) - F_{sc}(x)) - \frac{1}{\pi} \sqrt{\frac{n^3}{p}} \int_{-1}^1 f(2x) \frac{4x^3 - 3x}{\sqrt{1 - x^2}} dx, \quad (3.2)$$

and its random part

$$Q_n^{(1)}(f) \triangleq n \int_{-\infty}^{+\infty} f(x) d(F^{\mathbf{B}}(x) - EF^{\mathbf{B}}(x)). \quad (3.3)$$

Chapter 3. Central limit theorem for the linear spectral statistics

Let $\{T_k\}$ be the family of Chebyshev polynomials, which is defined as $T_0(x) = 1, T_1(x) = x$ and $T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x)$. To give an alternative way of calculating the asymptotic covariance of $X(f)$ in Theorem 3 below, for any $f \in \mathcal{M}$ and any integer $k > 0$, we define

$$\begin{aligned}\Psi_k(f) &\triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} f(2 \cos \theta) e^{ik\theta} d\theta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(2 \cos \theta) \cos k\theta d\theta = \frac{1}{\pi} \int_{-1}^1 f(2x) T_k(x) \frac{1}{\sqrt{1-x^2}} dx.\end{aligned}$$

The main result is formulated in the following theorem.

Theorem 3. *Suppose that*

(a) $\mathbf{X} = (X_{ij})_{p \times n}$ where $\{X_{ij} : i = 1, 2, \dots, p; j = 1, 2, \dots, n\}$ are i.i.d. real random variables with $EX_{11} = 0, EX_{11}^2 = 1$ and $\nu_4 = EX_{11}^4 < \infty$.

(b1) $n^3/p = O(1)$ as $n \rightarrow \infty$.

Then, for any $f_1, \dots, f_k \in \mathcal{M}$, the finite dimensional random vector $(Q_n(f_1), \dots, Q_n(f_k))$ converges weakly to a Gaussian vector $(X(f_1), \dots, X(f_k))$ with mean function

$$\begin{aligned}EX(f) &= \frac{1}{\pi} \int_{-1}^1 f(2x) \left[2(\nu_4 - 3)x^2 - (\nu_4 - \frac{5}{2}) \right] \frac{1}{\sqrt{1-x^2}} dx \\ &\quad + \frac{1}{4}(f(2) + f(-2))\end{aligned}\tag{3.4}$$

and covariance function

$$\text{cov}(X(f_1), X(f_2))$$

$$= (\nu_4 - 3)\Psi_1(f_1)\Psi_1(f_2) + 2 \sum_{k=1}^{\infty} k\Psi_k(f_1)\Psi_k(f_2),\tag{3.5}$$

$$= \frac{1}{4\pi^2} \int_{-2}^2 \int_{-2}^2 f'_1(x) f'_2(y) H(x, y) dx dy,\tag{3.6}$$

where

$$H(x, y) = (\nu_4 - 3)\sqrt{4 - x^2}\sqrt{4 - y^2} + 2 \log \left(\frac{4 - xy + \sqrt{(4 - x^2)(4 - y^2)}}{4 - xy - \sqrt{(4 - x^2)(4 - y^2)}} \right).$$

Remark 3. *Bai and Yin's result (1998) suggests that for large p and n with $p/n \rightarrow \infty$, the matrix $\sqrt{n}\mathbf{B}$ is close to a $n \times n$ Wigner matrix although its entries are not independent but weakly dependent. It is then reasonable to conjecture the CLT for LSS of $\mathbf{B}$ from that of Wigner matrix in Bai and Yao (2005). More precisely, write the matrix $\mathbf{B}$ as $\mathbf{B} = \frac{1}{\sqrt{n}}(w_{ij})$, where $w_{ii} = (\mathbf{s}_i^T \mathbf{s}_i - p)/\sqrt{p}$, $w_{ij} = \mathbf{s}_i^T \mathbf{s}_j/\sqrt{p}$ for $i \neq j$ and $\mathbf{s}_j$ is the j -th column of $\mathbf{X}$. Then one has*

$$\text{Var}(w_{11}) = \nu_4 - 1, \quad \text{Var}(w_{12}) = 1, \quad E(w_{12}^2 - 1)^2 = \frac{1}{p}(\nu_4^2 - 3).$$

By the definition of $\Psi_k(f)$, we can rewrite (3.4) as

$$EX(f) = \frac{1}{4}(f(2) + f(-2)) - \frac{1}{2}\Psi_0(f) + (\nu_4 - 3)\Psi_2(f).$$

Then (3.4), (3.5) and (3.6) are consistent with (1.4), (1.5) and (1.6) in Bai and Yao (2005), respectively, if we take the values of $\sigma^2 = \nu_4 - 1$, $\kappa = 2$ (real variables case) and $\beta = 0$. However, we remark that for general case, the mean correction term of $Q_n(f)$ in (3.2) can not be guessed from Bai and Yao (2005)'s result.

Remark 4. *If we interchange the roles of p and n , Birke and Dette (2005) established the CLT for $Q_n(f)$ when $f = x^2$ and $X_{ij} \sim N(0, 1)$. To see how Theorem 3 recovers their result, first, the fact that $f = x^2$ is even implies that $\Psi_3(f) = 0$. There fore the means in Theorem 3.4 of Birke and Dette (2005) are both equal to one. Second, the variance in Theorem 3.4 of Birke and Dette (2005) is also consistent with (3.5) in Theorem 3. Indeed, on the*

Chapter 3. Central limit theorem for the linear spectral statistics

one hand, $\text{cov}(sV)$ of Birke and Dette (2005) equals 4 when $y = 0$. On the other hand $\nu_4 = 3$ in (3.5) if $f = x^2, X_{ij} \sim N(0, 1)$. Also, for $k \geq 3$,

$$\begin{aligned}\Psi_k(f) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} 4 \cos^2 \theta \cos k\theta d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} 2(\cos 2\theta + 1) \cos k\theta d\theta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (\cos(k-2)\theta + \cos(k+2)\theta + 2\cos k\theta) d\theta = 0.\end{aligned}$$

Meanwhile,

$$\begin{aligned}\Psi_1(f) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} 4 \cos^3 \theta d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\cos 3\theta + 3\cos \theta) d\theta = 0. \\ \Psi_2(f) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} 4 \cos^2 \theta \cos 2\theta d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\cos 4\theta + 1 + 2\cos 2\theta) d\theta = 1.\end{aligned}$$

Therefore $\text{cov}(X(x^2), X(x^2)) = 4$, which equals $\text{cov}(sV)$.

It is noted that Theorem 1.1 is established under the restriction $n^3/p = O(1)$. The next theorem release this condition to the general framework $n/p \rightarrow 0$. For this purpose, we renew $Q_n(f)$ as

$$\begin{aligned}G_n(f) &\triangleq n \int_{-\infty}^{+\infty} f(x) d(F^{\mathbf{B}}(x) - F_{sc}(x)) \\ &\quad - \frac{n}{2\pi i} \oint_{|m|=\rho} f(-m - m^{-1}) \mathcal{X}_n(m) \frac{1-m^2}{m^2} dm,\end{aligned}\tag{3.7}$$

where

$$\begin{aligned}\mathcal{X}_n(m) &\triangleq \frac{-\mathcal{B} + \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}}}{2\mathcal{A}}, \\ \mathcal{A} &= m - \sqrt{\frac{n}{p}}(1 + m^2), \\ \mathcal{B} &= m^2 - 1 - \sqrt{\frac{n}{p}}m(1 + 2m^2), \\ \mathcal{C} &= \frac{m^3}{n} \left(\frac{m^2}{1-m^2} - \nu_4 - 2 \right) - \sqrt{\frac{n}{p}}m^4,\end{aligned}\tag{3.8}$$

and $\sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}}$ is the complex number whose imaginary has the same sign as the imaginary part of $\mathcal{B}$. The integral's contour is $|m| = \rho$ with $\rho < 1$.

Theorem 4. *Suppose that*

(a) $\mathbf{X} = (X_{ij})_{p \times n}$ where $\{X_{ij} : i = 1, 2, \dots, p; j = 1, 2, \dots, n\}$ are i.i.d. real random variables with $EX_{11} = 0, EX_{11}^2 = 1$ and $\nu_4 = EX_{11}^4 < \infty$.

(b2) $n/p \rightarrow 0$ as $n \rightarrow \infty$.

Then, for any $f_1, \dots, f_k \in \mathcal{M}$, the finite dimensional random vector $(G_n(f_1), \dots, G_n(f_k))$ converges weakly to a Gaussian vector $(Y(f_1), \dots, Y(f_k))$ with mean function $EY(f) = 0$ and covariance function $\text{cov}(Y(f), Y(g))$ the same as the right side of (3.5) and (3.6).

Remark 5. *It is observed from Theorem 3 and 4 that the two linear functionals of the eigenvalues of $\mathbf{B}$, $Q_n(f)$ under assumption $n^3/p = O(1)$ and $G_n(f)$ under assumption $n/p \rightarrow 0$, have the same asymptotic covariances, but different mean correction terms. The proofs for the two theorems are almost the same except the differences in identifying the mean correction terms (which is showed in Proposition 3).*

Remark 6. *If $n^3/p = O(1)$, Theorem 4 is consistent with theorem 3. Indeed, since $n^3/p = O(1)$, we have $4\mathcal{AC} = o(1)$, $\mathcal{B} = m^2 - 1$. By (3.8),*

$$\begin{aligned} n\mathcal{X}_n(m) &= n \cdot \frac{-\mathcal{B} + \sqrt{\mathcal{B}^2 - 4\mathcal{AC}}}{2\mathcal{A}} = \frac{-2nC}{\mathcal{B} + \sqrt{\mathcal{B}^2 - 4\mathcal{AC}}} \\ &= \frac{m^3}{1 - m^2} \left(\frac{m^2}{1 - m^2} - \nu_4 - 2 \right) + \sqrt{\frac{n^3}{p}} \frac{m^4}{1 - m^2} + o(1). \end{aligned}$$

Hence, by the same calculation in Section 5.1 in Bai and Yao (2005), we

Chapter 3. Central limit theorem for the linear spectral statistics

have

$$\begin{aligned}
& -\frac{n}{2\pi i} \oint_{|m|=\rho} f(-m-m^{-1}) \mathcal{X}_n(m) \frac{1-m^2}{m^2} dm \\
& = -\frac{1}{2\pi i} \oint_{|m|=\rho} f(-m-m^{-1}) m \left[\frac{m^2}{1-m^2} \right. \\
& \quad \left. - \nu_4 - 2 + \sqrt{\frac{n^3}{p}} m \right] dm + o(1) \\
& = -\left[\frac{1}{4}(f(2) + f(-2)) - \frac{1}{2}\Psi_0(f) + (\nu_4 - 3)\Psi_2(f) \right] \\
& \quad - \sqrt{\frac{n^3}{p}} \Psi_3(f) + o(1),
\end{aligned} \tag{3.9}$$

which is consistent with the result in Theorem 3.

Remark 7. Note that $\mathcal{X}_n(m)$ and $\underline{\mathcal{X}}_n(m) \triangleq \frac{-\mathcal{B} - \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}}}{2\mathcal{A}}$ are the two roots of the equation $\mathcal{A}x^2 + \mathcal{B}x + \mathcal{C} = 0$. Since $n/p \rightarrow 0$, a easy calculation shows $X_n(m) = o(1)$ and $\underline{X}_n(m) = \frac{1-m^2}{m} + o(1)$. Hence, for those who may potentially utilize this result, one may implement the mean correction in (3.7) as

$$\mathcal{X}_n(m) = \min \left\{ \frac{-\mathcal{B} + \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}}}{2\mathcal{A}}, \frac{-\mathcal{B} - \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}}}{2\mathcal{A}} \right\},$$

where $\rho = 0.4$ (say), $m = \rho e^{i\theta}$, $\theta \in [-2\pi, 2\pi]$ and $\min(\cdot, \cdot)$ selects the complex number with smaller absolute value.

3.2 Truncation and strategy for the proofs of Theorem 3 and Theorem 4

As illustrated in Remark 5, the proofs for Theorem 3 and Theorem 4 only differs in deriving the mean correction terms which is shown in Proposition 3. The other proofs are exactly the same except the notations differences. So we only show the details of the proof of Theorem 3 in the following.

3.2.1 Truncation

In this section, we will truncate the underlying random variables as in Pan and Gao (2012). Choose δ_n satisfying

$$\lim_{n \rightarrow \infty} \delta_n^{-4} E|X_{11}|^4 I(|X_{11}| > \delta_n \sqrt[4]{np}) = 0, \quad \delta_n \downarrow 0, \quad \delta_n \sqrt[4]{np} \uparrow \infty. \quad (3.10)$$

In what follows, we will use δ to represent δ_n for convenience. We first truncate the variables $\hat{X}_{ij} = X_{ij} I(|X_{ij}| < \delta \sqrt[4]{np})$ and then normalize it as $\tilde{X}_{ij} = (\hat{X}_{ij} - E\hat{X}_{ij})/\sigma$, where σ is the standard deviation of $\hat{X}_{ij}$. Let $\hat{\mathbf{X}} = (\hat{X}_{ij})$ and $\tilde{\mathbf{X}} = (\tilde{X}_{ij})$. Define $\hat{\mathbf{B}}, \tilde{\mathbf{B}}$ and $\hat{Q}_n(f), \tilde{Q}_n(f)$ similarly by means of (3.1) and (3.2), respectively. We then have

$$\begin{aligned} P(\mathbf{B} \neq \hat{\mathbf{B}}) &\leq np P(|X_{11}| \geq \delta_n \sqrt[4]{np}) \\ &\leq K \delta_n^{-4} E|X_{11}|^4 I(|X_{11}| > \delta_n \sqrt[4]{np}) \\ &= o(1). \end{aligned}$$

It follows from (3.10) that

$$\begin{aligned} |1 - \sigma^2| &\leq 2 |EX_{11}^2 I(|X_{11}| > \delta \sqrt[4]{np})| \\ &\leq 2(np)^{-1/2} \delta^{-2} E|X_{11}|^4 I(|X_{11}| > \delta \sqrt[4]{np}) \\ &= o((np)^{-1/2}), \end{aligned}$$

and

$$|E\hat{X}_{11}| \leq \delta^{-3} (np)^{-3/4} E|X_{11}|^4 I(|X_{11}| > \delta \sqrt[4]{np}) = o((np)^{-3/4}).$$

Therefore

$$\begin{aligned} Etr(\tilde{\mathbf{X}} - \hat{\mathbf{X}})^T (\tilde{\mathbf{X}} - \hat{\mathbf{X}}) &\leq \sum_{i,j} E|\hat{X}_{ij} - \tilde{X}_{ij}|^2 \\ &\leq Kpn \left(\frac{(1 - \sigma)^2}{\sigma^2} E|\hat{X}_{11}|^2 + \frac{1}{\sigma^2} |E\hat{X}_{11}|^2 \right) = o(1), \end{aligned}$$

Chapter 3. Central limit theorem for the linear spectral statistics

and

$$Etr\hat{\mathbf{X}}^T\hat{\mathbf{X}} \leq \sum_{i,j} E|\hat{X}_{ij}|^2 \leq Knp, \quad Etr\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T \leq \sum_{i,j} E|\tilde{X}_{ij}|^2 \leq Knp.$$

Recalling that the notation $\lambda_j(\cdot)$ represents the j -th largest eigenvalue, we then have $\lambda_j(\mathbf{X}^T\mathbf{X}) = \sqrt{np}\lambda_j(\mathbf{B}) + p$. Similar equalities also hold if $\mathbf{X}, \mathbf{B}$ are replaced by $\hat{\mathbf{X}}, \hat{\mathbf{B}}$ or $\tilde{\mathbf{X}}, \tilde{\mathbf{B}}$. Consequently, applying the argument used in Theorem 11.36 in Bai and Silverstein (2006) and Cauchy-Schwarz's inequality, we have

$$\begin{aligned} & E \left| \tilde{Q}_n(f) - \hat{Q}_n(f) \right| \\ & \leq n \sum_{j=1}^n E |f(\lambda_j(\hat{\mathbf{B}})) - f(\lambda_j(\tilde{\mathbf{B}}))| \\ & \leq K_f n \sum_{j=1}^n E |\lambda_j(\hat{\mathbf{B}}) - \lambda_j(\tilde{\mathbf{B}})| \leq \frac{K_f}{\sqrt{np}} \sum_{j=1}^n E |\lambda_j(\hat{\mathbf{X}}^T\hat{\mathbf{X}}) - \lambda_j(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})| \\ & \leq \frac{K_f}{\sqrt{np}} E \left[tr(\tilde{\mathbf{X}} - \hat{\mathbf{X}})^T(\tilde{\mathbf{X}} - \hat{\mathbf{X}}) \cdot 2 \left(tr\hat{\mathbf{X}}^T\hat{\mathbf{X}} + tr\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} \right) \right]^{1/2} \\ & \leq \frac{2K_f}{\sqrt{np}} \left[Etr(\tilde{\mathbf{X}} - \hat{\mathbf{X}})^T(\tilde{\mathbf{X}} - \hat{\mathbf{X}}) \cdot \left(Etr\hat{\mathbf{X}}^T\hat{\mathbf{X}} + Etr\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} \right) \right]^{\frac{1}{2}} \\ & = o(1). \end{aligned}$$

where K_f is a bound on $|f'(x)|$. Thus the weak convergence of $Q_n(f)$ is not affected if we replace the original variables X_{ij} by the truncated and normalized variables $\tilde{X}_{ij}$. For convenience, we still use X_{ij} to denote $\tilde{X}_{ij}$, which satisfies the following additional assumption (c):

(c) The underlying variables satisfy

$$|X_{ij}| \leq \delta \sqrt[4]{np}, \quad EX_{ij} = 0, \quad EX_{ij}^2 = 1, \quad EX_{ij}^4 = \nu_4 + o(1),$$

where $\delta = \delta_n$ satisfies $\lim_{n \rightarrow \infty} \delta_n^{-4} E|X_{11}|^4 I(|X_{11}| > \delta_n \sqrt[4]{np}) = 0$, $\delta_n \downarrow 0$, and $\delta_n \sqrt[4]{np} \uparrow \infty$.

For any $\epsilon > 0$, define the event $F_n(\epsilon) = \{\max_{j \leq n} |\lambda_j(\mathbf{B})| \geq 2 + \epsilon\}$ where $\mathbf{B}$ is defined by the truncated and normalized variables satisfying Assumption (c). By Theorem 2 in Chen and Pan (2012), for any $\ell > 0$

$$P(F_n(\epsilon)) = o(n^{-\ell}). \quad (3.11)$$

Here we would point out that the result regarding the minimum eigenvalue of $\mathbf{B}$ can be obtained similarly by investigating the maximum eigenvalue of $-\mathbf{B}$.

3.2.2 Strategy of the proof

We shall follow the strategy of Bai and Yao (2005). Specifically speaking, assume that u_0, v are fixed and sufficiently small so that $\varsigma \subset \mathcal{S}$ (see the definition in the introduction). Let ς be the contour formed by the boundary of the rectangle with $(\pm u_0 \pm iv)$ where $u_0 > 2, 0 < v \leq 1$. By Cauchy's integral formula, with probability one,

$$Q_n(f) = -\frac{1}{2\pi i} \oint_{\varsigma} f(z) n \left[m_n(z) - m(z) + \sqrt{\frac{n}{p}} \frac{m^4(z)}{1 - m^2(z)} \right] dz,$$

where $m_n(z), m(z)$ denote the Stieltjes transform of $F^{\mathbf{B}}(x)$ and $F(x)$, respectively, and we also use the fact that

$$\begin{aligned} & -\sqrt{\frac{n^3}{p}} \frac{1}{2\pi i} \oint_{\varsigma} f(z) \frac{m^4(z)}{1 - m^2(z)} dz \\ &= \sqrt{\frac{n^3}{p}} \frac{1}{2\pi i} \oint_{|m|=\rho} f(-m - m^{-1}) \sqrt{\frac{n}{p}} m^2 dm = -\sqrt{\frac{n^3}{p}} \Psi_3(f). \end{aligned}$$

which can be obtained by the calculations of Section 5.1 in Bai and Yao (2005).

Let

$$M_n(z) = n \left[m_n(z) - m(z) + \sqrt{\frac{n}{p}} \frac{m^4(z)}{1 - m^2(z)} \right], \quad z \in \varsigma.$$

Chapter 3. Central limit theorem for the linear spectral statistics

For $z \in \varsigma$, write $M_n(z) = M_n^{(1)}(z) + M_n^{(2)}(z)$ where

$$M_n^{(1)}(z) = n[m_n(z) - Em_n(z)], \quad M_n^{(2)}(z) = n \left[Em_n(z) - m(z) + \sqrt{\frac{n}{p}} \frac{m^4(z)}{1 - m^2(z)} \right].$$

Split the contour ς as the union of $\varsigma_u, \varsigma_l, \varsigma_r, \varsigma_0$ where $\varsigma_l = \{z = -u_0 + iv, \xi_n n^{-1} < |v| < v_1\}$, $\varsigma_r = \{z = u_0 + iv, \xi_n n^{-1} < |v| < v_1\}$, $\varsigma_0 = \{z = \pm u_0 + iv, |v| \leq \xi_n n^{-1}\}$ and $\varsigma_u = \{z = u \pm iv_1, |u| \leq u_0\}$ and where ξ_n is a slowly varying sequence of positive constants and v_1 is a positive constant which is independent of n . Throughout this paper, let $\mathbb{C}_1 = \{z : z = u + iv, u \in [-u_0, u_0], |v| \geq v_1\}$.

