Universality of some models of random matrices and random processes

Dissertation

zur Erlangung des Doktorgrades

an der Fakultät für Mathematik

der Universität Bielefeld

von Alexey Naumov

November 2012

Gedruckt auf alterungsbeständigem Papier nach DIN ISO-9706.

Abstract

Many limit theorems in probability theory are universal in the sense that the limiting distribution of a sequence of random variables does not depend on the particular distribution of these random variables. This universality phenomenon motivates many theorems and conjectures in probability theory. For example the limiting distribution in the central limit theorem for the suitably normalized sums of independent random variables which satisfy some moment conditions are independent of the distribution of the summands.

In this thesis we establish universality-type results for two classes of random objects: random matrices and stochastic processes.

In the first part of the thesis we study ensembles of random matrices with dependent elements. We consider ensembles of random matrices $\mathbf{X}_n$ with independent vectors of entries $(X_{ij}, X_{ji})_{i \neq j}$. Under the assumption that $\max(\mathbb{E}X_{12}^4, \mathbb{E}X_{21}^4) < \infty$ it is proved that the empirical spectral distribution of eigenvalues converges in probability to the uniform distribution on an ellipse. The axes of the ellipse are determined by the correlation between X_{12} and X_{21}. This result is called Elliptic Law. Here the limit distribution is universal, that is it doesn't depend on the distribution of the matrix elements. These ensembles generalize ensembles of symmetric random matrices and ensembles of random matrices with independent entries.

We also generalize ensembles of random symmetric matrices and consider symmetric matrices $\mathbf{X}_n = \{X_{ij}\}_{i,j=1}^n$ with a random field type dependence, such that $\mathbb{E}X_{ij} = 0$, $\mathbb{E}X_{ij}^2 = \sigma_{ij}^2$, where σ_{ij} may be different numbers. Assuming that the average of the normalized sums of variances in each row converges to one and Lindeberg condition holds true we prove that the empirical spectral distribution of eigenvalues converges to Wigner's semicircle law.

In the second part of the thesis we study some classes of stochastic processes. For martingales with continuous parameter we provide very general sufficient conditions for the strong law of large numbers and prove analogs of the Kolmogorov and Brunk–Prokhrov strong laws of large numbers. For random processes with independent homogeneous increments we prove analogs of the Kolmogorov and Zygmund–Marcinkiewicz strong laws of large numbers. A new generalization of the Brunk–Prokhorov strong law of large numbers is given for martingales with discrete times. Along with the almost surely convergence, we also prove the convergence in average.

Acknowledgments

I want to thank all who have helped me with the dissertation. Especially I would like to thank Prof. Dr. Friedrich Götze, Prof. Dr. Alexander Tikhomirov, Prof. Dr. Vladimir Ulyanov, Prof. Dr. Viktor Kruglov and my family.

My research has been supported by the German Research Foundation (DFG) through the International Research Training Group IRTG 1132, CRC 701 "Spectral Structures and Topological Methods in Mathematics", Deutscher Akademischer Austauschdienst (DAAD) and Asset management company "Aton".

Contents

Chapter 1

Introduction

Many limit theorems in probability theory are universal in the sense that the limiting distribution of a sequence of random variables does not depend on the particular distribution of these random variables. This universality phenomenon motivates many theorems and conjectures in probability theory. For example let us consider a sequence of independent identically distributed random variables $X_i, i \geq 1$. Assume that $\mathbb{E}|X_1| < \infty$. Then

$$\lim_{n\to\infty} \frac{X_1 + ... + X_n - n\mathbb{E}X_1}{n} = 0 \text{ almost surely.} \tag{1.0.1}$$

This result is called the strong law of large numbers. If we additionally assume that $\mathbb{E}X_1^2 = \sigma^2 < \infty$ and normalize the sum $X_1 + ... + X_n$ by the factor $\sigma\sqrt{n}$ then the central limit theorem holds

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_1 + ... + X_n - n\mathbb{E}X_1}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{2\pi}\int_{-\infty}^{x} e^{-z^2/2}dz. \tag{1.0.2}$$

We see here explicitly that the right-hand sides of (1.0.1) and (1.0.2) are universal, independent of the distribution of X_i. Results (1.0.1) and (1.0.2) were first proved for independent Bernoulli random variables, $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 0) = 1/2$, and then extended to all distributions with finite first and second moment respectively. Of course, the strong law of large numbers and the central limit theorem are only one of many similar universality-type results now known in probability theory.