Proposition 1. *Under Assumptions (b), (c), the empirical process $\{M_n(z), z \in \mathbb{C}_1\}$ converges weakly to a Gaussian process $\{M(z), z \in \mathbb{C}_1\}$ with the mean function*

$$\Delta(z) = m^3(z) \left(m'(z) + \nu_4 - 2 \right) \left(1 + m'(z) \right) \quad (3.12)$$

and the covariance function

$$\Lambda(z_1, z_2) = m'(z_1) m'(z_2) \left[\nu_4 - 3 + 2 \left(1 - m(z_1) m(z_2) \right)^{-2} \right]. \quad (3.13)$$

As in Bai and Yao (2005), the process of $\{M(z), z \in \mathbb{C}_1\}$ can be extended to $\{M(z), \Re(z) \notin [-2, 2]\}$ due to the facts that i) $M(z)$ is symmetric, e.g. $M(\bar{z}) = \overline{M(z)}$; ii) the mean and the covariance function of $M(z)$ are independent of v_1 and they are continuous except for $\Re(z) \notin [-2, 2]$. By Proposition 1 and the continuous mapping theorem,

$$-\frac{1}{2\pi i} \int_{\varsigma_u} f(z) M_n(z) dz \xrightarrow{d} -\frac{1}{2\pi i} \int_{\varsigma_u} f(z) M(z) dz.$$

Thus, to prove Theorem 3, it is also necessary to prove the following proposition.

Proposition 2. *Let $z \in \mathbb{C}_1$. Under Assumptions (b), (c), there exists some event U_n with $P(U_n) \rightarrow 0$, as $n \rightarrow \infty$, such that*

$$\lim_{v_1 \downarrow 0} \limsup_{n \rightarrow \infty} E \left| \int_{\bigcup_{i=l,r,0} \varsigma_i} M_n^{(1)}(z) I(U_n^c) dz \right|^2 = 0, \quad (3.14)$$

$$\lim_{v_1 \downarrow 0} \limsup_{n \rightarrow \infty} \left| \int_{\bigcup_{i=l,r,0} \varsigma_i} E M_n(z) I(U_n^c) dz \right| = 0, \quad (3.15)$$

and

$$\lim_{v_1 \downarrow 0} E \left| \int_{\varsigma_i} M^{(1)}(z) dz \right|^2 = 0, \quad \lim_{v_1 \downarrow 0} E \left| \int_{\varsigma_i} M(z) dz \right|^2 = 0. \quad (3.16)$$

Since $E|M^{(1)}(z)|^2 = \Lambda(z, \bar{z})$ and $E|M(z)|^2 = \Lambda(z, \bar{z}) + |EM(z)|^2$, (3.16) can be easily obtained from Proposition 1. For $i = 0$, if we choose $U_n = F_n(\epsilon)$ with the $\epsilon = (u_0 - 2)/2$, then when U_n^c happens, $\forall z \in \varsigma_0$, we have $|m_n(z)| \leq 2/(u_0 - 2)$ and $|m(z)| \leq 1/(u_0 - 2)$. Thus

$$\left| \int_{\varsigma_0} M_n^{(1)}(z) I(U_n^c) dz \right| \leq n \left(\frac{4}{u_0 - 2} \right)^2 \|\varsigma_0\| \leq \frac{4\xi_n}{(u_0 - 2)^2},$$

where $\|\varsigma_0\|$ represents the length of ς_0 . Furthermore,

$$\left| \int_{\varsigma_0} M_n(z) I(U_n^c) dz \right| \leq n \left(\frac{2}{u_0 - 2} + \frac{1}{u_0 - 2} + K \frac{n}{p} \right)^2 \|\varsigma_0\|,$$

These imply that (3.15) and (3.14) are true for $z \in \varsigma_0$ by noting that $\xi_n \rightarrow 0$ as $p \rightarrow \infty$. The proof of (3.14) for $i = l, r$ are given in Section 3.5.

Section 3.3 and 3.4 are devoted to the proof of Proposition 1. The main steps are encapsulated in the following:

- According to Theorem 8.1 in Billingsley (1968), to establish the convergence of the process $\{M_n(z), z \in \mathbb{C}_1\}$, it suffices to prove the finite dimensional convergence of the random part $M_n^{(1)}(z)$ and its tightness, and the convergence of the non-random part $M_n^{(2)}(z)$ converges to zero.

Chapter 3. Central limit theorem for the linear spectral statistics

- For $M_n^{(1)}(z)$, we first write it into the martingale expression, namely $M_n^{(1)}(z) = \sum_{k=1}^n E_k(\alpha_k(z)) + o_{L_1}(1)$. Applying the central limit theorem of martingales, we next show that the covariance of $M_n^{(1)}(z)$ is

$$\begin{aligned}\Lambda_n(z_1, z_2) &\triangleq \sum_{k=1}^n E_{k-1} [E_k \alpha_k(z_1) \cdot E_k \alpha_k(z_2)] \\ &= \frac{\partial^2}{\partial z_2 \partial z_1} \left[m(z_1) m(z_2) \tilde{\Lambda}_n(z_1, z_2) + o_{i.p.}(1) \right]\end{aligned}$$

(see the definitions of the notation in Section 3.3.1 and 3.3.2). By Vitali's theorem, we then convert the problem to the limit of $\tilde{\Lambda}_n(z_1, z_2)$.

- We next show that $\tilde{\Lambda}_n(z_1, z_2) = \frac{2}{n} \sum_{k=1}^n \mathbb{Z}_k + \nu_4 - 1 + o_{L_1}(1)$ and derive the asymptotic expression of $\mathbb{Z}_k$. Then the limit of covariance $\Lambda_n(z_1, z_2)$ will be obtained.
- For $M_n^{(2)}(z)$, a expansion of $Em_n(z) - m(z)$ is obtained in (3.59) and then discuss this equation under assumption $n^3/p = O(1)$ and $n/p \rightarrow 0$ respectively.

3.3 Convergence of $M_n^{(1)}(z)$

To prove Proposition 1, we need to establish i) the finite-dimensional convergence and the tightness of $M_n^{(1)}(z)$; ii) the convergence of the mean function $EM(z)$. This section is devoted to the first target. Throughout this section we assume $z \in \mathbb{C}_1$. and K denotes a constant which may change from line to line and may depend on v_1 but be independent of n .

3.3.1 Simplification of $M_n^{(1)}(z)$

The aim of this subsection is to simplify $M_n^{(1)}(z)$ so that $M_n^{(1)}(z)$ can be written in the form of martingales. Some moment bounds are also proved.

Define $\mathbf{D} = \mathbf{B} - z\mathbf{I}_n$. Let $\mathbf{s}_k$ be the k -th column of $\mathbf{X}$ and $\mathbf{X}_k$ be a $p \times (n-1)$ matrix constructed from $\mathbf{X}$ by deleting the k -th column. We then similarly define $\mathbf{B}_k = \frac{1}{\sqrt{np}}(\mathbf{X}_k^T \mathbf{X}_k - p\mathbf{I}_{n-1})$ and $\mathbf{D}_k = \mathbf{B}_k - z\mathbf{I}_{n-1}$. The k -th diagonal element of $\mathbf{D}$ is $a_{kk}^{diag} = \frac{1}{\sqrt{np}}(\mathbf{s}_k^T \mathbf{s}_k - p) - z$ and the k -th row of $\mathbf{D}$ with the k -th element deleted is $\mathbf{q}_k^T = \frac{1}{\sqrt{np}}\mathbf{s}_k^T \mathbf{X}_k$. The Stieltjes transform of $F^{\mathbf{B}}$ has the form $m_n(z) = \frac{1}{n}tr\mathbf{D}^{-1}$. The limiting Stieltjes transform $m(z)$ satisfies

$$m(z) = -\frac{1}{z + m(z)}, \quad |m(z)| \leq 1 \quad (3.17)$$

(one may see Bai and Yao (2005)).

Define the σ -field $\mathcal{F}_k = \sigma(s_1, s_2, \dots, s_k)$ and the conditional expectation $E_k(\cdot) = E(\cdot|\mathcal{F}_k)$. By the matrix inversion formula we have (see (3.9) of Bai 1993)

$$tr(\mathbf{D}^{-1} - \mathbf{D}_k^{-1}) = -\frac{(1 + \mathbf{q}_k^T \mathbf{D}_k^{-2} \mathbf{q}_k)}{-a_{kk}^{diag} + \mathbf{q}_k^T \mathbf{D}_k^{-1} \mathbf{q}_k}. \quad (3.18)$$

We then obtain

$$\begin{aligned} M_n^{(1)}(z) &= tr\mathbf{D}^{-1} - Etr\mathbf{D}^{-1} \\ &= \sum_{k=1}^n (E_k - E_{k-1})tr(\mathbf{D}^{-1} - \mathbf{D}_k^{-1}) = \sum_{k=1}^n \varrho_k \\ &= (E_k - E_{k-1})\iota_k - E_k\kappa_k, \end{aligned} \quad (3.19)$$

Chapter 3. Central limit theorem for the linear spectral statistics

where

$$\begin{aligned}
\varrho_k &= -(E_k - E_{k-1})\beta_k (1 + \mathbf{q}_k^T \mathbf{D}_k^{-2} \mathbf{q}_k) \\
\iota_k &= -\beta_k^{tr} \beta_k \eta_k (1 + \mathbf{q}_k^T \mathbf{D}_k^{-2} \mathbf{q}_k), \\
\eta_k &= \frac{1}{\sqrt{np}} (\mathbf{s}_k^T \mathbf{s}_k - p) - \gamma_{k1}, \quad \beta_k = \frac{1}{-a_{kk}^{diag} + \mathbf{q}_k^T \mathbf{D}_k^{-1} \mathbf{q}_k}, \\
\beta_k^{tr} &= \frac{1}{z + \frac{1}{np} tr \mathbf{M}_k^{(1)}}, \quad \mathbf{M}_k^{(s)} = \mathbf{X}_k \mathbf{D}_k^{-s} \mathbf{X}_k^T, s = 1, 2, \\
\gamma_{ks} &= \mathbf{q}_k^T \mathbf{D}_k^{-s} \mathbf{q}_k - (np)^{-1} tr \mathbf{M}_k^{(s)}, \quad \kappa_k = \beta_k^{tr} \gamma_{k2}.
\end{aligned}$$

In the above equality, ϱ_k is obtained by (3.18) and the last equality uses the facts that

$$\beta_k = \beta_k^{tr} + \beta_k \beta_k^{tr} \eta_k, \quad (3.20)$$

and

$$(E_k - E_{k-1}) \left[\beta_k^{tr} \left(1 + \frac{1}{np} tr \mathbf{M}_k^{(2)} \right) \right] = 0, \quad E_{k-1} \kappa_k = 0.$$

We remind the readers that the variable z has been dropped from the expressions such as $\mathbf{D}^{-1}$, $\mathbf{D}_k^{-1}$, β_k , γ_{ks} and so on. When necessary, we will also indicate them as $\mathbf{D}^{-1}(z)$, $\mathbf{D}_k^{-1}(z)$, $\beta_k(z)$, $\gamma_{ks}(z)$, etc..

We next provide some useful bounds. It follows from the definitions of $\mathbf{D}$ and $\mathbf{D}_k$ that

$$\begin{aligned}
\mathbf{D}^{-1} \mathbf{X}^T \mathbf{X} &= p \mathbf{D}^{-1} + \sqrt{np} (\mathbf{I}_n + z \mathbf{D}^{-1}), \\
\mathbf{D}_k^{-1} \mathbf{X}_k^T \mathbf{X}_k &= p \mathbf{D}_k^{-1} + \sqrt{np} (\mathbf{I}_{n-1} + z \mathbf{D}_k^{-1}).
\end{aligned} \quad (3.21)$$

Since the eigenvalues of $\mathbf{D}^{-1}$ have the form $1/(\lambda_j(\mathbf{B}) - z)$, $\|\mathbf{D}^{-1}\| \leq 1/v_1$ and similarly $\|\mathbf{D}_k^{-1}\| \leq 1/v_1$. From Theorem 11.4 in Bai and Silverstein (2006), we note that $-\beta_k(z)$ is the k -th diagonal element of $\mathbf{D}^{-1}$ so that $|\beta_k| \leq 1/v_1$. Moreover, considering the imaginary parts of $1/\beta_k^{tr}$ and $1/\beta_k$ and by (3.21) we have

$$|\beta_k^{tr}| \leq 1/v_1, \quad |1 + \frac{1}{np} tr \mathbf{M}_k^{(s)}| \leq (1 + 1/v_1^{2s}), s = 1, 2, \quad (3.22)$$

and

$$\left| \left(1 + \mathbf{q}_k^T \mathbf{D}_k^{-2} \mathbf{q}_k\right) \beta_k \right| \leq \frac{1 + \mathbf{q}_k^T \mathbf{D}_k^{-1} \overline{\mathbf{D}_k}^{-1} \mathbf{q}_k}{v_1(1 + \mathbf{q}_k^T \mathbf{D}_k^{-1} \overline{\mathbf{D}_k}^{-1} \mathbf{q}_k)} = 1/v_1. \quad (3.23)$$

Applying (3.20), we split ι_k as

$$\begin{aligned} \iota_k &= - \left(1 + \frac{1}{np} \text{tr} \mathbf{M}_k^{(2)}\right) (\beta_k^{tr})^2 \eta_k - \gamma_{k1} (\beta_k^{tr})^2 \eta_k \\ &\quad - \left(1 + \frac{1}{np} \mathbf{q}_k^T \mathbf{D}_k^{-2} \mathbf{q}_k\right) (\beta_k^{tr})^2 \beta_k \eta_k^2 \\ &= \iota_{k1} + \iota_{k2} + \iota_{k3}. \end{aligned}$$

As will be seen, ι_{k1}, ι_{k2} could be negligible by Lemma 2 below.

By Lemma 2, (3.22) and (3.23), we have

$$\begin{aligned} E \left| \sum_{k=1}^n (E_k - E_{k-1}) \iota_{k3} \right|^2 &\leq \sum_{k=1}^n E \left| \left(1 + \frac{1}{np} \mathbf{s}_k^T \mathbf{M}_k^{(2)} \mathbf{s}_k\right) (\beta_k^{tr})^2 \beta_k \eta_k^2 \right|^2 \\ &\leq K \delta^4, \end{aligned}$$

and that

$$\begin{aligned} E \left| \sum_{k=1}^n (E_k - E_{k-1}) \iota_{k2} \right|^2 &\leq \sum_{k=1}^n E \left| \gamma_{k1} (\beta_k^{tr})^2 \eta_k \right|^2 \\ &\leq K \sum_{k=1}^n (E |\gamma_{k1}|^4 E |\eta_k|^4)^{1/2} \\ &\leq \frac{Kn}{p} + K \delta^2. \end{aligned}$$

Therefore $M_n^{(1)}(z)$ is simplified as

$$\begin{aligned} M_n^{(1)}(z) &= \sum_{k=1}^n E_k \left[- \left(1 + \frac{1}{np} \text{tr} \mathbf{M}_k^{(2)}\right) (\beta_k^{tr})^2 \eta_k - \kappa_k \right] + o_{L_2}(1) \\ &= \sum_{k=1}^n E_k (\alpha_k(z)) + o_{L_2}(1), \end{aligned} \quad (3.24)$$

Chapter 3. Central limit theorem for the linear spectral statistics

where $\alpha_k(z)$ represents the term in the square bracket. Thus, to prove finite dimensional convergence of $M_n^{(1)}(z)$, $z \in \mathbb{C}_1$ we need only consider the sum

$$\sum_{j=1}^l a_j \sum_{k=1}^n E_k(\alpha_k(z_j)) = \sum_{k=1}^n \sum_{j=1}^l a_j E_k(\alpha_k(z_j)), \quad (3.25)$$

where $a_1, \dots, a_l$ are complex numbers and l is any positive integer.

3.3.2 Application of central limit theorem of martingales

In order to establish the central limit theorem for the martingale (3.25) we have to check the following two conditions :

Condition 1. *Lyapunov condition.* For some $a > 2$,

$$\sum_{k=1}^n E_{k-1} \left[\left| E_k \left(\sum_{j=1}^l a_j E_k(\alpha_k(z_j)) \right) \right|^a \right] \xrightarrow{i.p.} 0.$$

Condition 2. *The covariance*

$$\Lambda_n(z_1, z_2) \triangleq \sum_{k=1}^n E_{k-1} [E_k \alpha_k(z_1) \cdot E_k \alpha_k(z_2)] \quad (3.26)$$

converges in probability to $\Lambda(z_1, z_2)$ whose explicit form will be given in (3.54).

Condition 1 is satisfied by choosing $a = 4$, using Lemma 2, and the fact that via (3.22)

$$|\alpha_k(z)| = \left| \left(1 + \frac{1}{np} \text{tr} \mathbf{M}_k^{(2)} \right) (\beta_k^{tr})^2 \eta_k + \beta_k^{tr} \gamma_k \right| \leq \frac{1 + v_1^{-2}}{v_1^2} |\eta_k| + \frac{1}{v_1} |\gamma_k|.$$

Consider Condition 2 now. Note that

$$\alpha_k(z) = - \left(1 + \frac{1}{np} \text{tr} \mathbf{M}_k^{(2)} \right) (\beta_k^{tr})^2 \eta_k - \gamma_k \beta_k^{tr} = \frac{\partial}{\partial z} \left(\beta_k^{tr} \eta_k \right).$$

By the dominated convergence theorem, we have

$$\begin{aligned} \Lambda_n(z_1, z_2) &= \frac{\partial^2}{\partial z_2 \partial z_1} \sum_{k=1}^n E_{k-1} \left[E_k \left(\beta_k^{tr}(z_1) \eta_k(z_1) \right) \right. \\ &\quad \left. \times E_k \left(\beta_k^{tr}(z_2) \eta_k(z_2) \right) \right]. \end{aligned} \quad (3.27)$$

By (3.21), (3.18), (3.23), (3.17) and the fact $m_n(z) \xrightarrow{a.s.} m(z)$, and the dominated convergence theorem again, for any fixed t ,

$$\begin{aligned} E \left| \frac{1}{np} \text{tr} \mathbf{M}_k^{(1)} - m(z) \right|^t &\rightarrow 0, \\ E |\beta_k^{tr}(z) + m(z)|^t &\rightarrow 0, \quad \text{as } n \rightarrow \infty. \end{aligned} \quad (3.28)$$

Substituting (3.28) into (3.27) yields

$$\begin{aligned} \Lambda_n(z_1, z_2) &= \frac{\partial^2}{\partial z_2 \partial z_1} \left[m(z_1) m(z_2) \sum_{k=1}^n E_{k-1} \left(E_k \eta_k(z_1) \cdot E_k \eta_k(z_2) \right) \right. \\ &\quad \left. + o_{i.p.}(1) \right] \\ &= \frac{\partial^2}{\partial z_2 \partial z_1} \left[m(z_1) m(z_2) \tilde{\Lambda}_n(z_1, z_2) + o_{i.p.}(1) \right] \end{aligned} \quad (3.29)$$

By Vitali's theorem (see Titchmarsh, 1939, p. 168), it is enough to find the limit of $\tilde{\Lambda}_n(z_1, z_2)$. To this end, with notation $E_k(\mathbf{M}_k^{(1)}(z)) = \left(a_{ij}(z) \right)_{n \times n}$, write

$$\begin{aligned} E_k \eta_k(z) &= \frac{1}{\sqrt{np}} \sum_{j=1}^p (X_{jk}^2 - 1) - \frac{1}{np} \sum_{i \neq j} X_{ik} X_{jk} a_{ij}(z) \\ &\quad - \frac{1}{np} \sum_{i=1}^p (X_{ik}^2 - 1) a_{ii}(z). \end{aligned}$$

By the above formula and independence between $\{X_{ik}\}_{i=1}^p$ and $E_k(\mathbf{M}_k^{(1)})$, a straightforward calculation yields

$$E_{k-1} [E_k \eta_k(z_1) \cdot E_k \eta_k(z_2)] = \frac{1}{n} E(X_{11}^2 - 1)^2 + A_1 + A_2 + A_3 + A_4, \quad (3.30)$$

Chapter 3. Central limit theorem for the linear spectral statistics

where

$$\begin{aligned} A_1 &= -\frac{1}{np\sqrt{np}}E(X_{11}^2 - 1)^2 \sum_{i=1}^p a_{ii}(z_1), \\ A_2 &= -\frac{1}{np\sqrt{np}}E(X_{11}^2 - 1)^2 \sum_{i=1}^p a_{ii}(z_2), \\ A_3 &= \frac{2}{n^2 p^2} \sum_{i \neq j}^p a_{ij}(z_1) a_{ij}(z_2), \\ A_4 &= \frac{1}{n^2 p^2} E(X_{11}^2 - 1)^2 \sum_{i=1}^p a_{ii}(z_1) a_{ii}(z_2). \end{aligned}$$

Note that $a_{ii}(z)$ is precisely $E_k a_{ii}^{(1)}$ in (3.89). From (3.89) we then obtain for $j = 1, 2, 4$

$$E \left| \sum_{k=1}^n A_j \right| \rightarrow 0.$$

Also, we conclude from (3.89) that

$$\sum_{k=1}^n A_3 = \frac{2}{n} \sum_{k=1}^n \mathbb{Z}_k - \frac{2}{n^2 p^2} \sum_{k=1}^n \sum_{i=1}^p a_{ii}(z_1) a_{ii}(z_2) = \frac{2}{n} \sum_{k=1}^n \mathbb{Z}_k + o_{L_1}(1),$$

where

$$\mathbb{Z}_k = \frac{1}{np^2} \text{tr} E_k \mathbf{M}_k^{(1)}(z_1) \cdot E_k \mathbf{M}_k^{(1)}(z_2).$$

Summarizing the above we see that

$$\tilde{\Lambda}_n(z_1, z_2) = \frac{2}{n} \sum_{k=1}^n \mathbb{Z}_k + \nu_4 - 1 + o_{L_1}(1). \quad (3.31)$$

3.3.3 The asymptotic expression of $\mathbb{Z}_k$

The goal is to derive an asymptotic expression of $\mathbb{Z}_k$ in the purpose to obtain the limit of $\tilde{\Lambda}_n(z_1, z_2)$.

3.3.3.1 Decomposition of $\mathbb{Z}_k$

To evaluate $\mathbb{Z}_k$, we need two different decompositions of $E_k \mathbf{M}_k^{(1)}(z)$. With slight abuse of notation, let $\{\mathbf{e}_i, i = 1, \dots, k-1, k+1, \dots, n\}$ be the $n-1$ dimensional unit vectors with the i -th (or $(i-1)$ -th) element equal to 1 and the remaining equal to 0 according as $i < k$ (or $i > k$). Write $\mathbf{X}_k = \mathbf{X}_{ki} + \mathbf{s}_i \mathbf{e}_i^T$. Define

$$\begin{aligned} \mathbf{D}_{ki,r} &= \mathbf{D}_k - \mathbf{e}_i \mathbf{h}_i^T = \frac{1}{\sqrt{np}} (\mathbf{X}_{ki}^T \mathbf{X}_k - p \mathbf{I}_{(i)}) - z \mathbf{I}_{n-1}, \\ \mathbf{D}_{ki} &= \mathbf{D}_k - \mathbf{e}_i \mathbf{h}_i^T - \mathbf{r}_i \mathbf{e}_i^T = \frac{1}{\sqrt{np}} (\mathbf{X}_{ki}^T \mathbf{X}_{ki} - p \mathbf{I}_{(i)}) - z \mathbf{I}_{n-1}, \\ \mathbf{h}_i^T &= \frac{1}{\sqrt{np}} \mathbf{s}_i^T \mathbf{X}_{ki} + \frac{1}{\sqrt{np}} (\mathbf{s}_i^T \mathbf{s}_i - p) \mathbf{e}_i^T, \quad \mathbf{r}_i = \frac{1}{\sqrt{np}} \mathbf{X}_{ki}^T \mathbf{s}_i \\ \zeta_i &= \frac{1}{1 + \vartheta_i}, \quad \vartheta_i = \mathbf{h}_i^T \mathbf{D}_{ki,r}^{-1}(z) \mathbf{e}_i, \quad \mathbf{M}_{ki} = \mathbf{X}_{ki} \mathbf{D}_{ki}^{-1}(z) \mathbf{X}_{ki}^T. \end{aligned} \tag{3.32}$$

Here $\mathbf{I}_{(i)}$ is obtained from $\mathbf{I}_{n-1}$ with the i -th (or $(i-1)$ -th) diagonal element replaced by zero if $i < k$ (or $i > k$). With respect to the above notations we would point out that, for $i < k$ (or $i > k$), the matrix $\mathbf{X}_{ki}$ is obtained from $\mathbf{X}_k$ with the entries on the i -th (or $(i-1)$ -th) column replaced by zero; $\mathbf{h}_i^T$ is the i -th (or $(i-1)$ -th) row of $\mathbf{B}_k$ and $\mathbf{r}_i$ is the i -th (or $(i-1)$ -th) column of $\mathbf{B}_k$ with the i -th (or $(i-1)$ -th) element replaced by zero. $(\mathbf{X}_{ki}^T \mathbf{X}_k - p \mathbf{I}_{(i)})$ is obtained from $(\mathbf{X}_k^T \mathbf{X}_k - p \mathbf{I}_{n-1})$ with the entries on the i -th (or $(i-1)$ -th) row and i -th (or $(i-1)$ -th) column replaced by zero.

The notation defined above may depend on k . When we obtain bounds or limits for them such as $\frac{1}{n} \text{tr} \mathbf{D}_{ki}^{-1}$ the results hold uniformly in k .

Observing the structure of the matrices $\mathbf{X}_{ki}$ and $\mathbf{D}_{ki}^{-1}$, we have some crucial identities,

$$\mathbf{X}_{ki} \mathbf{e}_i = \mathbf{0}, \quad \mathbf{e}_i^T \mathbf{D}_{ki,r}^{-1} = \mathbf{e}_i^T \mathbf{D}_{ki}^{-1} = -z^{-1} \mathbf{e}_i, \tag{3.33}$$

where $\mathbf{0}$ is a p dimensional vector with all the elements equal to 0. By (3.33)

Chapter 3. Central limit theorem for the linear spectral statistics

and the frequently used formulas

$$\begin{aligned} \mathbf{Y}^{-1} - \mathbf{W}^{-1} &= -\mathbf{W}^{-1}(\mathbf{Y} - \mathbf{W})\mathbf{Y}^{-1}, \quad (\mathbf{Y} + \mathbf{a}\mathbf{b}^T)^{-1}\mathbf{a} = \frac{\mathbf{Y}^{-1}\mathbf{a}}{1 + \mathbf{b}^T\mathbf{Y}^{-1}\mathbf{a}}, \\ \mathbf{b}^T(\mathbf{Y} + \mathbf{a}\mathbf{b}^T)^{-1} &= \frac{\mathbf{b}^T\mathbf{Y}^{-1}}{1 + \mathbf{b}^T\mathbf{Y}^{-1}\mathbf{a}}, \end{aligned} \quad (3.34)$$

we have

$$\begin{aligned} \mathbf{D}_k^{-1} - \mathbf{D}_{ki,r}^{-1} &= -\zeta_i \mathbf{D}_{ki,r}^{-1} \mathbf{e}_i \mathbf{h}_i^T \mathbf{D}_{ki,r}^{-1} \\ \mathbf{D}_{ki,r}^{-1} - \mathbf{D}_{ki}^{-1} &= \frac{1}{z\sqrt{np}} \mathbf{D}_{ki}^{-1} \mathbf{X}_{ki}^T \mathbf{s}_i \mathbf{e}_i^T. \end{aligned} \quad (3.35)$$

We first claim the following decomposition of $E_k \mathbf{M}_k^{(1)}(z)$, for $i < k$,

$$\begin{aligned} E_k \mathbf{M}_k^{(1)}(z) &= E_k \mathbf{M}_{ki} - E_k \left(\frac{\zeta_i}{znp} \mathbf{M}_{ki} \mathbf{s}_i \mathbf{s}_i^T \mathbf{M}_{ki} \right) \\ &\quad + E_k \left(\frac{\zeta_i}{z\sqrt{np}} \mathbf{M}_{ki} \right) \mathbf{s}_i \mathbf{s}_i^T + \mathbf{s}_i \mathbf{s}_i^T E_k \left(\frac{\zeta_i}{z\sqrt{np}} \mathbf{M}_{ki} \right) \\ &\quad - E_k \left(\frac{\zeta_i}{z} \right) \mathbf{s}_i \mathbf{s}_i^T \\ &= B_1(z) + B_2(z) + B_3(z) + B_4(z) + B_5(z). \end{aligned} \quad (3.36)$$

Indeed, by the decomposition of $\mathbf{X}_k$, write

$$\mathbf{M}_k^{(1)} = \mathbf{X}_{ki} \mathbf{D}_k^{-1} \mathbf{X}_{ki}^T + \mathbf{X}_{ki} \mathbf{D}_k^{-1} \mathbf{e}_i \mathbf{s}_i^T + \mathbf{s}_i \mathbf{e}_i^T \mathbf{D}_k^{-1} \mathbf{X}_{ki}^T + \mathbf{s}_i \mathbf{e}_i^T \mathbf{D}_k^{-1} \mathbf{e}_i \mathbf{s}_i^T.$$

Applying (3.32), (3.33) and (3.35), we obtain

$$\begin{aligned} \mathbf{X}_{ki} \mathbf{D}_k^{-1} \mathbf{X}_{ki}^T &= \mathbf{X}_{ki} \mathbf{D}_{ki,r}^{-1} \mathbf{X}_{ki}^T - \zeta_i \mathbf{X}_{ki}^T \mathbf{D}_{ki,r}^{-1} \mathbf{e}_i \mathbf{h}_i^T \mathbf{D}_{ki,r}^{-1} \mathbf{X}_{ki}^T \\ &= \mathbf{M}_{ki} - \frac{\zeta_i}{z\sqrt{np}} \mathbf{M}_{ki} \mathbf{s}_i \cdot \frac{1}{\sqrt{np}} \mathbf{s}_i^T \mathbf{X}_{ki} \mathbf{D}_{ki,r}^{-1} \mathbf{X}_{ki}^T \\ &= \mathbf{M}_{ki} - \frac{\zeta_i}{znp} \mathbf{M}_{ki} \mathbf{s}_i \mathbf{s}_i^T \mathbf{M}_{ki}. \end{aligned}$$

Similarly,

$$\begin{aligned} \mathbf{X}_{ki} \mathbf{D}_k^{-1} \mathbf{e}_i \mathbf{s}_i^T &= \frac{\zeta_i}{z\sqrt{np}} \mathbf{M}_{ki} \mathbf{s}_i \mathbf{s}_i^T, \quad \mathbf{s}_i \mathbf{e}_i^T \mathbf{D}_k^{-1} \mathbf{X}_{ki}^T = \frac{\zeta_i}{z\sqrt{np}} \mathbf{s}_i \mathbf{s}_i^T \mathbf{M}_{ki}, \\ \mathbf{s}_i \mathbf{e}_i^T \mathbf{D}_k^{-1} \mathbf{e}_i \mathbf{s}_i^T &= \zeta_i \mathbf{s}_i \mathbf{e}_i^T \mathbf{D}_{ki,r}^{-1} \mathbf{e}_i \mathbf{s}_i^T = -\frac{\zeta_i}{z} \mathbf{s}_i \mathbf{s}_i^T. \end{aligned}$$

Summarizing the above and noting $E_k(\mathbf{s}_i) = \mathbf{s}_i$ for $i < k$ yield (3.36), as claimed.

On the other hand, write

$$\mathbf{D}_k = \sum_{i=1(\neq k)}^n \mathbf{e}_i \mathbf{h}_i^T - z \mathbf{I}_{n-1}.$$

Multiplying by $\mathbf{D}_k^{-1}$ on both sides, we have

$$z \mathbf{D}_k^{-1} = -\mathbf{I}_{n-1} + \sum_{i=1(\neq k)}^n \mathbf{e}_i \mathbf{h}_i^T \mathbf{D}_k^{-1}. \quad (3.37)$$

Therefore, by (3.33), (3.35) and the fact that $\mathbf{X}_k \mathbf{X}_k^T = \sum_{i \neq k} \mathbf{s}_i \mathbf{s}_i^T$, we have

$$\begin{aligned} z E_k(\mathbf{M}_k^{(1)}(z)) &= -E_k(\mathbf{X}_k \mathbf{X}_k^T) + \sum_{i=1(\neq k)}^n E_{k-1}(\mathbf{X}_k \mathbf{e}_i \mathbf{h}_i^T \mathbf{D}_k^{-1} \mathbf{X}_k^T) \\ &= -E_k\left(\sum_{i=1(\neq k)}^n \mathbf{s}_i \mathbf{s}_i^T\right) + \sum_{i=1(\neq k)}^n E_k(\zeta_i \mathbf{s}_i \mathbf{h}_i^T \mathbf{D}_{k,i,r}^{-1} (\mathbf{X}_{ki}^T + \mathbf{e}_i \mathbf{s}_i^T)) \\ &= -(n-k) \mathbf{I}_{n-1} - \sum_{i < k} \mathbf{s}_i \mathbf{s}_i^T + \sum_{i=1(\neq k)}^n E_k\left(\frac{\zeta_i}{\sqrt{np}} \mathbf{s}_i \mathbf{s}_i^T \mathbf{M}_{ki}\right) \\ &\quad + \sum_{i=1(\neq k)}^n E_k\left(\zeta_i \vartheta_i \mathbf{s}_i \mathbf{s}_i^T\right). \end{aligned} \quad (3.38)$$

Consequently, by splitting $E_k(\mathbf{M}_k^{(1)}(z_2))$ as in (3.36) for $i < k$ and $z_1 E_k(\mathbf{M}_k^{(1)}(z_1))$ as in (3.38), we obtain

$$\begin{aligned} z_1 \mathbb{Z}_k &= \frac{z_1}{np^2} \text{tr} E_k \mathbf{M}_k^{(1)}(z_1) \cdot E_k \mathbf{M}_k^{(1)}(z_2) \\ &= C_1(z_1, z_2) + C_2(z_1, z_2) + C_3(z_1, z_2) + C_4(z_1, z_2), \end{aligned} \quad (3.39)$$

Chapter 3. Central limit theorem for the linear spectral statistics

where

$$\begin{aligned}
C_1(z_1, z_2) &= -\frac{1}{np^2}(n-k)\text{tr} E_k \mathbf{M}_k^{(1)}(z_2), \\
C_2(z_1, z_2) &= -\frac{1}{np^2} \sum_{i < k} \mathbf{s}_i^T \left(\sum_{j=1}^5 B_j(z_2) \right) \mathbf{s}_i = \sum_{j=1}^5 C_{2j}, \\
C_3(z_1, z_2) &= \frac{1}{np^2} \sum_{i < k} E_k \left[\frac{\zeta_i(z_1)}{\sqrt{np}} \mathbf{s}_i^T \mathbf{M}_{ki}(z_1) \left(\sum_{j=1}^5 B_j(z_2) \right) \mathbf{s}_i \right] \\
&\quad + \frac{1}{np^2} \sum_{i > k} E_k \left[\frac{\zeta_i(z_1)}{\sqrt{np}} \mathbf{s}_i^T \mathbf{M}_{ki}(z_1) E_k \mathbf{M}_k^{(1)}(z_2) \mathbf{s}_i \right] = \sum_{j=1}^6 C_{3j}, \\
C_4(z_1, z_2) &= \frac{1}{np^2} \sum_{i < k} E_k \left[\zeta_i(z_1) \vartheta_i(z_1) \mathbf{s}_i^T \left(\sum_{j=1}^5 B_j(z_2) \right) \mathbf{s}_i \right] \\
&\quad + \frac{1}{np^2} \sum_{i > k} E_k \left[\zeta_i(z_1) \vartheta_i(z_1) \mathbf{s}_i^T E_k \mathbf{M}_k^{(1)}(z_2) \mathbf{s}_i \right] = \sum_{j=1}^6 C_{4j},
\end{aligned}$$

where C_{2j} corresponds to B_j , $j = 1, \dots, 5$, for example $C_{21} = -\frac{1}{np^2} \sum_{i < k} \mathbf{s}_i^T \left(B_1(z_2) \right) \mathbf{s}_i$, and C_{3j} and C_{4j} are similarly defined. Here both $C_3(z_1, z_2)$ and $C_4(z_1, z_2)$ are broken up into two parts in terms of $i > k$ or $i < k$. As will be seen, the terms in (3.39) tend to 0 in L_1 , except C_{25}, C_{34}, C_{45} . Next let us demonstrate the details.

3.3.3.2 Conclusion of the asymptotic expansion of $\mathbb{Z}_k$

The purpose is to analyze each term in $C_j(z_1, z_2)$, $j = 1, 2, 3, 4$. We first claim the limits of ζ_i, ϑ_i which appear in $C_j(z_1, z_2)$ for $j = 2, 3, 4$:

$$\vartheta_i \xrightarrow{L_4} m(z)/z, \quad \zeta_i(z) \xrightarrow{L_4} -zm(z), \quad \text{as } n \rightarrow \infty. \quad (3.40)$$

Indeed, by (3.33) and (3.35), we have

$$\vartheta_i = \frac{1}{znp} \mathbf{s}_i^T \mathbf{M}_{ki} \mathbf{s}_i - \frac{1}{z\sqrt{np}} (\mathbf{s}_i^T \mathbf{s}_i - p). \quad (3.41)$$

Replacing $\mathbf{M}_k^{(m)}$ in $\gamma_{km}(z)$ by $\mathbf{M}_{ki}$, by a proof similar to (3.84), we have

$$E \left| \frac{1}{np} \mathbf{s}_i^T \mathbf{M}_{ki} \mathbf{s}_i - \frac{1}{np} \text{tr} \mathbf{M}_{ki} \right|^4 \leq K \left(\frac{1}{n^2} + \frac{1}{np} \right). \quad (3.42)$$

By (3.21), we then have $\vartheta_i - \frac{1}{zn} \text{tr} \mathbf{D}_{ki}^{-1} \xrightarrow{L_4} 0$. To investigate the distance between $\text{tr} \mathbf{D}_{ki}^{-1}$ and $\text{tr} \mathbf{D}_k^{-1}$, let $\dot{\mathbf{B}}_{ki}$ be the matrix constructed from $\mathbf{B}_k$ by deleting its i -th (or $(i-1)$ -th) row and i -th (or $(i-1)$ -th) column and write $\dot{\mathbf{D}}_{ki} \triangleq \dot{\mathbf{D}}_{ki}(z) = \dot{\mathbf{B}}_{ki} - z \mathbf{I}_{n-2}$ if $i < k$ (or $i > k$). We observe that $\dot{\mathbf{D}}_{ki}^{-1}$ can be obtained from $\mathbf{D}_{ki}^{-1}$ by deleting the i -th (or $(i-1)$ -th) row and i -th (or $(i-1)$ -th) column if $i < k$ (or $i > k$). Then $\text{tr} \mathbf{D}_{ki}^{-1} - \text{tr} \dot{\mathbf{D}}_{ki}^{-1} = -\frac{1}{z}$. By an identity similar to (3.18) and an inequality similar to the bound (3.23), we also have $|\text{tr} \mathbf{D}_k^{-1} - \text{tr} \dot{\mathbf{D}}_{ki}| \leq 1/v_1$. Hence $|\text{tr} \mathbf{D}_k^{-1} - \text{tr} \mathbf{D}_{ki}^{-1}| \leq (1/v_1 + 1/|z|)$. From (3.18) we have $|\text{tr} \mathbf{D}_k^{-1} - \text{tr} \dot{\mathbf{D}}| \leq 1/v_1$ as well. As $\frac{1}{n} \text{tr} \mathbf{D}^{-1} \xrightarrow{L_t} m(z)$ for any fixed t by the Helly-bray theorem and the dominated convergence theorem, we obtain the first conclusion of (3.40).

Since the imaginary part of $(z\zeta_i)^{-1}$ is $(\text{Im}(z) + \frac{1}{np} \text{Im}(\mathbf{s}_i^T \mathbf{M}_{ki} \mathbf{s}_i))$ whose absolute value is greater than v_1 , we have $|\zeta_i| \leq |z|/v_1$. Consequently, via (3.17), we complete the proof of the second consequence of (3.40), as claimed.

Consider $C_1(z_1, z_2)$ first. By (3.21),

$$E|C_1(z_1, z_2)| = E \left| -\frac{1}{n^2 p} (n-k) \text{tr} E_k \mathbf{M}_k^{(1)}(z_2) \right| \leq \frac{K}{np^2} n^2 p \rightarrow 0. \quad (3.43)$$

Before proceeding, we introduce the inequalities for further simplification in the following. By Lemma 8.10 in Bai and Silverstein (2006) and (3.21), for any matrix $\mathbf{C}$ independent of $\mathbf{s}_i$,

$$\begin{aligned} E|\mathbf{s}_i^T \mathbf{M}_{ki} \mathbf{C} \mathbf{s}_i|^2 &\leq K \left(E|\mathbf{s}_i^T \mathbf{M}_{ki} \mathbf{C} \mathbf{s}_i - \text{tr} \mathbf{M}_{ki} \mathbf{C}|^2 + K E|\text{tr} \mathbf{M}_{ki} \mathbf{C}|^2 \right) \\ &\leq K p^2 n^2 E \|\mathbf{C}\|^2, \end{aligned} \quad (3.44)$$

Chapter 3. Central limit theorem for the linear spectral statistics

where we also use the fact that, via (3.21),

$$\begin{aligned}
\left| \text{tr} \mathbf{M}_{ki} \mathbf{C} \overline{\mathbf{C}} \mathbf{M}_{ki} \right| &= \left| \text{tr} \mathbf{D}_{ki}^{-1/2} \mathbf{X}_{ki}^T \mathbf{C} \overline{\mathbf{C}} \mathbf{X}_{ki} \overline{\mathbf{D}}_{ki}^{-1} \mathbf{X}_{ki}^T \mathbf{X}_{ki} \mathbf{D}_{ki}^{-1/2} \right| \\
&\leq n \|\mathbf{D}_{ki}^{-1/2} \mathbf{X}_{ki}^T\|^2 \cdot \|\mathbf{C}\|^2 \cdot \|\mathbf{X}_{ki} \overline{\mathbf{D}}_{ki}^{-1} \mathbf{X}_{ki}^T\| \\
&= n \cdot \|\mathbf{C}\|^2 \cdot \|\mathbf{D}_{ki}^{-1} \mathbf{X}_{ki}^T \mathbf{X}_{ki}\| \\
&= n \cdot \|\mathbf{C}\|^2 \cdot \|p \mathbf{D}_{ki}^{-1} + \sqrt{np}(I_{n-1} + z \mathbf{D}_{ki}^{-1})\|^2 \\
&\leq K n p^2 \|\mathbf{C}\|^2.
\end{aligned}$$

For $i > k$, since $E_k \mathbf{M}_k$ is independent of $\mathbf{s}_i$, we similarly have

$$E |\mathbf{s}_i^T E_k \mathbf{M}_k \mathbf{C} \mathbf{s}_i|^2 \leq K n^2 p^2. \quad (3.45)$$

Applying Cauchy-Schwarz's inequality, (3.44) with $\mathbf{C} = \mathbf{I}_{n-1}$ and the fact that $|\zeta_i|$ is bounded by $|z|/v_1$ we have

$$E |C_{2j}| \leq K \sqrt{\frac{n}{p}}, \quad j = 1, 2, 3, 4. \quad (3.46)$$

Using (3.44) with $\mathbf{C} = E_k \mathbf{M}_{ki}(z_2)$ or $\mathbf{C} = E_k \mathbf{M}_k$ in (3.44), we also have

$$E |C_{3j}| \leq K \sqrt{\frac{n}{p}}, \quad j = 1, 2, 3, 4. \quad (3.47)$$

By (3.45), (3.40) and (3.44) with $\mathbf{C} = \mathbf{I}_{n-1}$, we obtain

$$E |C_{4j}| \leq K \frac{n}{p}, \quad j = 1, 2, 3, 4, 6. \quad (3.48)$$

Consider C_{32} now. Define $\check{\zeta}_i$ and $\check{\mathbf{M}}_{ki}$, the analogues of $\zeta_i(z)$ and $\mathbf{M}_{ki}(z)$ respectively, by $\left(\mathbf{s}_1, \dots, \mathbf{s}_k, \check{\mathbf{s}}_{k+1}, \dots, \check{\mathbf{s}}_n \right)^T$, where $\check{\mathbf{s}}_{k+1}, \dots, \check{\mathbf{s}}_n$ are *i.i.d.* copies of $\mathbf{s}_{k+1}, \dots, \mathbf{s}_n$ and independent of $\mathbf{s}_1, \dots, \mathbf{s}_n$. Then $\check{\zeta}_i, \check{\mathbf{M}}_{ki}$ have the same properties as $\zeta_i(z), \mathbf{M}_{ki}(z)$, respectively. Therefore, $|\check{\zeta}_i| \leq |z|/v_1$ and $\|\check{\mathbf{M}}_{ki}\| \leq Kp$.