In this thesis we establish universality-type results for two classes of random objects: random matrices and stochastic processes.

In the first part of the thesis we study the limit theorems for the random matrices with dependent entries. We prove that the empirical spectral distribution converges to some limit and this limit does not depend on the particular distribution of the random matrix elements.

In the second part of the thesis we study stochastic processes and prove the strong law of large numbers for martingales. For martingales with continuous parameter we provide very general sufficient conditions for the strong law of large numbers and prove analogs of the famous strong law of large numbers. Along with the almost surely convergence we prove the convergence in average.

1.1 Universality in random matrix theory

The study of random matrices, and in particular the properties of their eigenvalues, has emerged from applications, first in data analysis and later from statistical models for heavy-nuclei atoms. Recently Random matrix theory (RMT) has found its numerous application in many other areas, for example, in numeric analysis, wireless communications, finance, biology. It also plays an important role in different areas of pure mathematics. Moreover, the technics used in the study of random matrices has its sources in other branches of mathematics.

In this work we will be mostly interested in the behavior of an empirical spectral distribution of a random matrix. In the sections below we define main objects and introduce different ensembles of random matrices. We also give a brief survey of the most useful methods to investigate convergence of a sequence of empirical spectral distributions.

1.1.1 Empirical spectral distribution

Suppose $\mathbf{A}$ is an $n \times n$ matrix with eigenvalues $\lambda_i, 1 \le i \le n$. If all these eigenvalues are real, we can define a one-dimensional empirical spectral distribution of the matrix $\mathbf{A}$:

$$\mathcal{F}^{\mathbf{A}}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\lambda_i \le x), \tag{1.1.1}$$

where $\mathbb{I}(B)$ denotes the indicator of an event B. If the eigenvalues λ_i are not all real, we can define a two-dimensional empirical spectral distribution of the matrix $\mathbf{A}$:

$$\mathcal{F}^{\mathbf{A}}(x, y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\operatorname{Re}\lambda_i \le x, \operatorname{Im}\lambda_i \le y), \tag{1.1.2}$$

We also denote by $F^{\mathbf{A}}(x) = \mathbb{E}\mathcal{F}^{\mathbf{A}}(x)$ and $F^{\mathbf{A}}(x,y) = \mathbb{E}\mathcal{F}^{\mathbf{A}}(x,y)$ an expected empirical distribution functions of the matrix $\mathbf{A}$.

One of the main problems in RMT is to investigate the convergence of a sequence of empirical distributions $\{\mathcal{F}^{\mathbf{A}_n}\}$ (or $F^{\mathbf{A}_n}$) for a given sequence of random matrices $\mathbf{A}_n$. Under convergence of $\{F^{\mathbf{A}_n}\}$ to some limit F we mean the convergence in vague topology. Under convergence of $\{\mathcal{F}^{\mathbf{A}_n}\}$ to the limit F we mean the convergence almost surely or in probability in vague topology. If it doesn't confuse we shall omit the phrase "in vague topology". The limit distribution F, which is usually non-random, is called the limiting spectral distribution of the sequence $\mathbf{A}_n$.

Sometimes it is more convenient to work with measures then with corresponding distribution functions. We define an empirical spectral measure of eigenvalues of the matrix $\mathbf{A}$:

$$\mu^{\mathbf{A}}(B) = \frac{1}{n}\#\{1 \le i \le n : \lambda_i \in B\}, \quad B \in \mathcal{B}(\mathbb{T}),$$

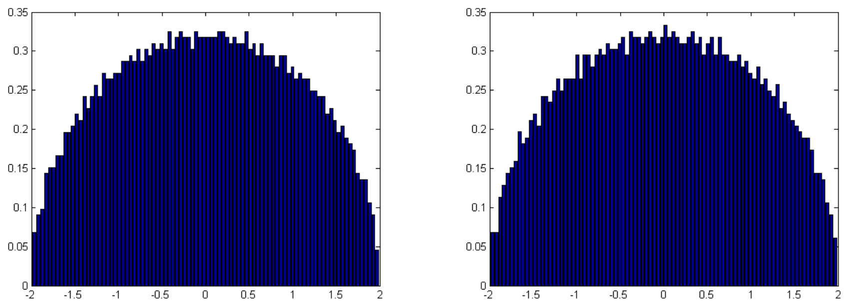

Figure 1.1: Empirical density of the eigenvalues of the symmetric matrix $n^{-1/2}\mathbf{X}_n$ for $n = 3000$, entries are Gaussian random variables. On the left, each entry is an i.i.d. Gaussian normal random variable. On the right, each entry is an i.i.d. Bernoulli random variable, taking the values $+1$ and -1 each with probability $1/2$.

where $\mathbb{T} = \mathbb{C}$ or $\mathbb{T} = \mathbb{R}$ and $\mathcal{B}(\mathbb{T})$ is a Borel σ-algebra of $\mathbb{T}$.