Applying (3.44) with $\mathbf{C} = \check{\mathbf{M}}_{ki}(z_1)$, we have

$$\begin{aligned} E|C_{32}| &= E \left| \frac{1}{np^2} \sum_{i < k} E_k E_k \left(\frac{\zeta_i(z_1)}{\sqrt{np}} \mathbf{s}_i^T \mathbf{M}_{ki}(z_1) \frac{\check{\zeta}_i(z_2)}{z_2 np} \check{\mathbf{M}}_{ki}(z_2) \mathbf{s}_i \mathbf{s}_i^T \check{\mathbf{M}}_{ki}(z_2) \mathbf{s}_i \right) \right| \\ &\leq \frac{K}{n^2 p^3 \sqrt{np}} \sum_{i < k} E^{\frac{1}{2}} \left| \mathbf{s}_i^T \mathbf{M}_{ki}(z_1) \check{\mathbf{M}}_{ki}(z_2) \mathbf{s}_i \right|^2 \cdot E^{\frac{1}{2}} \left| \mathbf{s}_i^T \check{\mathbf{M}}_{ki}(z_2) \mathbf{s}_i \right|^2 \\ &\leq K \sqrt{\frac{n}{p}}. \end{aligned} \tag{3.49}$$

Thirdly, consider C_{25} . In view of (3.40), it is straightforward to check that

$$C_{25} = -\frac{k}{n} m(z_2) + o_{L_1}(1) \tag{3.50}$$

Further, consider C_{34} . By (3.40) and (3.44), we have

$$\begin{aligned} C_{34} &= \frac{1}{np^2} \sum_{i < k} E_k \left[\frac{\zeta_i(z_1)}{\sqrt{np}} \mathbf{s}_i^T \mathbf{M}_{ki}(z_1) B_4(z_2) \mathbf{s}_i \right] \\ &= \frac{1}{np^2} \sum_{i < k} E_k \left[\frac{\zeta_i(z_1)}{\sqrt{np}} \mathbf{s}_i^T \mathbf{M}_{ki}(z_1) E_k \left(\frac{\zeta_i(z_2)}{z_2 \sqrt{np}} \mathbf{M}_{ki}(z_2) \right) \mathbf{s}_i \mathbf{s}_i^T \mathbf{s}_i \right] \\ &= z_1 m(z_1) m(z_2) \frac{1}{n^2 p^2} \sum_{i < k} \mathbf{s}_i^T E_k \mathbf{M}_{ki}(z_1) \cdot E_k \mathbf{M}_{ki}(z_2) \mathbf{s}_i + o_{L_1}(1) \tag{3.51} \\ &= z_1 m(z_1) m(z_2) \frac{1}{n^2 p^2} \sum_{i < k} \text{tr} \left(E_k \mathbf{M}_{ki}(z_1) \cdot E_k \mathbf{M}_{ki}(z_2) \right) + o_{L_1}(1) \\ &= z_1 m(z_1) m(z_2) \frac{k}{n} \mathbb{Z}_k + o_{L_1}(1). \end{aligned}$$

where the last step uses the fact that via (3.36), (3.44), (3.33) and a tedious but elementary calculation

$$\frac{1}{np^2} \left| \text{tr} \left(E_k \mathbf{M}_{ki}(z_1) \cdot E_k \mathbf{M}_{ki}(z_2) \right) - \text{tr} E_k \left(\mathbf{X}_k \mathbf{D}_k^{-1}(z_1) \mathbf{X}_k^T \right) \cdot E_k \left(\mathbf{X}_k \mathbf{D}_k^{-1}(z_2) \mathbf{X}_k^T \right) \right| \leq \frac{K}{n}.$$

Consider C_{45} finally. By (3.40), we have

$$C_{45} = -m^2(z_1) m(z_2) \frac{k}{n} + o_{L_1}(1). \tag{3.52}$$

Chapter 3. Central limit theorem for the linear spectral statistics

We conclude from (3.39), (3.43), (3.46)-(3.52) and the fact $m^2(z) + zm(z) + 1 = 0$ that

$$\begin{aligned} z_1 \mathbb{Z}_k &= -\frac{k}{n}m(z_2) - \frac{k}{n}m^2(z_1)m(z_2) + \frac{k}{n}z_1m(z_1)m(z_2)\mathbb{Z}_k + o_{L_1}(1) \\ &= \frac{k}{n}z_1m(z_1)m(z_2) + \frac{k}{n}z_1m(z_1)m(z_2)\mathbb{Z}_k + o_{L_1}(1), \end{aligned}$$

which is equivalent to

$$\mathbb{Z}_k = \frac{\frac{k}{n}m(z_1)m(z_2)}{1 - \frac{k}{n}m(z_1)m(z_2)} + o_{L_1}(1). \quad (3.53)$$

3.3.4 Verify Condition 2

The equality (3.53) ensures that

$$\begin{aligned} \frac{1}{n^2 p^2} \sum_{k=1}^n \text{tr} E_k \mathbf{M}_k^{(1)}(z_1) \cdot E_k \mathbf{M}_k^{(1)}(z_2) &= \frac{1}{n} \sum_{k=1}^n \mathbb{Z}_k \\ \rightarrow \int_0^1 \frac{tm(z_1)m(z_2)}{1 - tm(z_1)m(z_2)} dt &= -1 - \left(m(z_1)m(z_2)\right)^{-1} \log \left(1 - m(z_1)m(z_2)\right). \end{aligned}$$

Thus, via (3.31), we obtain

$$\tilde{\Lambda}_n(z_1, z_2) \xrightarrow{i.p.} \nu_4 - 3 - 2 \left(m(z_1)m(z_2)\right)^{-1} \log \left(1 - m(z_1)m(z_2)\right).$$

Consequently, by (3.29)

$$\begin{aligned} \Lambda(z_1, z_2) &= \frac{\partial^2}{\partial z_1 \partial z_2} \left[(\nu_4 - 3)m(z_1)m(z_2) - 2 \log \left(1 - m(z_1)m(z_2)\right) \right] \\ &= m'(z_1)m'(z_2) \left[\nu_4 - 3 + 2 \left(1 - m(z_1)m(z_2)\right)^{-2} \right]. \end{aligned} \quad (3.54)$$

3.3.5 Tightness of $M_n^{(1)}(z)$

This section is to prove tightness of $M_n^{(1)}(z)$ for $z \in \mathbb{C}_1$. By (3.22) and Lemma 2,

$$E \left| \sum_{k=1}^n \sum_{j=1}^l a_j E_{k-1}(\alpha_k(z_j)) \right|^2 \leq K \sum_{k=1}^n \sum_{j=1}^l |a_j|^2 E |\alpha_k(z_j)|^2 \leq K,$$

which ensures condition (i) of Theorem 12.3 of Billingsley (1968). Condition (ii) of Theorem 12.3 of Billingsley (1968) will be verified by proving

$$\frac{E|M_n^{(1)}(z_1) - M_n^{(1)}(z_2)|^2}{|z_1 - z_2|^2} \leq K, \quad z_1, z_2 \in \mathbb{C}_1. \quad (3.55)$$

We employ the same notation as those in Section 3.3.1. Let

$$\begin{aligned} \Upsilon_{k1} &= \frac{1}{np} \mathbf{s}_k^T \mathbf{X}_k \mathbf{D}_k^{-1}(z_1) \left(\mathbf{D}_k^{-1}(z_1) + \mathbf{D}_k^{-1}(z_2) \right) \mathbf{D}_k^{-1}(z_2) \mathbf{X}_k^T \mathbf{s}_k \\ &\quad - \frac{1}{np} \text{tr} \mathbf{X}_k \mathbf{D}_k^{-1}(z_1) \left(\mathbf{D}_k^{-1}(z_1) + \mathbf{D}_k^{-1}(z_2) \right) \mathbf{D}_k^{-1}(z_2) \mathbf{X}_k^T, \\ \Upsilon_{k2} &= \frac{1}{np} \left(\mathbf{s}_k^T \mathbf{X}_k \mathbf{D}_k^{-1}(z_2) \mathbf{D}_k^{-1}(z_1) \mathbf{X}_k^T \mathbf{s}_k - \text{tr} \mathbf{X}_k \mathbf{D}_k^{-1}(z_2) \mathbf{D}_k^{-1}(z_1) \mathbf{X}_k^T \right), \\ d_{k1}(z) &= \beta_k(z) \left(1 + \frac{1}{np} \mathbf{s}_k^T \mathbf{M}_k^{(2)}(z) \mathbf{s}_k \right), \\ d_{k2}(z) &= 1 + \frac{1}{np} \text{tr} \mathbf{M}_k^{(2)}(z), \\ d_{k3} &= 1 + \frac{1}{np} \text{tr} \mathbf{X}_k \mathbf{D}_k^{-1}(z_2) \mathbf{D}_k^{-1}(z_1) \mathbf{X}_k^T, \\ d_{k4} &= \frac{1}{np} \text{tr} \mathbf{X}_k \mathbf{D}_k^{-1}(z_1) \left(\mathbf{D}_k^{-1}(z_1) + \mathbf{D}_k^{-1}(z_2) \right) \mathbf{D}_k^{-1}(z_2) \mathbf{X}_k^T. \end{aligned}$$

As in (3.19) we write

$$\begin{aligned} M_n^{(1)}(z_1) - M_n^{(1)}(z_2) &= - \sum_{k=1}^n (E_k - E_{k-1}) (d_{k1}(z_1) - d_{k1}(z_2)) \\ &= - (z_1 - z_2) \sum_{k=1}^n (E_k - E_{k-1}) \left[\beta_k(z_1) (\Upsilon_{k1} + d_{k4}) - \beta_k(z_1) d_{k1}(z_2) (\Upsilon_{k2} + d_{k3}) \right] \\ &= - (z_1 - z_2) \sum_{k=1}^n (E_k - E_{k-1}) \left[(l_1 + l_2) + l_3 - \beta_k(z_1) \beta_k(z_2) d_{k2} d_{k3} - \right. \\ &\quad \left. \beta_k(z_1) \beta_k(z_2) d_{k3} \gamma_k(z_2) \right] \\ &= - (z_1 - z_2) \sum_{k=1}^n (E_k - E_{k-1}) (l_1 + l_2 + l_3 + l_4 + l_5 + l_6), \end{aligned}$$

Chapter 3. Central limit theorem for the linear spectral statistics

where

$$\begin{aligned}
l_1 &= \beta_k(z_1)\Upsilon_{k1}, \quad l_2 = \beta_k(z_1)\beta_k^{tr}(z_1)\eta_k(z_1)d_{k4}, \\
l_3 &= -\beta_k(z_1)\Upsilon_{k2}d_{k1}(z_1), \\
l_4 &= -\beta_k(z_1)\beta_k^{tr}(z_1)\eta_k(z_1)\beta_k(z_2)d_{k2}(z_2)d_{k3}, \\
l_5 &= -\beta_k^{tr}(z_1)\beta_k(z_2)\beta_k^{tr}(z_2)\eta_k(z_2)d_{k2}(z_2)d_{k3}, \\
l_6 &= -\beta_k(z_1)\beta_k(z_2)d_{k3}\gamma_k(z_2),
\end{aligned}$$

Here the last step uses (3.20) for $\beta_k(z_1)$ and the facts that

$$\begin{aligned}
\mathbf{D}_k^{-2}(z_1) - \mathbf{D}_k^{-2}(z_2) &= (z_1 - z_2)\mathbf{D}_k^{-1}(z_1)\left(\mathbf{D}_k^{-1}(z_1) + \mathbf{D}_k^{-1}(z_2)\right)\mathbf{D}_k^{-1}(z_2), \\
\beta_k(z_1) - \beta_k(z_2) &= (z_2 - z_1)\beta_k(z_1)\beta_k(z_2)\Upsilon_{k2} + (z_2 - z_1)\beta_k(z_1)\beta_k(z_2)d_{k3}, \\
(E_k - E_{k-1})\beta_k^{tr}(z_1)d_{k4} &= 0, \quad (E_k - E_{k-1})\beta_k^{tr}(z_1)\beta_k^{tr}(z_2)d_{k2}(z_2)d_{k3} = 0.
\end{aligned}$$

By (3.21) and Lemma 8.10 in Bai and Silverstein (2006), without any tedious calculations, one may verify that

$$|d_{kj}(z)| \leq K, \quad j = 1, 2, 3, 4, \quad \text{and} \quad E|\Upsilon_{kj}|^2 \leq Kp^{-1}, \quad j = 1, 2,$$

The above inequalities, together with Burkholder's inequality, imply (3.55).

3.4 Uniform convergence of $EM_n(z)$

To finish the proof of Proposition 1, it remains to derive an asymptotic expansion of $n(Em_n(z) - m(z))$ for $z \in \mathbb{C}_1$ (defined in Section 3.2.2). In order to unify the proof of Theorem 3 and Theorem 4, we derive the asymptotic expansion of $n(Em_n(z) - m(z))$ under both assumptions $n^3/p = O(1)$ and $n/p \rightarrow 0$ in Proposition 3. For the purpose of proving (3.15), we will prove a stronger result in Proposition 3, namely uniform convergence of

$n(Em_n(z) - m(z))$ for $z \in \varsigma_n = \bigcup_{i=l,r,u} \varsigma_i$. For z located in the wider range ς_n , the bounds or limits in Section 3.2 (e.g. Lemma 2, (3.28), (3.40)), can not be applied directly. Hence in Section 3.8, we re-establish these and other useful results. Throughout this section, we assume $z \in \varsigma_n$ and use the same notations as that in Section 3.2.

Proposition 3. *Suppose the Assumption (c) is satisfied.*

(i) *Under Assumption (b1): $n^3/p = O(1)$, we have the following asymptotic expansion*

$$\begin{aligned} n \left[Em_n(z) - m(z) + \sqrt{\frac{n}{p}} m^4(z) (1 + m'(z)) \right] \\ = m^3(z) (m'(z) + \nu_4 - 2) (1 + m'(z)) + o(1), \end{aligned} \quad (3.56)$$

uniformly for $z \in \varsigma_n = \bigcup_{i=l,r,u} \varsigma_i$.

(ii) *Under Assumption (b2): $n/p \rightarrow 0$, we have the following asymptotic expansion*

$$n \left[Em_n(z) - m(z) - \mathcal{X}_n(m(z)) \right] = o(1), \quad (3.57)$$

uniformly for $z \in \varsigma_n = \bigcup_{i=l,r,u} \varsigma_i$, where $\mathcal{X}_n(m)$ is defined in (3.8).

By the definition of $M_n(z)$, $EM(z) = 0$ in Proposition 1 is a direct consequence of Proposition 3. This, together with (3.54) and the tightness of $M_n^{(1)}(z)$ in Section 3.3.5, implies Proposition 1. It remains to prove Proposition 3. To facilitate statements, let

$$\begin{aligned} \omega_n &= \frac{1}{n} \sum_{k=1}^n m(z) \beta_k \bar{\mu}_k, \quad \bar{\epsilon}_n = \frac{1}{z + E \frac{1}{np} \text{tr} \mathbf{X} \mathbf{D}^{-1}(z) \mathbf{X}^T}, \\ \bar{\mu}_k &= \frac{1}{\sqrt{np}} (\mathbf{s}_k^T \mathbf{s}_k - p) - \mathbf{q}_k^T \mathbf{D}_k^{-1}(z) \mathbf{q}_k + E \frac{1}{np} \text{tr} \mathbf{X} \mathbf{D}^{-1}(z) \mathbf{X}^T. \end{aligned}$$

Here, $\omega_n, \bar{\epsilon}_n, \bar{\mu}_k$ all depend on z and n , and $\bar{\epsilon}_n$ are non-random.

Chapter 3. Central limit theorem for the linear spectral statistics

Lemma 1. *Let $z \in \varsigma_n$. We have*

$$nE\omega_n = m^3(z) \left(m'(z) + \nu_4 - 2 \right) + o(1).$$

Assuming Lemma 1 is true for the moment, whose proof is given in Section 1 below, let us demonstrate how to get Proposition 3. By (3.8) in Bai (1993), we obtain

$$m_n(z) = \frac{1}{n} \text{tr} \mathbf{D}^{-1}(z) = -\frac{1}{n} \sum_{k=1}^n \beta_k. \quad (3.58)$$

Applying (3.17), (3.58), (3.21) and taking the difference of β_k and $\frac{1}{z+m(z)}$, we observe that

$$\begin{aligned} Em_n(z) - m(z) &= -\frac{1}{n} \sum_{k=1}^n E\beta_k + \frac{1}{z + m(z)} \\ &= E \frac{1}{n} \sum_{k=1}^n \beta_k m(z) \left[\bar{\mu}_k - (Em_n(z) - m(z)) - \sqrt{\frac{n}{p}} (1 + zEm_n(z)) \right] \\ &= E\omega_n + m(z)Em_n(z)(Em_n(z) - m(z)) \\ &\quad + \sqrt{\frac{n}{p}} m(z)Em_n(z)(1 + zEm_n(z)). \end{aligned} \quad (3.59)$$

Under Assumption $n^3/p = O(1)$: subtracting $m(z)Em_n(z)(Em_n(z) - m(z))$ on both side of (3.59) and then dividing $\frac{1}{n}(1 - m(z)Em_n(z))$, we have

$$\begin{aligned} &n(Em_n(z) - m(z)) \\ &= \frac{nE\omega_n}{1 - m(z)Em_n(z)} + \sqrt{\frac{n^3}{p}} \frac{m(z)Em_n(z)(1 + zEm_n(z))}{1 - m(z)Em_n(z)} \\ &= \frac{m^3(z)}{1 - m^2(z)} \left(m'(z) + \nu_4 - 2 \right) - \sqrt{\frac{n^3}{p}} \frac{m^4(z)}{1 - m^2(z)} + o\left(\sqrt{\frac{n^3}{p}}\right), \end{aligned}$$

where we use (3.62) below, Lemma 1, (3.17) and the fact that $m'(z) = \frac{m^2(z)}{1 - m^2(z)}$. Proposition 3(i) is proved.

Under Assumption $n/p \rightarrow 0$: Let Em_n, m denote $Em_n(z), m(z)$ to simplify the notation below. By (3.17), we have

$$\begin{aligned}
Em_n - m &= E\omega_n + m^2(Em_n - m) + m(Em_n - m)^2 \\
&\quad + \sqrt{\frac{n}{p}}m(Em_n - m)(1 + zm) + \sqrt{\frac{n}{p}}m^2(1 + zm) \\
&\quad + \sqrt{\frac{n}{p}}zm(Em_n - m)^2 + \sqrt{\frac{n}{p}}zm^2(Em_n - m) \\
&= \mathcal{A}(Em_n - m)^2 + (\mathcal{B} + 1)(Em_n - m) + \mathcal{C}_n.
\end{aligned}$$

where $\mathcal{A}, \mathcal{B}$ are defined in (3.8) and

$$\mathcal{C}_n = E\omega_n - \sqrt{\frac{n}{p}}m^4.$$

Rearranging the above equation, we observe that $(Em_n - m)$ satisfies the equation $\mathcal{A}x^2 + \mathcal{B}x + \mathcal{C}_n = 0$. Solving the equation, we obtain

$$x_{(1)} = \frac{-\mathcal{B} + \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}_n}}{2\mathcal{A}}, \quad x_{(2)} = \frac{-\mathcal{B} - \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}_n}}{2\mathcal{A}},$$

where $\sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}_n}$ is the complex number whose imaginary part has the same sign as the imaginary part of $\mathcal{B}$. By the assumption $n/p \rightarrow 0$ and Lemma 1, we have $4\mathcal{A}\mathcal{C}_n \rightarrow 0$. Then $x_{(1)} = o(1)$ and $x_{(2)} = \frac{1-m^2}{m} + o(1)$. Since $Em_n - m = o(1)$, we choose $Em_n - m = x_{(1)}$. Applying Lemma 1 and the define of $\mathcal{X}_n(m)$ in (3.8), we have

$$\begin{aligned}
&n \left[Em_n(z) - m(z) - \mathcal{X}_n(m(z)) \right] \\
&= \frac{-4\mathcal{A} \left[nE\omega_n - m^3(z) \left(m'(z) + \nu_4 - 2 \right) \right]}{2\mathcal{A} \left(\sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}_n} + \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}} \right)} \\
&\rightarrow 0.
\end{aligned}$$

Hence Proposition 3(ii) is proved. Hence, the proof of Proposition 3 is complete. Now it remains to prove Lemma 1.

3.4.1 Proof of Lemma 1

From the definitions of β_k , $\bar{\epsilon}_n$ and $\bar{\mu}_k$, we obtain

$$\beta_k = \bar{\epsilon}_n + \beta_k \bar{\epsilon}_n \bar{\mu}_k. \quad (3.60)$$

By (3.60), we further write β_k as $\beta_k = \bar{\epsilon}_n + \bar{\epsilon}_n^2 \bar{\mu}_k + \bar{\epsilon}_n^3 \bar{\mu}_k^2 + \beta_k \bar{\epsilon}_n^3 \bar{\mu}_k^3$, which ensures that

$$\begin{aligned} nE\omega_n &= m(z) \bar{\epsilon}_n \sum_{k=1}^n E(\bar{\mu}_k) + m(z) \bar{\epsilon}_n^2 \sum_{k=1}^n E(\bar{\mu}_k^2) \\ &\quad + m(z) \bar{\epsilon}_n^3 \sum_{k=1}^n E(\bar{\mu}_k^3) + m(z) \bar{\epsilon}_n^3 \sum_{k=1}^n E(\beta_k \bar{\mu}_k^4) \\ &\triangleq H_1 + H_2 + H_3 + H_4, \end{aligned} \quad (3.61)$$

where $H_j, j = 1, 2, 3, 4$ are defined in the obvious way. As will be seen, H_3 and H_4 are both negligible and the contribution to the limit of $nE\omega_n$ comes from H_1 and H_2 . Now, we analyze $H_j, j = 1, \dots, 4$ one by one.

Consider H_4 first. Since $F_n \xrightarrow{a.s.} F$ as $n \rightarrow \infty$, we conclude from (3.91), (3.92), (3.94) below and the dominated convergence theorem that, for any fixed positive integer t

$$E|m_n(z) - m(z)|^t \rightarrow 0. \quad (3.62)$$

It follows from (3.17) and (3.62) that

$$\bar{\epsilon}_n = \frac{1}{z + m(z) + o(1)} = -m(z) + o(1). \quad (3.63)$$

By Lemma 3 and Lemma 4 in the last section,

$$\begin{aligned} E|\beta_k \bar{\mu}_k^4| &\leq KE|\bar{\mu}_k^4|I(U_n^c) + E|\beta_k \bar{\mu}_k^4|I(U_n) \\ &\leq K\left(\frac{\delta^4}{n} + \frac{n}{p^2}\right) + o(n^{-\ell}) \leq K\delta^4 n^{-1}, \end{aligned}$$

which, together with (3.63), further implies

$$H_4 = o(1). \quad (3.64)$$

It follows from Lemma 4 and (3.63) that

$$H_3 = o(1). \quad (3.65)$$

Consider H_1 next. We have, via (3.21) and (3.18),

$$\begin{aligned} H_1 &= m(z)\bar{\epsilon}_n \sum_{k=1}^n \left(E \frac{1}{np} \text{tr} \mathbf{X} \mathbf{D}^{-1} \mathbf{X}^T - E \frac{1}{np} \text{tr} \mathbf{M}_k^{(1)} \right) \\ &= \left(1 + z \sqrt{\frac{n}{p}} \right) m(z) \bar{\epsilon}_n \frac{1}{n} \sum_{k=1}^n E \left(\text{tr} \mathbf{D}^{-1} - \text{tr} \mathbf{D}_k^{-1} \right) + \sqrt{\frac{n}{p}} m(z) \bar{\epsilon}_n \\ &= - \left(1 + z \sqrt{\frac{n}{p}} \right) m(z) \bar{\epsilon}_n \frac{1}{n} \sum_{k=1}^n E \left[\beta_k \left(1 + \frac{1}{np} \mathbf{s}_k^T \mathbf{M}_k^{(2)} \mathbf{s}_k \right) \right] \\ &\quad + \sqrt{\frac{n}{p}} m(z) \bar{\epsilon}_n. \end{aligned} \quad (3.66)$$

Applying (3.96), (3.99) below and (3.62), it is easy to see

$$1 + \frac{1}{np} \mathbf{s}_k^T \mathbf{M}_k^{(2)} \mathbf{s}_k = 1 + \left(\frac{1}{np} \text{tr} \mathbf{M}_k^{(1)} \right)' + o_{L_4}(1) = 1 + m'(z) + o_{L_4}(1),$$

This, together with (3.63), Lemma 3 and (3.58), ensures that

$$H_1 = -m^2(z) \left(1 + m'(z) \right) E m_n(z) + o(1) \rightarrow -m^3(z) \left(1 + m'(z) \right). \quad (3.67)$$

Consider H_2 , by the previous estimation of $E\bar{\mu}_k$ included in H_1 , we obtain

$$E\bar{\mu}_k^2 = E(\bar{\mu}_k - E\bar{\mu}_k)^2 + O(n^{-2}). \quad (3.68)$$

Furthermore, by the definition, a direct calculation yields

$$E(\bar{\mu}_k - E\bar{\mu}_k)^2 = S_1 + S_2, \quad (3.69)$$

Chapter 3. Central limit theorem for the linear spectral statistics

where

$$\begin{aligned} S_1 &= \frac{1}{n} E (X_{11}^2 - 1)^2 + E \gamma_{k1}^2, \quad S_2 = S_{21} + S_{22}, \\ S_{21} &= \frac{1}{n^2 p^2} E \left(\text{tr} \mathbf{M}_k^{(1)} - E \text{tr} \mathbf{M}_k^{(1)} \right)^2, \\ S_{22} &= -\frac{2}{np\sqrt{np}} E \left[(\mathbf{s}_k^T \mathbf{s}_k - p) (\mathbf{s}_k^T \mathbf{M}_k^{(1)} \mathbf{s}_k - E \text{tr} \mathbf{M}_k^{(1)}) \right]. \end{aligned}$$

We claim that

$$nS_1 \rightarrow \nu_4 - 1 + 2m'(z), \quad nS_{21} \rightarrow 0, \quad nS_{22} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (3.70)$$

Indeed, with notation $\mathbf{M}_k^{(1)} = (a_{ij}^{(1)})_{p \times p}$, $i, j = 1, \dots, p$, as illustrated in (3.89), we have $\frac{1}{n^2 p^2} \sum_{k=1}^n \sum_{i=1}^p E |a_{ii}^{(1)}|^2 \rightarrow 0$. Via this, (3.62) and (3.21), a simple calculation yields

$$\begin{aligned} nE\gamma_{k1}^2 &= \frac{1}{np^2} E \left(\sum_{i \neq j} X_{ik} X_{jk} a_{ij}^{(1)} + \sum_{i=1}^p (X_{ik}^2 - 1) a_{ii}^{(1)} \right)^2 \\ &= \frac{1}{np^2} E \left(\sum_{i \neq j} \sum_{s \neq t} X_{ik} X_{jk} X_{sk} X_{tk} a_{ij}^{(1)} a_{st}^{(1)} \right) \\ &\quad + \frac{1}{np^2} \sum_{i=1}^p E \left[(X_{ik}^2 - 1)^2 (a_{ii}^{(1)})^2 \right] \\ &= \frac{2}{np^2} E \left(\sum_{i,j} a_{ij}^{(1)} a_{ji}^{(1)} \right) + o(1) = \frac{2}{np^2} E \text{tr} (\mathbf{M}_k^{(1)})^2 + o(1) \\ &= \frac{2}{n} E \text{tr} \mathbf{D}_k^{-2} + o(1) \rightarrow 2m'(z). \end{aligned}$$

Since $E|X_{11}^2 - 1|^2 = \nu_4 - 1$, we have proved the first result of (3.70). By Burkholder's inequality, Lemma 5, (3.21), (3.91) and (3.96)

$$n|S_{21}| = K(1 + z\sqrt{\frac{n}{p}})^2 \frac{1}{n} E |M^{(1)}(z)|^2 + Kn^{-1} \leq Kn^{-1}. \quad (3.71)$$

Furthermore

$$\begin{aligned}
n|S_{22}| &= \frac{2}{p\sqrt{np}} \left| E \left(\sum_{t=1}^p (X_{tk}^2 - 1) \right) \cdot \left(\sum_{i,j} X_{ik} X_{jk} a_{ij}^{(1)} \right) \right| \\
&= \frac{2}{p\sqrt{np}} |E(X_{11}^2 - 1)X_{11}^2 \cdot E \operatorname{tr} \mathbf{M}_k^{(1)}| \\
&\leq K \sqrt{\frac{n}{p}} + o(n^{-\ell}) \rightarrow 0.
\end{aligned}$$

Therefore, the proof of the second result of (3.70) is complete. We then conclude from (3.70), (3.68), (3.69) and (3.63) that

$$H_2 \rightarrow m^3(z) \left(2m'(z) + \nu_4 - 1 \right). \quad (3.72)$$

Finally, by (3.61), (3.64), (3.65), (3.67) and (3.72), we obtain

$$nE\omega_n = m^3(z) \left(m'(z) + \nu_4 - 2 \right) + o(1).$$

Lemma 1 is thus proved. This finished the proof of Proposition 1.

3.5 Proof of Proposition 2

Recall the definition of U_n below Proposition 2 or in Section 3.8. For $i = l, r$, by Lemma 5

$$E \left| \int_{\varsigma_i} M_n^{(1)}(z) I(U_n^c) dz \right|^2 \leq \int_{\varsigma_i} E |M_n^{(1)}(z)|^2 dz \leq K \|\varsigma_i\| \rightarrow 0, \quad \text{as } n \rightarrow \infty, v_1 \rightarrow 0.$$

Moreover,

$$\left| \int_{\varsigma_i} EM_n(z) I(U_n^c) dz \right| \leq \int_{\varsigma_i} |EM_n(z)| dz \rightarrow 0, \quad \text{as } n \rightarrow \infty, v_1 \rightarrow 0,$$

where the convergence follows from Proposition 3.

3.6 Calculation of the mean and covariance

To complete the proof of Theorem 3, it remains to calculate the mean function and covariance function of $X(f)$. In the computation, we use the same strategy as that in Bai and Yao (2005).

3.6.1 The Mean

In this subsection, we aim to obtain the formula (3.4). It follows from Proposition 1, (3.15) and (3.16) that

$$\begin{aligned} EQ_n(f) &= -\frac{1}{2\pi i} \oint_{\zeta} f(z) EM_n(z) dz \\ &= -\frac{1}{2\pi i} \int_{\zeta} f(z) \Delta(z) dz + o(1) \triangleq EX(f) + o(1). \end{aligned} \quad (3.73)$$

Recalling that $z = -\frac{1}{m} - m$, we may choose a contour

$$\zeta' = \{z = -(\rho e^{i\theta} + \rho^{-1} e^{-i\theta}), 0 \leq \theta < 2\pi\},$$

such that when z runs a cycle along ζ' , m runs a cycle along $|m| = \rho$. However, when z runs along ζ' in the positive direction, m runs along $|m| = \rho$ in the negative direction. On the other hand, by observing $|m| \leq 1$ (see (3.3) in Bai 1993), we may choose $\rho < 1$ so that the contour ζ' is still included in $\mathcal{S}$. By changing the variable z to m , the Cauchy theorem, the facts that $m' = \frac{m^2}{1-m^2}$ and that $dz = \frac{1-m^2}{m^2} dm$, we have

$$\begin{aligned} EX(f) &= \frac{1}{2\pi i} \oint_{|m|=\rho} f(-m - m^{-1}) m \left(\nu_4 - 2 + \frac{m^2}{1-m^2} \right) dm \\ &= \frac{1}{2\pi i} \oint_{|m|=1} f(-m - m^{-1}) m (\nu_4 - 2) dm \\ &\quad + \frac{1}{2\pi i} \oint_{|m|=\rho} f(-m - m^{-1}) \frac{m^3}{1-m^2} dm \\ &= R_1 + R_2. \end{aligned} \quad (3.74)$$

In order to evaluate R_2 , for any ϵ small enough, we introduce a new contour $\mathcal{C}(\epsilon) = \mathcal{C}_1(\epsilon) + \mathcal{C}_2(\epsilon) + \mathcal{C}_3(\epsilon)$, where

$$\mathcal{C}_1(\epsilon) = \{m : m = -e^{i\theta}, -\pi + \epsilon \leq \theta \leq -\epsilon, \epsilon \leq \theta \leq \pi - \epsilon\},$$

$$\mathcal{C}_2(\epsilon) = \{m : m + \cos \epsilon = \sin \epsilon \cdot e^{i\theta}, \frac{\pi}{2} \geq \theta \geq -\frac{\pi}{2}\},$$

and

$$\mathcal{C}_3(\epsilon) = \{m : m - \cos \epsilon = \sin \epsilon \cdot e^{i\theta}, \frac{3\pi}{2} \geq \theta \geq \frac{\pi}{2}\}.$$

Using the Cauchy's theorem first and then letting $\epsilon \rightarrow 0$, we obtain

$$\begin{aligned} R_2 &= \lim_{\epsilon \rightarrow 0} \frac{1}{2\pi i} \left(\int_{\mathcal{C}_1(\epsilon)} + \int_{\mathcal{C}_2(\epsilon)} + \int_{\mathcal{C}_3(\epsilon)} \right) f(-m - m^{-1}) \frac{m^3}{1 - m^2} dm \\ &= \lim_{\epsilon \rightarrow 0} \left(R_{21}(\epsilon) + R_{22}(\epsilon) + R_{23}(\epsilon) \right). \end{aligned} \quad (3.75)$$

Since $\mathcal{C}_1(\epsilon)$ is symmetric, by letting $\epsilon \rightarrow 0$ and $\cos \theta = x$, we have

$$\begin{aligned} R_{21}(\epsilon) &= \frac{1}{2\pi i} \left(\int_{-\pi+\epsilon}^{\epsilon} + \int_{\epsilon}^{\pi-\epsilon} \right) f(-m - m^{-1}) \frac{m^3}{1 - m^2} dm \\ &= \frac{1}{2\pi} \left(\int_{-\pi+\epsilon}^{\epsilon} + \int_{\epsilon}^{\pi-\epsilon} \right) f(2 \cos \theta) \frac{e^{4i\theta}}{1 - e^{2i\theta}} d\theta \\ &= -\frac{1}{\pi} \int_{\epsilon}^{\pi-\epsilon} f(2 \cos \theta) \frac{1}{2} (4 \cos^2 \theta - 1) d\theta \\ &\rightarrow -\frac{1}{\pi} \int_0^{\pi} f(2 \cos \theta) \frac{1}{2} (4 \cos^2 \theta - 1) d\theta \\ &= -\frac{1}{\pi} \int_{-1}^1 f(2x) \left(2x^2 - \frac{1}{2}\right) \frac{1}{\sqrt{1-x^2}} dx. \end{aligned} \quad (3.76)$$

For $z \in \mathcal{C}_2(\epsilon)$, letting $\epsilon \rightarrow 0$, we then have $m \rightarrow -1$ and

$$\frac{dm}{1+m} = \frac{i \sin(\epsilon) \cdot e^{i\theta} d\theta}{1 - \cos \epsilon + \sin \epsilon \cdot e^{i\theta}} \rightarrow i d\theta.$$

Chapter 3. Central limit theorem for the linear spectral statistics

We conclude from the dominated convergence theorem that

$$\begin{aligned}
 \lim_{\epsilon \rightarrow 0} R_{22}(\epsilon) &= \frac{1}{2\pi i} \int_{\frac{\pi}{2}}^{-\frac{\pi}{2}} \lim_{\epsilon \rightarrow 0} f(-m - m^{-1}) \frac{m^3}{1 - m^2} dm \\
 &= \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f(2) \frac{i}{2} d\theta \\
 &= \frac{f(2)}{4}.
 \end{aligned} \tag{3.77}$$

For $z \in \mathcal{C}_3$, let $\epsilon \rightarrow 0$, we similarly have $m \rightarrow 1$ and $\frac{dm}{1+m} \rightarrow -id\theta$. Thus

$$\lim_{\epsilon \rightarrow 0} R_{23}(\epsilon) = \frac{f(-2)}{4}. \tag{3.78}$$

Appealing to a method similar to (3.76), we have

$$R_1 = \frac{1}{\pi} \int_{-1}^1 f(2x) \frac{(\nu_4 - 2)(2x^2 - 1)}{\sqrt{1 - x^2}} dx \tag{3.79}$$

and via Cauchy's Theorem

$$\begin{aligned}
 & - \frac{1}{2\pi i} \sqrt{\frac{n^3}{p}} \oint_{\mathcal{C}} f(z) m^4(z) (1 + m'(z)) dz \\
 &= \frac{1}{2\pi i} \sqrt{\frac{n^3}{p}} \oint_{|m|=1} f(-m - m^{-1}) m^2 dm \\
 &= -\frac{1}{\pi} \sqrt{\frac{n^3}{p}} \int_{-1}^1 f(2x) \frac{4x^3 - 3x}{\sqrt{1 - x^2}} dx.
 \end{aligned} \tag{3.80}$$

It follows from (3.74)-(3.78) that

$$EX(f) = \frac{1}{\pi} \int_{-1}^1 f(2x) \left[2(\nu_4 - 3)x^2 - (\nu_4 - \frac{5}{2}) \right] \frac{1}{\sqrt{1 - x^2}} dx + (f(-2) + f(2))/4,$$

which gives the formula (3.4).

3.6.2 The covariance

To finish the proof of Theorem 3, it remains to prove (3.5) and (3.6), the two representations of the covariance function $\text{cov}(X(f), X(g))$. Let $\varsigma_i, i = 1, 2$

be two disjoint contours with vertices $\pm(2 + \varpi_i) \pm i\tau_i$. The positive values of ϖ_i, τ_i are appropriately selected such that the two contours are contained in $\mathcal{S}$. Then

$$\begin{aligned} & \text{cov}(Q_n(f), Q_n(g)) \\ &= -\frac{1}{4\pi^2} \oint_{\varsigma_1} \oint_{\varsigma_2} f(z_1)g(z_2)EM_n^{(1)}(z_1) \cdot M_n^{(1)}(z_2)dz_1dz_2 \\ &= \frac{1}{4\pi^2} \int_{\varsigma_1} \int_{\varsigma_2} f(z_1)g(z_2)\Lambda(z_1, z_2)dz_1dz_2 + o(1) \\ &\triangleq \text{cov}(X(f), X(g)) + o(1). \end{aligned}$$

Rewrite $\text{cov}(X(f), X(g))$ by integration by parts and apply (3.54)

$$\begin{aligned} & \text{cov}(X(f), X(g)) \\ &= -\frac{1}{4\pi^2} \oint_{\varsigma_1} f(z_1) \left(\oint_{\varsigma_2} g(z_2)\Lambda(z_1, z_2)dz_2 \right) dz_1 \\ &= -\frac{1}{4\pi^2} \oint_{\varsigma_1} \oint_{\varsigma_2} G(z_1, z_2)dz_1dz_2, \end{aligned}$$

where

$$G(z_1, z_2) = f'(z_1)g'(z_2) \left[(\nu_4 - 3)m(z_1)m(z_2) - 2 \log \left(1 - m(z_1)m(z_2) \right) \right].$$

Let $\tau_i \rightarrow 0$ first and then $\varpi_i \rightarrow 0$. It is a simple matter to verify that the integral along the vertical edges of the two contours converges to 0 when $\tau_i \rightarrow 0$. Therefore

$$\begin{aligned} \text{cov}(X(f), X(g)) &= -\frac{1}{4\pi^2} \int_{-2}^2 \int_{-2}^2 [G(x^-, y^-) - G(x^-, y^+) \\ &\quad - G(x^+, y^-) + G(x^+, y^+)] dx dy, \end{aligned}$$

where $x^\pm \triangleq x \pm i0$, $y^\pm \triangleq y \pm i0$. Since f' and g' are both continuous we obtain $f(x^\pm) = f(x)$ and $g(y^\pm) = g(y)$. Recalling that $m(x \pm i0) = \frac{1}{2}(-x \pm i\sqrt{4 - x^2})$ we have

$$\begin{aligned} & m(x^-)m(y^-) - m(x^+)m(y^-) - m(x^-)m(y^+) + m(x^+)m(y^+) \\ &= -\sqrt{4 - x^2}\sqrt{4 - y^2}, \end{aligned}$$

Chapter 3. Central limit theorem for the linear spectral statistics

and

$$\begin{aligned} & \log \left(1 - m(x^-)m(y^-) \right) - \log \left(1 - m(x^+)m(y^-) \right) \\ & - \log \left(1 - m(x^-)m(y^+) \right) + \log \left(1 - m(x^+)m(y^+) \right) \\ & = -\log \left(\frac{4 - xy - \sqrt{(4 - x^2)(4 - y^2)}}{4 - xy + \sqrt{(4 - x^2)(4 - y^2)}} \right). \end{aligned}$$

Consequently,

$$\text{cov}(X(f), X(g)) = \frac{1}{4\pi^2} \int_{-2}^2 \int_{-2}^2 f'(x)g'(y)H(x, y)dx dy,$$

where

$$H(x, y) = (\nu_4 - 3)\sqrt{4 - x^2}\sqrt{4 - y^2} + 2 \log \left(\frac{4 - xy + \sqrt{(4 - x^2)(4 - y^2)}}{4 - xy - \sqrt{(4 - x^2)(4 - y^2)}} \right),$$

which recovers the formula (3.6).

Now, we proceed to get the second representation, formula (3.5). As before, we define the contours $\varsigma'_i, i = 1, 2$ as

$$\begin{aligned} \varsigma'_1 &= \{-(\rho e^{i\theta_1} + \rho^{-1}e^{-i\theta_1}), 0 \leq \theta < 2\pi\}, \\ \varsigma'_2 &= \{-(e^{i\theta_1} + e^{-i\theta_1}), 0 \leq \theta < 2\pi\}, \end{aligned}$$

where $\rho < 1$. Thus $|m(z_1)m(z_2)| = \rho < 1$ with $z_1 \in \varsigma'_1, z_2 \in \varsigma'_2$. Note that $m' = \frac{m^2}{1-m^2}$, $dz = \frac{1-m^2}{m^2}dm$ and that for $j = 1, 2$ when z_j runs along ς'_j in the positive direction, m_j runs along $|m_j| = \rho_j$ ($\rho_1 = \rho$ and $\rho_2 = 1$) in the negative direction. By Cauchy's Theorem we then obtain

$$\begin{aligned} \text{cov}(X(f), X(g)) &= -\frac{1}{4\pi^2} \oint_{\varsigma'_1} \oint_{\varsigma'_2} f(z_1)g(z_2)\Lambda(z_1, z_2)dz_1dz_2 \\ &= -\frac{1}{4\pi^2} \oint_{|m_1|=\rho} \oint_{|m_2|=1} f(-m_1 - m_2^{-1})g(-m_2 - m_2^{-1}) \\ &\quad \times \left(\nu_4 - 3 + \frac{2}{(1 - m_1m_2)^2} \right) dm_1dm_2. \end{aligned} \tag{3.81}$$

Setting $m_1 = -\rho e^{i\theta_1}$, $m_2 = -e^{i\theta}$, we change the variables m_1, m_2 to θ_1, θ_2 .

It follows from Taylor's expansion that

$$\begin{aligned}
 (3.81) &= -\frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(\rho e^{i\theta_1} + \rho^{-1} e^{-i\theta_2}) g(2 \cos \theta_2) \\
 &\quad \times \left[(\nu_4 - 3) \rho e^{i(\theta_1 + \theta_2)} + 2 \sum_{k=1}^{\infty} k \rho^k e^{ik(\theta_1 + \theta_2)} \right] d\theta_1 d\theta_2 \\
 &= (\nu_4 - 3) \rho \Psi_1(g) + 2 \sum_{k=1}^{\infty} k \rho^k \Psi_k(f, \rho) \Psi_k(g),
 \end{aligned}$$

where $\Psi_k(g)$ is defined in Section 2 and

$$\Psi_k(f, \rho) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\rho e^{i\theta} + \rho^{-1} e^{-i\theta}) d\theta.$$

Applying integration by parts yields for $k \geq 3$

$$\begin{aligned}
 \Psi_k(f, \rho) &= \frac{\rho^{-1}}{k} \Psi_{k-1}(f', \rho) - \frac{\rho}{k} \Psi_{k+1}(f', \rho) \\
 &= \frac{\rho^2}{k(k+1)} \Psi_{k+1}(f'', \rho) - \frac{2}{k^2 - 1} \Psi_k(f'', \rho) + \frac{\rho^{-2}}{k(k+1)} \Psi_{k-2}(f'', \rho).
 \end{aligned}$$

Since f'' is uniformly bounded in $\mathcal{S}$, we have $|\Psi_k(f, \rho)| \leq \frac{K}{k(k+1)}$ uniformly for all ρ close to 1. Using the dominated convergence theorem and letting $\rho \rightarrow 1$, we then have

$$\Psi_k(f, \rho) \rightarrow \Psi(f), \quad \text{as } \rho \rightarrow 1,$$

which ensures (3.5).

3.7 Calibration

3.7.1 Calibration of the mean correction term in (3.7)

Theorem 4 provides a CLT for $G_n(f)$ under the general framework $p/n \rightarrow \infty$, where bounded fourth moment of the variable is enough to ensure the

Chapter 3. Central limit theorem for the linear spectral statistics

validness of the CLT. However, simulation results show that the CLT for $G_n(f)$ may be sensitive to the skewness and the kurtosis of the variable for some particular functions f , e.g. $f(x) = \frac{1}{2}x(x^2 - 3)$, especially for the mean correction term in (3.7). This phenomenon can be explained by the convergence rate of $EG_n(f)$ to zero. If we assume $EX_{11}^8 < \infty$, then by the arguments in Section 3.4, we would have $|EG_n(f)| = O(\sqrt{n/p}) + O(1/\sqrt{n})$. Therefore, the terms (e.g. S_{22} below (3.69)) with coefficient $(\nu_4 - 1)\sqrt{n/p}$ tend to zero theoretically, but may cause inaccuracy in finite sample simulation. For example, if $n = 100, p = n^2$ and the variables X_{ij} are from normalized $\exp(1)$, $(\nu_4 - 1)\sqrt{n/p}$ could be 0.8. In this section, we will pick up such terms and give a calibration of the mean correction term in (3.2).

From Section 3.4, we observe that the convergence rate of $|EG_n(f)|$ relies on the rate of $|nE\omega_n - m^3(z)(m'(z) + \nu_4 - 2)|$ in Lemma 1. By the arguments in Section 3.4, only S_{22} below (3.69) has the coefficient $(\nu_4 - 1)\sqrt{n/p}$. A simply calculation implies that

$$S_{22} = -\frac{2}{p\sqrt{np}}E(X_{11}^2 - 1)X_{11}^2 Etr\mathbf{M}_k^{(1)} = -2(\nu_4 - 1)\sqrt{n/pm}(z) + o(1).$$

Hence, the limit of $nE\omega_n$ is calibrated as

$$nE\omega_n = m^3(z) \left[\nu_4 - 2 + m'(z) - 2(\nu_4 - 1)\sqrt{n/pm}(z) \right] + o(1). \quad (3.82)$$

We then calibrate $G_n(f)$ as

$$\begin{aligned} G_n^{Calib}(f) &\triangleq n \int_{-\infty}^{+\infty} f(x) d(F^{\mathbf{B}}(x) - F(x)) \\ &\quad - \frac{n}{2\pi i} \oint_{|m|=\rho} f(-m - m^{-1}) \mathcal{X}_n^{Calib}(m) \frac{1 - m^2}{m^2} dm. \end{aligned} \quad (3.83)$$

where according to (3.82)

$$\mathcal{X}_n^{Calib}(m) \triangleq \frac{-\mathcal{B} + \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}^{Calib}}}{2\mathcal{A}},$$

$$\mathcal{C}^{Calib} = \frac{m^3}{n} \left[\nu_4 - 2 + \frac{m^2}{1 - m^2} - 2(\nu_4 - 1)\sqrt{n/pm} \right] - \sqrt{\frac{n}{p}}m^4,$$

$\mathcal{A}, \mathcal{B}$ are defined in (3.8) and $\sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}^{Calib}}$ is the complex number whose imaginary part has the same sign as the imaginary part of the $\mathcal{B}$. Theorem 4 still holds if we replace $G_n(f)$ with $G_n^{Calib}(f)$.

3.7.2 Simulations for checking the accuracy of CLT

We provide simulations to check the accuracy of the CLT in Theorem 4 with $G_n(f)$ replaced by the calibrated expression $G_n^{Calib}(f)$ in (3.83). We pay particular attention to the mean correlation term in (3.83). Two combinations of (p, n) , $p = n^2, n^{2.5}$, and the test function $f(x) = \frac{1}{2}x(x^2 - 3)$ are considered in the simulation. To inspect the impact of the skewness and the kurtosis of the variables, we use three types of random variables, $N(0, 1)$, normalized $exp(1)$ and normalized $t(6)$. The skewnesses of these variables are 0, 2 and 0 while the fourth moments of these variables are 3, 9 and 6, respectively. The empirical mean and empirical standard deviation from 1000 independent replications are shown in Table 3.1.

It is observed from Table 3.1 that both the empirical means and standard deviations for $N(0, 1)$ random variables are very accurate. The empirical means for normalized $exp(1)$ and normalized $t(6)$ also show their good accuracy. We note the standard deviations for normalized $exp(1)$ and normalized $t(6)$ random variables are not good when n is small (e.g. $n=50$). But it gradually tends to 1 as the sample size n increases.

QQplots are employed to illustrate accuracy of the normal approximation in Figure 3.1, 3.2 corresponding to the scenarios $p = n^2, p = n^{2.5}$,

respectively. In each figure, QQplots from left to right in the upper and lower panel corresponds to $n = 50, 100$ and $n = 150, 200$, respectively with random variables generated from $N(0, 1)$ (red), normalize $\exp(1)$ (green) and normalized $t(6)$ (blue). We observe the same thing that the normal approximation is very accurate for normal variables while the approximation is gradually better when n increases for normalized $\exp(1)$ and $t(6)$ variables.

3.8 Some Lemmas

Lemma 2. *Let $z \in \mathbb{C}_1$. Under Assumptions (b), (c), we have*

$$E|\gamma_{ks}|^2 \leq Kn^{-1}, \quad E|\gamma_{ks}|^4 \leq K\left(\frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np}\right), \quad (3.84)$$

$$E|\eta_k|^2 \leq Kn^{-1}, \quad E|\eta_k|^4 \leq K\frac{\delta^4}{n} + K\left(\frac{1}{n^2} + \frac{p}{n^2} + \frac{1}{np}\right). \quad (3.85)$$

Proof: From Lemma 5 in Pan and Zhou (2011), we obtain

$$\begin{aligned} E|\mathbf{s}_k^T \mathbf{H} \mathbf{s}_k - \text{tr} \mathbf{H}|^4 &\leq K \left(EX_{11}^4 \right)^2 E(\text{tr} \mathbf{H} \mathbf{H})^2 \\ &\leq KE \left(\text{tr} \mathbf{M}_k^{(s)} \overline{\mathbf{M}}_k^{(s)} \right)^2 \leq Kn^2 p^4, \end{aligned} \quad (3.86)$$

where $\mathbf{H} = \mathbf{M}_k^{(s)} - \text{diag}(a_{11}^{(s)}, \dots, a_{nn}^{(s)})$ and $a_{jj}^{(s)}$ is the j -th diagonal element of the matrix $\mathbf{M}_k^{(s)}$. To get the third inequality in (3.86), by (3.21) and the uniform bound for $\|\mathbf{D}_k^{-1}\|$, we obtain

$$\begin{aligned} |\text{tr} \mathbf{M}_k^{(s)} \overline{\mathbf{M}}_k^{(s)}| &= |\text{tr} \mathbf{D}_k^{-s} \mathbf{X}_k^T \mathbf{X}_k \overline{\mathbf{D}}_k^{-s} \mathbf{X}_k^T \mathbf{X}_k| \leq \frac{n}{v_1^{2(s-1)}} \|\mathbf{D}_k^{-1} \mathbf{X}_k^T \mathbf{X}_k\|^2 \\ &\leq \frac{n}{v_1^{2(s-1)}} \|p \mathbf{D}_k^{-1} + \sqrt{np}(\mathbf{I}_{n-1} + z \mathbf{D}_k^{-1})\|^2 \leq \frac{Kn^2 p^4}{v_1^{2s}}. \end{aligned} \quad (3.87)$$

Let $\mathbb{E}_j(\cdot) = E(\cdot | X_{1k}, X_{2k}, \dots, X_{jk})$, $j = 1, \dots, p$. Since $\{X_{jk}\}_{j=1}^k$ are independent of $a_{jj}^{(s)}$, $(X_{jk}^2 - 1)a_{jj}^{(s)} = (\mathbb{E}_j - \mathbb{E}_{j-1})(X_{jk}^2 - 1)a_{jj}^{(s)}$. By Burkholder's

inequality and Assumption (c)

$$\begin{aligned}
E \left| \sum_{j=1}^p (X_{jk}^2 - 1) a_{jj}^{(s)} \right|^4 &= E \left| \sum_{j=1}^p (\mathbb{E}_j - \mathbb{E}_{j-1}) (X_{jk}^2 - 1) a_{jj}^{(s)} \right|^4 \\
&\leq KE \left(\sum_{j=1}^n E |X_{11}|^4 |a_{jj}^{(s)}|^2 \right)^2 + K \sum_{j=1}^p E |X_{11}|^8 E |a_{jj}|^4 \\
&\leq Kn^5 p^2 + n^3 p^3,
\end{aligned} \tag{3.88}$$

where we use the fact that, with $\mathbf{w}_j^T$ being the j -th row of $\mathbf{X}_k$,

$$\begin{aligned}
E |a_{jj}^{(s)}|^4 &= E |\ddot{\mathbf{e}}_j^T \mathbf{X}_k \mathbf{D}_k^{-s} \mathbf{X}_k^T \ddot{\mathbf{e}}_j|^4 = E |\mathbf{w}_j^T \mathbf{D}_k^{-s} \mathbf{w}_j|^4 \\
&\leq v_1^{-4s} E \|\mathbf{w}_j^T\|^8 \leq Kn^4 + Kn^2 p.
\end{aligned} \tag{3.89}$$

Here for $j = 1, \dots, p$, $\ddot{\mathbf{e}}_j$ denotes the p -dimensional unit vector with the j -th element being 1 and all the remaining being zero. It follows from (3.86) and (3.88) that

$$\begin{aligned}
E |\gamma_{ks}|^4 &\leq \frac{K}{n^4 p^4} E \left| \sum_{j=1}^p (X_{jk}^2 - 1) a_{jj}^{(s)} \right|^4 + \frac{K}{n^4 p^4} E |\mathbf{s}_k^T \mathbf{H} \mathbf{s}_k - \text{tr} \mathbf{H}|^4 \\
&\leq K \left(\frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np} \right).
\end{aligned}$$

Moreover, applying Lemma 8.10 in Bai and Silverstein (2006) we have

$$\begin{aligned}
E |\eta_k|^4 &\leq \frac{K}{n^2 p^2} E |\mathbf{s}_k^T \mathbf{s}_k - n|^4 + K E |\gamma_{k1}(z)|^4 \\
&\leq K \frac{\delta^4}{p} + K \left(\frac{1}{n^2} + \frac{p}{n^2} + \frac{1}{np} \right).
\end{aligned}$$

The bounds of the absolute second moments for γ_{ks}, η_k follow from a direct application of Lemma 8.10 in Bai and Silverstein (2006), (3.21) and the uniform bound for $\|\mathbf{D}_k^{-1}\|$.

□

Chapter 3. Central limit theorem for the linear spectral statistics

In the following, we will prove some primary results in the purpose of proving Lemma 1 where $z \in \varsigma_n$. One trouble for $z \in \varsigma_l \cup \varsigma_r$ is the unbounded spectral norm of $\mathbf{D}^{-1}(z)$ which means the bound (3.22) or Lemma 2 can not be employed directly. Therefore we re-establish these facts below.

Let the event $U_n = \{\max_{j \leq n} |\lambda_j(\mathbf{B})| \geq u_0/2 + 1\}$ and $U_{nk} = \{\max_{j \leq n} |\lambda_j(\mathbf{B}_k)| \geq 1 + u_0/2\}$. The Cauchy interlacing theorem ensures that

$$\lambda_1(\mathbf{B}) \geq \lambda_1(\mathbf{B}_k) \geq \lambda_2(\mathbf{B}) \geq \lambda_2(\mathbf{B}_k) \geq \cdots \geq \lambda_{n-1}(\mathbf{B}_k) \geq \lambda_n(\mathbf{B}). \quad (3.90)$$

Thus $U_{nk} \subset U_n$. By (3.11) for any $\ell > 0$

$$P(U_{nk}) \leq P(U_n) = o(n^{-\ell}). \quad (3.91)$$

We claim that

$$\max\{\|\mathbf{D}^{-1}(z)\|, \|\mathbf{D}^{-1}(z)\|, |\beta_k|\} \leq \xi_n^{-1}n; \quad (3.92)$$

$$\frac{I(U_n^c)}{|\lambda_j(\mathbf{B}) - z|} \leq K, j = 1, 2, \dots, n, \quad \frac{I(U_{nk}^c)}{|\lambda_j(\mathbf{B}_k) - z|} \leq K, i = 1, 2, \dots, (n-1); \quad (3.93)$$

$$\|\mathbf{D}^{-1}(z)\|I(U_n^c) \leq 2/(u_0 - 2), \quad \|\mathbf{D}_k^{-1}(z)\|I(U_{nk}^c) \leq 2/(u_0 - 2). \quad (3.94)$$

Indeed, the quantities in (3.92) are bounded $|1/\text{Im}(z)| \leq \xi_n^{-1}n$ while (3.93) holds because $I(U_n^c)/|\lambda_j(\mathbf{B}) - z|$ (or $I(U_n^c)/|\lambda_j(\mathbf{B}_k) - z|$) is bounded by v_1^{-1} when $z \in \varsigma_u$ and bounded by $2/(u_0 - 2)$ when $z \in \varsigma_l \cup \varsigma_r$. The estimates in (3.94) hold because of the eigenvalues of $\mathbf{D}^{-1}I(U_n^c)$ (or $\mathbf{D}^{-1}I(U_{nk}^c)$) having the form $I(U_n^c)/(\lambda_j(\mathbf{B}) - z)$ (or $I(U_n^c)/(\lambda_j(\mathbf{B}_k) - z)$).

Lemma 3. *Let $z \in \varsigma_n$. The following bound*

$$|\beta_k|I(U_n^c) \leq K, \quad (3.95)$$

holds.

Proof: In view of (3.18), to prove (3.95), we need to find an upper bound for $\left| \text{tr} \mathbf{D}^{-1} - \text{tr} \mathbf{D}_k^{-1} \right| I(U_n^c)$ and a lower bound for $|1 + \mathbf{q}_k^T \mathbf{D}_k^{-2} \mathbf{q}_k| I(U_n^c)$. It follows from (3.93) and (3.90) that

$$\begin{aligned}
& \left| \text{tr} \mathbf{D}^{-1} - \text{tr} \mathbf{D}_k^{-1} \right| I(U_n^c) \\
& \leq \left| \sum_{j=1}^n \frac{1}{\lambda_j(\mathbf{B}) - z} - \sum_{j=1}^{n-1} \frac{1}{\lambda_j(\mathbf{B}_k) - z} \right| I(U_n^c) \\
& \leq \left(\sum_{j=1}^{n-1} \frac{|\lambda_j(\mathbf{B}) - \lambda_j(\mathbf{B}_k)|}{|\lambda_j(\mathbf{B}) - z| |\lambda_j(\mathbf{B}_k) - z|} + \frac{1}{|\lambda_n(\mathbf{B}) - z|} \right) I(U_n^c) \quad (3.96) \\
& \leq K \left(\sum_{j=1}^{n-1} (\lambda_j(\mathbf{B}) - \lambda_j(\mathbf{B}_k)) + 1 \right) I(U_n^c) \\
& \leq K(\lambda_1(\mathbf{B}) - \lambda_n(\mathbf{B}) + 1) I(U_n^c) \leq K(u_0 + 3).
\end{aligned}$$

Let $\mathbf{u}_j(\mathbf{B}_k), j = 1, \dots, n-1$ be the eigenvectors corresponding to the eigenvalues $\lambda_j(\mathbf{B}_k), j = 1, \dots, n-1$. Then $\sum_{j=1}^{n-1} \frac{\mathbf{u}_j(\mathbf{B}_k) \mathbf{u}_j^T(\mathbf{B}_k)}{(\lambda_j(\mathbf{B}_k) - z)^2}$ is the spectral decomposition of $\mathbf{D}_k^{-2}$. We distinguish two cases:

- i When $z \in V_1 = \varsigma_u \cup \{z : |\text{Im}(z)| > (u_0 - 2)/4\}$, via (3.22), we then obtain

$$|\beta_k| I(U_n^c) \leq 1/|\text{Im}(z)| \leq \max\{v_1^{-1}, 4/(u_0 - 2)\} \leq K.$$

Thus, (3.95) is true for $z \in V_1$.

- ii When $z \in V_2 = (\varsigma_l \cup \varsigma_r) \cap \{z : |\text{Im}(z)| < (u_0 - 2)/4\}$, if U_n^c happens, we have $|\lambda_j(\mathbf{B}_k) - \Re(z)| \geq \frac{u_0 - 2}{2}$ since $\Re(z) = \pm u_0$ for $z \in V_2$. A direct calculation shows

$$\begin{aligned}
& \Re\left((1 + \mathbf{q}_k^T \mathbf{D}_k^{-2} \mathbf{q}_k) I(U_n^c)\right) \\
& = 1 + \sum_{j=1}^{n-1} \frac{(\lambda_j(\mathbf{B}_k) - \Re(z))^2 - |\text{Im}(z)|^2}{|\lambda_j(\mathbf{B}_k) - z|^4} (\mathbf{q}_k^T \mathbf{u}_j(\mathbf{B}_k))^2 I(U_n^c) \\
& > 1.
\end{aligned}$$

Chapter 3. Central limit theorem for the linear spectral statistics

Therefore, $|1 + \mathbf{q}^T \mathbf{D}_k^{-2} \mathbf{q}| I(U_n^c)$ has a lower bound which, together with (3.96), implies (3.95) is true for $z \in V_2$.

Since $\varsigma_n = V_1 \cup V_2$, we finish the proof of Lemma 3.

□

Lemma 4. *Let $z \in \varsigma_n$. The following bounds hold*

$$E|\bar{\mu}_k|^4 \leq K \frac{\delta^4}{n} + K \left(\frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np} \right), \quad (3.97)$$

and

$$|E\bar{\mu}_k^3| = o(n^{-1}). \quad (3.98)$$

Proof: Write

$$\begin{aligned} \bar{\mu}_k &= \frac{1}{\sqrt{np}} (\mathbf{s}_k^T \mathbf{s}_k - p) - \gamma_{k1} + \left(1 + z \sqrt{\frac{p}{n}} \right) \left(\frac{1}{n} \text{tr} \mathbf{D}^{-1}(z) - \frac{1}{n} \text{tr} \mathbf{D}_k^{-1}(z) \right) \\ &\quad - \left(1 + z \sqrt{\frac{p}{n}} \right) \left(\frac{1}{n} \text{tr} \mathbf{D}^{-1}(z) - E \frac{1}{n} \text{tr} \mathbf{D}^{-1}(z) \right) + \frac{1}{\sqrt{np}} \\ &= L_1 - \gamma_{k1} + L_3 + L_4 + L_5. \end{aligned}$$

When the event U_n^c happens, reviewing the proof of the second result of (3.84) and via (3.94), we also have

$$E|\gamma_{ks}|^4 I(U_n^c) \leq K \left(\frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np} \right), m = 1, 2.$$

Moreover, by (3.91) and (3.92)

$$E|\gamma_{ks}|^4 I(U_n) = o(n^{-\ell}).$$

It follows that

$$E|\gamma_{ks}|^4 \leq K \left(\frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np} \right), m = 1, 2. \quad (3.99)$$

Using Lemma 8.10 in Bai and Silverstein (2006), (3.91), (3.92) and (3.96) we then have

$$E|L_1|^4 \leq K\delta^4 n^{-1}, \quad E|L_3|^4 \leq Kn^{-4}, \quad E|L_5|^4 \leq Kn^{-2}p^{-2}. \quad (3.100)$$

As for L_4 , by Burkholder's inequality, (3.19) and (3.96), we have

$$\begin{aligned} E|L_4|^4 &\leq Kn^{-4} E \left| \sum_{k=1}^n (E_k - E_{k-1})(tr\mathbf{D}^{-1} - tr\mathbf{D}_k^{-1}) \right|^4 \\ &\leq Kn^{-4} \sum_{k=1}^n E \left| tr\mathbf{D}^{-1}(z) - tr\mathbf{D}_k^{-1}(z) \right|^4 \\ &\quad + Kn^{-1} E \left(\sum_{k=1}^n E_k \left| tr\mathbf{D}^{-1}(z) - tr\mathbf{D}_k^{-1}(z) \right|^2 \right)^2 \\ &\leq Kn^{-4} \sum_{k=1}^n E \left| tr\mathbf{D}^{-1}(z) - tr\mathbf{D}_k^{-1}(z) \right|^4 I(U_n^c) \\ &\quad + Kn^{-4} E \left(\sum_{k=1}^n E_k \left| tr\mathbf{D}^{-1}(z) - tr\mathbf{D}_k^{-1}(z) \right|^2 \right)^2 I(U_n^c) + o(n^{-\ell}) \\ &\leq Kn^{-2}. \end{aligned} \quad (3.101)$$

Therefore, the proof of (3.97) is complete. Also, the analysis above yields

$$\begin{aligned} E|L_1 - \gamma_{k1}|^4 &\leq K \left(\frac{\delta^2}{n} + \frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np} \right) \leq K\delta^2 n^{-1}, \\ E|L_3 + L_4 + L_5|^4 &\leq Kn^{-2}. \end{aligned} \quad (3.102)$$

It is also easy to verify that, for $z \in \varsigma_n$,

$$E \left| \frac{1}{\sqrt{np}} (\mathbf{s}_k^T \mathbf{s}_k - p) \right|^2 \leq Kn^{-1}, \quad E|\gamma_{km}|^2 \leq Kn^{-1}. \quad (3.103)$$

We proceed to prove (3.98). First of all

$$|EL_1^3| = \left| E \left(\sum_{j=1}^p (X_{jk}^2 - 1) \right)^3 \right| = \sum_{j=1}^p E(X_{jk}^2 - 1)^3 \leq K\delta^2/\sqrt{np}. \quad (3.104)$$

Chapter 3. Central limit theorem for the linear spectral statistics

For $s = 1, 2$, denoting $\mathbf{M}_k^{(s)} = (a_{ij}^{(s)})_{p \times p}$, we then have

$$\begin{aligned} E\gamma_{ks}^3 &= \frac{1}{n^3 p^3} E \left(\sum_{i \neq j} X_{ik} X_{jk} a_{ij}^{(s)} + \sum_{i=1}^n (X_{ik}^2 - 1) a_{ii}^{(s)} \right)^3 \\ &= J_1 + J_2 + J_3 + J_4, \end{aligned}$$

where

$$\begin{aligned} J_1 &= \frac{1}{n^3 p^3} E \left(\sum_{i \neq j, j \neq t, t \neq i} X_{ik}^2 X_{jk}^2 X_{tk}^2 a_{ij}^{(s)} a_{jt}^{(s)} a_{ti}^{(s)} \right) \\ &\quad + \frac{4}{n^3 p^3} E \left(\sum_{i \neq j} X_{ik}^3 X_{jk}^3 (a_{ij}^{(s)})^3 \right) \\ &\triangleq J_{11} + J_{12}, \\ J_2 &= \frac{1}{n^3 p^3} E \left(\sum_{i=1}^p (X_{ik}^2 - 1)^3 (a_{ii}^{(s)})^3 \right), \\ J_3 &= 3 \frac{1}{n^3 p^3} E \left(\sum_{i \neq j} X_{ik} (X_{ik}^2 - 1) X_{jk} (X_{jk}^2 - 1) a_{ij}^{(s)} a_{ii}^{(s)} a_{jj}^{(s)} \right), \\ J_4 &= 3 \frac{2}{n^3 p^3} E \left(\sum_{i \neq j} X_{ik}^2 (X_{ik}^2 - 1) X_{jk}^2 a_{ij}^{(s)} a_{ii}^{(s)} a_{ji}^{(s)} \right). \end{aligned}$$

The inequality (3.89) can be extended to the range $z \in \varsigma_n$ by a similar method as that in (3.99). Therefore,

$$\begin{aligned} |J_2| &\leq K \frac{1}{n^3 p^3} p \delta^2 \sqrt{np} (n^4 + n^2 p)^{3/4} \leq K \delta^2 n^{-1}, \\ |J_3| &\leq K \frac{1}{n^3 p^3} p^2 E \|\mathbf{w}_i\|^3 E \|\mathbf{w}_j\|^3 + o(n^{-\ell}) \\ &\leq K p^{-1} + o(n^{-\ell}), \\ J_4 &\leq K p^{-1} + o(n^{-\ell}), \end{aligned}$$

where $\mathbf{w}_j^T$ is the j -row of $\mathbf{X}_k$.

Consider J_1 now. We first note that $J_{12} = O(p^{-1})$. Split J_{12} as

$$\begin{aligned} J_{12} &= \frac{1}{n^3 p^3} E \text{tr}(\mathbf{X}_k \mathbf{D}_k^{-s} \mathbf{X}_k^T)^3 - \frac{1}{n^3 p^3} E \sum_{i \neq t} a_{ii}^{(s)} a_{it}^{(s)} a_{ti}^{(s)} \\ &\quad + \frac{1}{n^3 p^3} E \sum_{i \neq j} a_{ij}^{(s)} a_{jj}^{(s)} a_{ji}^{(s)} + \frac{1}{n^3 p^3} E \sum_{i \neq j} a_{ij}^{(s)} a_{ji}^{(s)} a_{ii}^{(s)} \\ &\quad + \frac{1}{n^3 p^3} E \sum_{i=1}^p (a_{ii}^{(s)})^3 \\ &\leq K n^{-2} + K p^{-1}. \end{aligned}$$

Thus, we obtain

$$|E \gamma_{ks}^3| \leq K(\delta^2 n^{-1} + p^{-1}). \quad (3.105)$$

It follows from (3.102), (3.103) and (3.105) that

$$\begin{aligned} |E \bar{\mu}_k^3| &\leq |E(L_1 - \gamma_{k1})^3| + |E(L_3 + L_4 + L_5)^3| + 3|E(L_1 - \gamma_{k1})(L_3 + L_4 + L_5)^2| \\ &\quad + 3|E(L_1 - \gamma_{k1})^2(L_3 + L_4 + L_5)| \\ &\leq |E L_1^3| + |E \gamma_{k3}^3| + 3E^{1/2} E L_1^4 \cdot E^{1/2} \gamma_{ks}^2 \\ &\quad + 3E^{1/2} L_1^2 \cdot E^{1/2} \gamma_{ks}^4 + K n^{-3/2} + K \delta n^{-1} \\ &= o(n^{-1}). \end{aligned}$$

The proof of Lemma 4 is completed. □

The following lemma will be used to prove the first result of (3.14) and (3.71).

Lemma 5. *For $z \in \varsigma_n$ we have*

$$E|M_n^{(1)}(z)| \leq K,$$

where $M_n^{(1)}(z) = n(m_n(z) - E m_n(z))$.

Chapter 3. Central limit theorem for the linear spectral statistics

Proof: Note that the expression $M_n^{(1)}(z)$ in (3.19) may not be suitable for $z \in \varsigma_n$, since β_k^{tr} or even $\beta_k^{tr} I(U_n^c)$ may be not bounded. For this reason, we introduce the following notations with the purpose to obtain a similar expression as (3.19). Let

$$\begin{aligned}\acute{\epsilon}_k &= \frac{1}{z + \frac{1}{np} Etr \mathbf{M}_k^{(1)}}, \\ \acute{\mu}_k &= \frac{1}{\sqrt{np}} (\mathbf{s}_k^T \mathbf{s}_k - p) - \gamma_{k1} - \left(\frac{1}{np} tr \mathbf{M}_k^{(1)} - \frac{1}{np} Etr \mathbf{M}_k^{(1)} \right).\end{aligned}$$

Hence,

$$\beta_k = \epsilon_k + \beta_k \acute{\epsilon}_k \acute{\mu}_k. \quad (3.106)$$

As in (3.19) and a few lines below it, by (3.106), we write

$$M^{(1)}(z) = \sum_{k=1}^n (E_k - E_{k-1}) (\acute{\iota}_{k1} + \acute{\iota}_{k2} + \acute{\iota}_{k3} + \acute{\kappa}_k),$$

where

$$\begin{aligned}\acute{\iota}_{k1}(z) &= - \left(1 + \frac{1}{np} tr \mathbf{M}_k^{(2)} \right) (\acute{\epsilon}_k)^2 \acute{\mu}_k, \quad \acute{\iota}_{k2}(z) = -\gamma_{k1} (\acute{\epsilon}_k)^2 \acute{\mu}_k, \\ \acute{\iota}_{k3}(z) &= - \left(1 + \frac{1}{np} \mathbf{q}_k^T \mathbf{D}_k^{-2}(z) \mathbf{q}_k \right) \beta_k (\acute{\epsilon}_k)^2 \acute{\mu}_k^2, \quad \acute{\kappa}_k = \acute{\epsilon}_k \gamma_{k2}(z).\end{aligned}$$

We next derive the bounds for $\acute{\epsilon}_k$ and the forth moment of $\acute{\mu}_k$. By (3.21), (3.96) and (3.62), we then have

$$\begin{aligned}E \frac{1}{np} tr \mathbf{M}_k^{(1)} &= E \left[\left(1 + z \sqrt{\frac{n}{p}} m_n(z) - \left(1 + z \sqrt{\frac{n}{p}} \right) \frac{1}{n} (tr \mathbf{D}^{-1} - tr \mathbf{D}_k^{-1}) + \frac{n-1}{\sqrt{np}} \right] \right. \\ &\quad \left. \rightarrow m(z). \right.\end{aligned}$$

Hence

$$|\acute{\epsilon}_k| = \left| \frac{1}{z + m(z) + o(1)} \right| \leq \left| \frac{2}{z + m(z)} \right| \leq 2. \quad (3.107)$$

On the other hand, via (3.21), (3.96) and (3.101)

$$E \left| \frac{1}{np} tr \mathbf{M}_k^{(1)} - E \frac{1}{np} tr \mathbf{M}_k^{(1)} \right|^4 \leq (1 + z \sqrt{\frac{n}{p}})^4 n^{-4} E |tr \mathbf{D}^{-1} - Etr \mathbf{D}^{-1}|^4 \leq K n^{-2},$$

and this, together with (3.99), implies

$$E|\dot{\mu}_k|^4 \leq K \frac{\delta^4}{n} + K \left(\frac{1}{n^2} + \frac{n}{p^2} + \frac{1}{np} \right). \quad (3.108)$$

Combining (3.107), (3.108), Lemma 3, together with (3.91), (3.92), (3.94) and Burkholder's inequality, we obtain

$$E|M_n^{(1)}(z)|^2 \leq K.$$

The proof of the lemma is complete.

□

Table 3.1: *Empirical mean and standard deviation*

$p = n^2$				
n	50	100	150	200
$N(0, 1)$	(-0.314, 1.227)	(-0.221, 1.038)	(-0.188, 1.051)	(-0.093, 0.940)
$exp(1)$	(-0.088, 2.476)	(-0.079, 1.447)	(-0.140, 1.400)	(-0.161, 1.154)
$t(6)$	(-0.084, 2.813)	(-0.077, 1.541)	(-0.095, 1.246)	(-0.0897, 1.104)
$p = n^{2.5}$				
n	50	100	150	200
$N(0, 1)$	(-0.068, 1.049)	(-0.053, 1.077)	(-0.0476, 0.944)	(-0.016, 1.045)
$exp(1)$	(-0.049, 1.879)	(-0.029, 1.390)	(-0.046, 1.162)	(-0.045, 1.156)
$t(6)$	(-0.075, 1.693)	(-0.050, 1.252)	(-0.044, 1.145)	(-0.027, 1.044)

Figure 3.1: Normal Q-Q plot with the sample generating from $N(0,1)$ (red) and normalized $\exp(1)$ (green). The sample size are set to $n = 50, 100, 150, 200$ corresponds to the figures from left to right while the dimension $p = n^2$.

Figure 3.2: Normal Q-Q plot with the sample generating from $N(0,1)$ (red) and normalized $\exp(1)$ (green). The sample size are set to $n = 50, 100, 150, 200$ corresponds to the figures from left to right while the dimension $p = n^{2.5}$.

Chapter 4

Application of CLTs to Hypothesis test

4.1 Simulation Study

We consider an application of CLTs in this section. Suppose that $\mathbf{y} = \Gamma \mathbf{s}$ is a p -dimensional vector with covariance matrix $\Sigma = \Gamma \Gamma^T$ with Γ being a $p \times p$ matrix whose eigenvalues are positive and the entries of $\mathbf{s}$ being i.i.d random variables with mean zero and variance one. We want to test

$$H_0 : \quad \Sigma = \mathbf{I}_p, \quad H_1 : \quad \Sigma \neq \mathbf{I}_p. \quad (4.1)$$

Based on the i.i.d. samples $\mathbf{y}_1, \dots, \mathbf{y}_n$ (from $\mathbf{y}$), many methods have been introduced to test (4.1) in terms of the relationship of p and n . For example, John (1971) and Nagao (1973) considered (4.1) for fixed dimension p , Ledoit and Wolf (2002), Fisher, Sun and Gallagher (2005) and Bai et al. (2009) considered (4.1) for $\frac{p}{n} \rightarrow c \in (0, \infty)$, and Srivastava (2005), Srivastava, Kollob and Rosenc (2011), Fisher (2012) and Chen, Zhang and Zhong (2010) proposed a test which can accommodate large p and small n .

We are interested in testing (4.1) in the setting $\frac{p}{n} \rightarrow \infty$, for large p and small n . As in Ledoit and Wolf (2002) and Birke and Dette (2005), we take $f = x^2$ in (3.2) or (3.7). By Theorem 3 or Theorem 4. We then propose the test statistic as follows:

$$\begin{aligned} L_n &= \frac{1}{2} \left[n \left(\int x^2 dF^{\mathbf{R}}(x) - \int x^2 dF(x) \right) - (\nu_4 - 2) \right] \\ &= \frac{1}{2} \left(\text{tr} \mathbf{R} \mathbf{R}^T - n - (\nu_4 - 2) \right), \end{aligned} \quad (4.2)$$

where $\mathbf{R} = \sqrt{\frac{p}{n}} \left(\frac{1}{p} \mathbf{Y}^T \mathbf{Y} - \mathbf{I}_n \right)$ and $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$. Since $\Gamma^T \Gamma = \mathbf{I}_p$ is equivalent to $\Gamma \Gamma^T = \mathbf{I}_p$, under null hypothesis H_0 , we have

$$L_n \xrightarrow{d} N(0, 1). \quad (4.3)$$

The numerical performance of the proposed statistic L_n is carried out by Monte Carlo simulations. Let $Z_{\alpha/2}$ and $Z_{1-\alpha/2}$, respectively, be the $100\alpha/2\%$ and $100(1 - \alpha/2)\%$ quantiles of the asymptotic null distribution of the test statistic L_n . With T replications of the data set simulated under the null hypothesis, we calculate the empirical size as

$$\hat{\alpha} = \frac{\{\# \mid L_n^{\text{null}} \leq Z_{\alpha/2}\} + \{\# \mid L_n^{\text{null}} > Z_{1-\alpha/2}\}}{T},$$

where $\#$ denotes the number and L_n^{null} represents the values of the test statistic L_n based on the data set simulated under the null hypothesis. The empirical power is calculated by

$$\hat{\beta} = \frac{\{\# \mid L_n^{\text{alter}} \leq Z_{\alpha/2}\} + \{\# \mid L_n^{\text{alter}} > Z_{1-\alpha/2}\}}{T},$$

where L_n^{alter} represents the values of the test statistic L_n based on the data set simulated under the alternative hypothesis. In our simulation, we fix $T = 1000$ as the number of replications and set the nominal significant level $\alpha = 5\%$. By asymptotic normality, we have $Z_{\alpha/2} = -1.96$ and

$$Z_{1-\alpha/2} = 1.96.$$

Our proposed test is intended for the situation “large p , small n ”. To inspect the impact caused by sample size and/ or dimension, we set

$$n = 20, 40, 60, 80,$$

$$p = 600, 1500, 3000, 5500, 8000, 10000.$$

The entries of $\mathbf{s}$ are generated from three types of distributions, Gaussian distribution, standardized Gamma(4,0.5) and Bernoulli distribution with $P(X_{ij} = \pm 1) = 0.5$.

The following two types of covariance matrices are considered in the simulations to investigate the empirical power of the test.

1. (Diagonal covariance). $\Sigma = \text{diag}(\sqrt{2}\mathbf{1}_{[\nu p]}, \mathbf{1}_{1-[\nu p]})$, where $\nu = 0.08$ or $\nu = 0.25$, $[a]$ denotes the largest integer that is not greater than a .
2. (Banded covariance). $\Sigma = \text{diag}\left(A_1, \text{diag}(\mathbf{1}_{p-[v_2 p]})\right)$, where A_1 is a $[v_2 p] \times [v_2 p]$ tridiagonal symmetric matrix with the diagonal elements equal to 1 and elements below and above the diagonal all equal to v_1 .

Since the performance of the test in Chen, Zhang and Zhong (2010) accommodate a wider class of variates and the less restriction of the ratio p/n , we below compare performance of our test with that of Chen, Zhang and Zhong (2010). To simplify notation, denote their test by the CZZ test. Table 4.1 reports empirical sizes of the proposed test and of the CZZ test for the preceding three distributions. We observe from Table 4.1 that the sizes of both tests are roughly the same, when the underlying variables are normally or bernoulli distributed when p increases. This is expected as our

result based on $p/n \rightarrow \infty$. When n, p are large enough the size of the test for Gamma variables is also around 5%.

Table 4.2 to Table 4.4 summarize the empirical power of the proposed tests as well as those of the CZZ test for both the diagonal and banded covariance matrix. Table 4.2 assumes the underlying variables are normally distributed while Table 4.3 and 4.4 assume the central gamma and the central bernoulli distributed underlining variables, respectively. For the diagonal covariance matrix, we observe that the proposed test consistently outperforms the CZZ test for all types of distributions, especially for “small” n . For example, when $n = 20$, even $n = 40, 60, 80$ for $\nu = 0.08$, the CZZ test results in power ranging from 0.2-0.8, while our test still gains very satisfying powers exceeding 0.932.

For the banded covariance matrix, we observe an interesting phenomenon. Our test seems to be more sensitive to the dimension p . When $p = 600, 1500, 3000$, the power of our test is not good for small $v_2 (=0.4)$. Fortunately, when $p = 5500, 8000, 10000$, the performance is much better, where the power is one or close to one. Similar results are also observed for $v_2 = 0.8$. We also note that large v_2 outperforms smaller v_2 because when v_2 becomes larger, the corresponding covariance matrix becomes more “different” from the identity matrix. As for the CZZ test, its power is mainly affected by n . But generally speaking, our test gains better power than the CZZ test for extremely larger p and small n .

Table 4.1: *Empirical sizes of CZZ test and the proposed identity test at 5% significance level for normal, gamma, bernoulli random vectors.*

p	CZZ Test				Proposed Test			
	n				n			
	20	40	60	80	20	40	60	80
Normal random vectors								
600	0.069	0.071	0.052	0.052	0.063	0.077	0.066	0.082
1500	0.057	0.059	0.061	0.059	0.055	0.058	0.058	0.062
3000	0.067	0.068	0.057	0.053	0.048	0.067	0.056	0.052
5500	0.064	0.06	0.067	0.058	0.054	0.055	0.071	0.068
8000	0.071	0.062	0.062	0.054	0.055	0.049	0.06	0.059
10000	0.055	0.059	0.063	0.06	0.037	0.058	0.057	0.054
Gamma random vectors								
600	0.055	0.073	0.056	0.062	0.103	0.119	0.125	0.123
1500	0.064	0.047	0.059	0.059	0.094	0.072	0.072	0.088
3000	0.069	0.071	0.059	0.052	0.066	0.074	0.071	0.061
5500	0.065	0.069	0.048	0.041	0.077	0.073	0.047	0.045
8000	0.069	0.065	0.07	0.053	0.078	0.075	0.063	0.059
10000	0.072	0.06	0.06	0.057	0.078	0.082	0.065	0.06
Bernoulli random vectors								
600	0.078	0.079	0.056	0.037	0.048	0.064	0.046	0.037
1500	0.065	0.050	0.051	0.053	0.039	0.040	0.049	0.050
3000	0.048	0.053	0.058	0.060	0.040	0.052	0.052	0.056
5500	0.059	0.061	0.059	0.042	0.040	0.052	0.060	0.040
8000	0.065	0.074	0.065	0.059	0.046	0.052	0.05	0.051
10000	0.07	0.057	0.047	0.048	0.044	0.037	0.038	0.047

Table 4.2: Empirical power of CZZ test and the proposed identity test at 5% significance level for normal random vectors. Two types of population covariance matrices are considered. In the first case $\Sigma_1 = \text{diag}(2 \times \mathbf{1}_{[\nu p]}, \mathbf{1}_{p-[\nu p]})$ for $\nu = 0.08$ and $\nu = 0.25$ respectively. In the second case $\Sigma_2 = \text{diag}(A_1, \text{diag}(\mathbf{1}_{p-[\nu_2 p]}))$, where A_1 is a $[\nu_1 p] \times [\nu_1 p]$ tridiagonal symmetric matrix with diagonal elements equal to 1 and elements beside diagonal all equal to v_1 for $v_1 = 0.5, v_2 = 0.8$ and $v_1 = 0.5, v_2 = 0.4$ respectively.

p	CZZ Test				Proposed Test			
	n				n			
	20	40	60	80	20	40	60	80
Normal random vectors								
$\nu = 0.08$								
600	0.186	0.392	0.648	0.826	0.932	1	1	1
1500	0.179	0.397	0.642	0.822	0.999	1	1	1
3000	0.197	0.374	0.615	0.867	1.000	1	1	1
5500	0.225	0.382	0.615	0.85	1	1	1	1
8000	0.203	0.391	0.638	0.843	1	1	1	1
10000	0.204	0.381	0.639	0.835	1	1	1	1
Normal random vectors								
$\nu = 0.25$								
600	0.571	0.952	0.997	1	1	1	1	1
1500	0.585	0.959	1.000	1	1	1	1	1
3000	0.594	0.961	1.000	1	1	1	1	1
5500	0.617	0.954	1	1	1	1	1	1
8000	0.607	0.957	0.999	1	1	1	1	1
10000	0.595	0.949	1	1	1	1	1	1
Normal random vectors								
$v_1 = 0.5, v_2 = 0.8$								
600	0.333	0.874	0.997	1	0.443	0.493	0.492	0.488
1500	0.310	0.901	0.999	1	0.987	0.997	0.997	0.998
3000	0.348	0.889	0.998	1	1.000	1.000	1.000	1.000
5500	0.382	0.871	0.998	1	1	1	1	1
8000	0.33	0.867	0.998	1	1	1	1	1
10000	0.359	0.868	0.998	1	1	1	1	1
Normal random vectors								
$v_1 = 0.5, v_2 = 0.4$								
600	0.142	0.364	0.668	0.896	0.078	0.089	0.069	0.102
1500	0.131	0.354	0.653	0.890	0.220	0.235	0.230	0.226
3000	0.139	0.361	0.662	0.899	0.635	0.660	0.647	0.684
5500	0.148	0.352	0.645	0.898	0.97	0.979	0.989	0.989
8000	0.152	0.36	0.688	0.905	0.981	0.978	0.986	0.989
10000	0.137	0.328	0.674	0.886	1	1	1	1

Table 4.3: Empirical power of CZZ test and the proposed identity test at 5% significance level for standard gamma random vectors. Two types of population covariance matrices are considered. In the first case $\Sigma_1 = \text{diag}(2 \times \mathbf{1}_{[\nu p]}, \mathbf{1}_{p-[\nu p]})$ for $\nu = 0.08$ and $\nu = 0.25$ respectively. In the second case $\Sigma_2 = \text{diag}(A_1, \text{diag}(\mathbf{1}_{p-[\nu_2 p]}))$, where A_1 is a $[\nu_2 p] \times [\nu_2 p]$ tridiagonal symmetric matrix with diagonal elements equal to 1 and elements beside diagonal all equal to v_1 for $v_1 = 0.5, v_2 = 0.8$ and $v_1 = 0.5, v_2 = 0.4$ respectively.

p	CZZ Test				Proposed Test			
	n				n			
	20	40	60	80	20	40	60	80
Gamma random vectors $\nu = 0.08$								
600	0.331	0.638	0.891	0.982	0.999	1	1	1
1500	0.356	0.636	0.901	0.979	1	1	1	1
3000	0.197	0.383	0.638	0.823	1	1	1	1
5500	0.178	0.361	0.658	0.845	1	1	1	1
8000	0.199	0.399	0.642	0.85	1	1	1	1
10000	0.216	0.353	0.636	0.843	1	1	1	1
Gamma random vectors $\nu = 0.25$								
600	0.621	0.943	1.000	1	1	1	1	1
1500	0.610	0.946	0.999	1	1	1	1	1
3000	0.579	0.946	0.997	1	1	1	1	1
5500	0.596	0.957	0.999	1	1	1	1	1
8000	0.616	0.962	0.999	1	1	1	1	1
10000	0.614	0.955	0.999	1	1	1	1	1
Gamma random vectors $v_1 = 0.5, v_2 = 0.8$								
600	0.192	0.871	0.998	0.972	0.122	0.413	0.423	0.133
1500	0.198	0.883	0.995	0.980	0.440	0.992	0.993	0.433
3000	0.343	0.885	0.995	1	1	1	1	1
5500	0.342	0.88	0.996	1	1	1	1	1
8000	0.349	0.877	0.998	1	1	1	1	1
10000	0.337	0.879	0.998	1	1	1	1	1
Gamma random vectors $v_1 = 0.5, v_2 = 0.4$								
600	0.117	0.353	0.650	0.780	0.087	0.111	0.114	0.120
1500	0.138	0.365	0.661	0.799	0.183	0.215	0.226	0.157
3000	0.129	0.349	0.646	0.89	0.593	0.621	0.627	0.61
5500	0.124	0.335	0.678	0.889	0.945	0.972	0.981	0.986
8000	0.142	0.369	0.668	0.901	0.999	1	1	1
10000	0.142	0.336	0.668	1	1	1	1	1

Table 4.4: Empirical power of CZZ test and the proposed identity test at 5% significance level for standard bernoulli random vectors. Two types of population covariance matrices are considered. In the first case $\Sigma_1 = \text{diag}(2 \times \mathbf{1}_{[\nu p]}, \mathbf{1}_{p-[\nu p]})$ for $\nu = 0.08$ and $\nu = 0.25$ respectively. In the second case $\Sigma_2 = \text{diag}(A_1, \text{diag}(\mathbf{1}_{p-[\nu_2 p]}))$, where A_1 is a $[\nu_2 p] \times [\nu_2 p]$ tridiagonal symmetric matrix with diagonal elements equal to 1 and elements beside diagonal all equal to v_1 for $v_1 = 0.5, v_2 = 0.8$ and $v_1 = 0.5, v_2 = 0.4$ respectively.

p	CZZ Test				Proposed Test			
	n				n			
	20	40	60	80	20	40	60	80
Bernoulli random vectors								
$\nu = 0.08$								
600	0.216	0.381	0.622	0.849	0.972	1	1	1
1500	0.198	0.401	0.632	0.837	1	1	1	1
3000	0.203	0.362	0.622	0.823	1	1	1	1
5500	0.196	0.354	0.627	0.829	1	1	1	1
8000	0.203	0.373	0.638	0.834	1	1	1	1
10000	0.213	0.397	0.637	0.822	1	1	1	1
Bernoulli random vectors								
$\nu = 0.25$								
600	0.594	0.952	0.998	1	1	1	1	1
1500	0.619	0.960	1.000	1	1	1	1	1
3000	0.594	0.964	0.999	1	1	1	1	1
5500	0.609	0.948	1.000	1	1	1	1	1
8000	0.589	0.952	1	1	1	1	1	1
10000	0.603	0.957	0.999	1	1	1	1	1
Bernoulli random vectors								
$v_1 = 0.5, v_2 = 0.8$								
600	0.356	0.870	0.996	1	0.507	0.512	0.526	0.558
1500	0.359	0.892	0.995	1	0.999	1	1	0.999
3000	0.343	0.877	0.998	1	1.000	1	1	1.000
5500	0.355	0.868	0.997	1	1.000	1.000	1.000	1.000
8000	0.332	0.873	0.997	1	1	1	1	1
10000	0.353	0.872	1	1	1	1	1	1
Bernoulli random vectors								
$v_1 = 0.5, v_2 = 0.4$								
600	0.153	0.348	0.643	0.901	0.092	0.086	0.079	0.085
1500	0.154	0.372	0.643	0.878	0.239	0.255	0.235	0.241
3000	0.141	0.339	0.649	0.882	0.682	0.680	0.680	0.674
5500	0.156	0.343	0.656	0.893	0.997	0.994	0.994	0.994
8000	0.144	0.353	0.664	0.904	1	1	1	1
10000	0.139	0.356	0.685	0.889	1	1	1	1

Chapter 5

Conclusion and future research

5.1 Conclusion

This thesis comes with new results that deal with large dimensional sample covariance matrices where both dimensions go to infinity but one of them dominates the other. Under the setting $p/n \rightarrow \infty$, we consider the large dimensional sample covariance matrix $\mathbf{B} := \frac{1}{\sqrt{np}}(\mathbf{X}^T \mathbf{X} - p\mathbf{I}_n)$. The largest eigenvalue of $\mathbf{B}$ is showed to almost surely converge to 2, which is the right end point of the support of the limit spectral distribution (semicircle law) of $\mathbf{B}$. To prove this result, it suffices to show that $\limsup_{n \rightarrow \infty} \lambda_{\max}(\mathbf{B}) \leq 2$ a.s.. If we replace the diagonal elements of $\mathbf{B}$ by 0 and denote the new matrix as $\underline{\mathbf{B}}$, we show the largest eigenvalues of $\mathbf{B}$ and $\underline{\mathbf{B}}$ are equivalent. Then graph method is adopted to prove $\limsup_{n \rightarrow \infty} \lambda_{\max}(\underline{\mathbf{B}}) \leq 2$ a.s.. This result not only ensures the applicability of Helly-Bray theorem but is also important for establishing the CLT for linear spectral statistics of $\mathbf{B}$.

Many important statistics in multivariate analysis can be written into the linear functionals of the empirical spectral distribution (or linear spectral statistics) of some random matrices. The hypothesis testing is the first

purpose to study the CLT of the linear spectral statistics. Define

$$Q_n(f) \triangleq n \int f(x) d(F^{\mathbf{A}}(x) - F(x)) - \frac{1}{\pi} \sqrt{\frac{n^3}{p}} \int_{-1}^1 f(2x) \frac{4x^3 - 3x}{\sqrt{1 - x^2}} dx,$$

where f is a analytic function. This linear functional is different from that in Bai and Silverstein (2004) and Bai and Yao (2005) as there is a extra mean correction term. We showed that the finite dimensional random vector $(Q_n(f_1), \dots, Q_n(f_k))$ converges to a Gaussian vector whose explicit mean and covariance are obtained. The strategy is to convert the CLT of $Q_n(f)$ to the CLT of Stieltjes transform of the ESD of $\mathbf{B}$ by Cauchy's integral formula. To prove CLT, we apply the central limit theorem for the martingales. We also provide an application of this result: testing whether the covariance matrix is identity or not. Simulation results reveal that the powers of our proposed test are better than the powers in Chen, Zhang and Zhong (2010) when the sample size n is very small, say $n = 40$. Besides, this asymptotic theory complements the results of Bai and Silverstein (2004) for the case $p/n \rightarrow c \in (0, \infty)$ and Bai and Yao (2005) for Wigner matrix.

5.2 Future research

In the study of the eigenvalues of the large normalized sample covariance matrix when $p/n \rightarrow \infty$, we assume that the underlying variables 1) are mean 0; 2) are independent and identically distributed. In real applications, these two assumptions are rarely satisfied. Hence, our future research works are intended to release these assumptions.

5.2.1 With sample mean

Suppose $\mathbf{s}_1, \dots, \mathbf{s}_n$ are the i.i.d. p -dimensional vectors with covariance matrix $\mathbf{I}_p$ but mean nonzero. It is known that the normal sample covariance matrix with sample mean in multivariate analysis is defined as

$$\mathcal{S} = \frac{1}{n} \sum_{k=1}^n (\mathbf{s}_k - \bar{\mathbf{s}})(\mathbf{s}_k - \bar{\mathbf{s}})^T,$$

where $\bar{\mathbf{s}} = \frac{1}{n} \sum_{k=1}^n \mathbf{s}_k$ is the sample mean. The common used sample covariance matrix in random matrix theory is

$$\mathbf{S} = \frac{1}{n} \sum_{k=1}^n \mathbf{s}_k \mathbf{s}_k^T.$$

A nature question is that whether the asymptotic results for eigenvalues of $\mathbf{S}$ apply for $\mathcal{S}$ as well. When $p/n \rightarrow c \in (0, \infty)$, the problem is solved by Pan (2013). It showed in Pan (2013) that the central limit theorem of linear spectral statistics of $\mathcal{S}$ and $\mathbf{S}$ have the same asymptotic variances but have different asymptotic means.

We consider the large normalized sample covariance matrix with sample mean being nonzero in the “large p , small n ” situation. Define

$$\mathcal{B} = \frac{1}{\sqrt{np}} \left((\mathbf{X} - \bar{\mathbf{s}} \mathbf{1}_n^T)^T (\mathbf{X} - \bar{\mathbf{s}} \mathbf{1}_n^T) - p \mathbf{I}_n \right)$$

where $\mathbf{X} = (\mathbf{s}_1, \dots, \mathbf{s}_n)$. In the future research work, we are interested in the CLT for the LSS of $\mathcal{B}$.

5.2.2 With covariance

The other assumption in this thesis is that the entries of the data matrix are independent and identically distributed. A concern may be that how can we obtain the theoretical power for the testing problem. So it is necessary

to consider a more general matrix which could incorporate the dependence structure of the data. This general matrix is defined by

$$\mathbb{B} = \frac{1}{\sqrt{np}}(\mathbf{X}^T \Sigma \mathbf{X} - tr \Sigma \cdot \mathbf{I}_n).$$

where Σ is the covariance matrix of the data. Spectral analysis of this general normalized sample covariance $\mathbb{B}$ may be more interesting, which is also one of our future research works.

Bibliography

- [1] Adamczak, R., Litvak, A., Pajor, A. and Tomczak-Jaegermann. N. (2010) Quantitative estimates of the convergence of the empirical covariance matrix in Log-concave ensembles. *J. Amer. Math. Soc.*, **23**, 535-561.
- [2] Anderson T.W. (2003) *An introduction to multivariate statistical analysis*. 3th edition. Wiley, New York.
- [3] Arnold L. (1967) On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.*, **20**, 262-268.
- [4] Arnold L. (1971) On Wigner's semicircle law for the eigenvalues of random matrices. *Z. Wahrsch. Verw. Gebiete.*, **19**, 191-198.
- [5] Bai, Z.D. (1993) Convergence rate of expected spectral distributions of large random matrices. Part I. Wigner matrices. *Ann. Probab.*, **2**, 625-648.
- [6] Bai, Z.D. (1999). Methodologies in spectral analysis of large-dimensional random matrices. a review. *Statist. Sinica*, **9**, 611-677.

-
- [7] Bai Z.D., Jiang, D.D., Yao J.F. and Zheng S.R. (2009) Corrections to LRT on Large Dimensional Covariance Matrix by RMT. *Ann. Stat.*, **37**, 3822-3840.
 - [8] Bai, Z.D. and Silverstein, J.W. (2006) *Spectral analysis of large dimensional random matrices*. Springer.
 - [9] Bai, Z.D. and Silverstein, J.W. (2004) CLT for Linear Spectral Statistics of Large-Dimensional Sample Covariance Matrices. *Ann. Probab.*, **32**, 553-605.
 - [10] Bai, Z.D., Silverstein, J.W. and Yin, Y.Q. (1988) A note on the largest eigenvalue of a large dimensional sample covariance matrix. *J. Multivariate Anal.*, **26**, 166-168.
 - [11] Bai, Z.D. and Yao, J.F. (2005) On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11(6)**, 1059-1092.
 - [12] Bai, Z.D. and Yin, Y.Q. (1988) Convergence to the semicircle law. *Ann. Probab.*, **16**, 863-875.
 - [13] Bai, Z.D. and Yin, Y.Q. (1993) limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.* **21**, 1275-1294.
 - [14] Birke, M. and Dette, H. (2005) A note on testing the covariance matrix for large dimension. *Statistics Probability Letters*, **74**, 281-289.
 - [15] Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley, New York.
 - [16] Billingsley, P. (1995) *Probability and Measure*. 3rd edition. Wiley, New York.

-
- [17] Chen, B.B. and Pan, G.M. (2012) Convergence of the largest eigenvalue of normalized sample covariance matrices when p and n both tend to infinity with their ratio converging to zero. *Bernoulli* , **18**, 1405-1420.
- [18] Costin, O. and Lebowitz, J. (1995) Gaussian fluctuations in random matrices. *Phys. Rev. Lett.*, **75**, 69-72.
- [19] Chen S.X., Zhang L.X. and Zhong P.S. (2010) Tests for High-dimensional Covariance Matrix. *J. Am. Statist. Ass.*, **105**, 810-819 .
- [20] Donoho, D.L. (2000) High-dimensional data analysis: The curses and blessings of dimensionality. *American Math. Society Lecture*, **Math. Challenges of the 21st Century**.
- [21] Fan, K. (1951) Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. Nat. Acad. Sci. U.S.A.* , **37**, 760-766.
- [22] Fan, J.Q. and Fan, Y.Y. (2008) High-dimensional classification using features annealed independence rules. *Ann.Stat.*, **36(6)**, 2605-2637.
- [23] Fisher, T.J. (2012) On testing for an identity covariance matrix when the dimensionality equals or exceeds the sample size. *Journal of Statistical Planning and Inference*, **1**, 312-326.
- [24] Fisher, T.J., Sun, X.Q. and Gallagher, C.M. (2010) A new test for sphericity of the covariance matrix for high dimensional data. *J. Multivariate Anal.*, **10**, 2554-2570.
- [25] Geman, S. (1980) A limit theorem for the norm of random matrices. *Ann. Probab.*, **8**, 252-261.

-
- [26] Grenander U. (1963) *Probabilities on Algebraic Structures*. John Wiley, New York-London.
- [27] John, S. (1971) Some Optimal Multivariate Tests. *Biometrika* , **59**, 123-127.
- [28] Jonsson, D. (1982) Some limit theorems of the eigenvalues of sample covariance matrix. *J. Multivariate Anal.*, **12**, 1-38.
- [29] Johansson, K. (1998) On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, **91**, 151-204.
- [30] Karoui, E.N. (2003) On the largest eigenvalue of Wishart matrices with identity covariance when n , p and p/n tend to infinity. *arXiv:math/0309355*.
- [31] Karoui, E.N. (2008) Operator norm consistent estimation of large dimensional sparse covariance matrices. *Ann. Stat.*, **6**, 2717-2756.
- [32] Ledoit, O. and Wolf, M. (2002) Some Hypothesis Tests for the Covariance Matrix When the Dimension Is Large Compare to the Sample Size. *Ann. Stat.*, **30**, 1081-1102.
- [33] Marčenko, V.A. and Pastur, L.A. (1967) Distribution for some sets of random matrices. *Math. USSR-Sb.*, **1**, 457-483.
- [34] Muirhead R.J. (1982) *Aspects of Multivariate Statistical Theory*. John Wiley & Sons.
- [35] Nagao, H. (1973) On Some Test Criteria for Covariance matrix. *Ann. Stat.*, **30**, 700-709.
- [36] Pan, G.M. (2013) Comparision between two types of large sample covariance matrices. *Annales Inst. Henri Poincaré*, Accepted.

-
- [37] Pan, G.M. and Gao, J.T. (2012) Asymptotic Theorey for Sample Covariance matrix under Cross-Sectional Dependence. Preprint.
 - [38] Pan, G.M., Gao, J.T. and Yang, Y.R. (2013) Testing Independence for a Large Number of High Dimensional Random Vectors. *Monash Econometrics and Business Statistics Working Papers*.
 - [39] Pan, G.M. and Zhou, W. (2008) Asymptotic distribution of signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.*, **18(3)**., 1232-1270.
 - [40] Pan, G.M. and Zhou, W. (2011) Central limit theorem for Hotelling's T^2 statistic under large dimension. *Ann. Appl. Probab.*, **21(5)**, 1860-1910.
 - [41] Silverstein, J.W. (1984) Some limit theorems on the eigenvectors of large dimensional sample covariance matrices. *J. Multivariate Anal.*, **15**, 295-324.
 - [42] Silverstein, J.W. (1984) The smallest eigenvalue of a large dimensional Wishart Matrix. *Ann. Probab.*, **13(4)**, 1364-1368.
 - [43] Silverstein, J.W. (1995) Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.*, **54**, 175-192.
 - [44] Silverstein, J.W. and Bai, Z.D. (1995) On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.*, **54**, 331-339.
 - [45] Srivastava, M. S. (2005) Some tests concerning the covariance matrix in high-dimensional data, *J. Japan Statist. Soc.*, **35**, 251-372.

-
- [46] Srivastava, M.S., Kollob, T. and Rosenc, D. von. (2011) Some tests for the covariance matrix with fewer observations than the dimension under non-normality. *J. Multivariate Anal.*, **6**, 1090-1103.
- [47] Titchmarsh, E. (1939) *The Theory of Functions*. 2nd edition. London: Oxford University Press.
- [48] Wigner, E.P. (1955) Characteristic vectors bordered matrices with infinite dimensions. *Ann. Math.*, **62**, 548-564.
- [49] Wigner, E.P. (1958) On the distributions of the roots for certain symmetric matrices. *Ann. Math.*, **67**, 325-327.
- [50] Yin, Y.Q. (1986). LSD's for a class of random matrices. *J. Multivariate Anal.*, **20**, 50-68.
- [51] 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.
- [52] 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. *Prob theory and its related fields.*, **20**, 50-68.
- [53] 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.