1.1.2 Ensembles of random matrices

In this thesis we will focus on square random matrices with real entries and assume that the size of the matrix tends to infinity. We shall restrict our attention to the following ensembles of random matrices: ensembles of symmetric random matrices, ensembles of random matrices with independent elements and ensembles of random matrices with correlated entries.

Ensembles of symmetric random matrices. Let $X_{jk}, 1 \le j \le k < \infty$, be a triangular array of random variables with $\mathbb{E}X_{jk} = 0$ and $\mathbb{E}X_{jk}^2 = \sigma_{jk}^2$, and let $X_{jk} = X_{kj}$ for $1 \le j < k < \infty$. We consider the random matrix

$$\mathbf{X}_n = \{X_{jk}\}_{j,k=1}^n.$$

Denote by $\lambda_1 \le ... \le \lambda_n$ the eigenvalues of the matrix $n^{-1/2}\mathbf{X}_n$ and define its spectral distribution function $\mathcal{F}^{\mathbf{X}_n}(x)$ by (1.1.1).

Let $g(x)$ and $G(x)$ denote the density and the distribution function of the standard semicircle law

$$g(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\,\mathbb{I}(|x| \le 2), \quad G(x) = \int_{-\infty}^{x} g(u)du.$$

For matrices with independent identically distributed (i.i.d.) elements, which have moments of all orders, Wigner proved in [44] that F_n converges to $G(x)$, later on called "Wigner's semicircle law". See Figure 1.1 for an illustration of Wigner's semicircle law.

The result has been extended in various aspects, i.e. by Arnold in [3]. In the non-i.i.d. case Pastur, [35], showed that Lindeberg's condition is sufficient for the convergence. In [25] Götze and Tikhomirov proved the semicircle law for matrices satisfying martingale-type conditions for the entries.

In the majority of papers it has been assumed that σ_{ij}^2 are equal for all $1 \le i < j \le n$. Recently Erdős, Yau and Yin and al. study ensembles of symmetric random matrices with independent elements which satisfy $n^{-1}\sum_{j=1}^n \sigma_{ij}^2 = 1$ for all $1 \le i \le n$. See for example the survey of results in [15].

In this thesis we study the following class of random matrices with martingale structure. Introduce the σ-algebras

$$\mathfrak{F}^{(i,j)} := \sigma\{X_{kl} : 1 \le k \le l \le n, (k,l) \ne (i,j)\}, 1 \le i \le j \le n.$$

For any $\tau > 0$ we introduce Lindeberg's ratio for random matrices as

$$L_n(\tau) := \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}|X_{ij}|^2 \mathbb{I}(|X_{ij}| \ge \tau\sqrt{n}).$$

We assume that the following conditions hold

$$\mathbb{E}(X_{ij}|\mathfrak{F}^{(i,j)}) = 0; \tag{1.1.3}$$

$$\frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}|\mathbb{E}(X_{ij}^2|\mathfrak{F}^{(i,j)}) - \sigma_{ij}^2| \to 0 \text{ as } n \to \infty; \tag{1.1.4}$$

$$\text{for any fixed } \tau > 0 \quad L_n(\tau) \to 0 \text{ as } n \to \infty; \tag{1.1.5}$$

$$\frac{1}{n}\sum_{i=1}^n \left| \frac{1}{n}\sum_{j=1}^n \sigma_{ij}^2 - 1 \right| \to 0 \text{ as } n \to \infty; \tag{1.1.6}$$

$$\max_{1 \le i \le n} \frac{1}{n}\sum_{j=1}^n \sigma_{ij}^2 \le C, \tag{1.1.7}$$

where C is some absolute constant.

Conditions (1.1.3) and (1.1.4) are analogues of the conditions in the martingale limit theorems, see [26]. Conditions (1.1.6) and (1.1.7) gives us that in average the sum of variances in each row and column is equal to one. Hence, the impact of each row and each column in average is the same for all rows and columns.

If the matrix elements $X_{jk}, 1 \le j \le k < \infty$ are independent then conditions (1.1.3) and (1.1.4) are automatically satisfied and a variant of the semicircle law for

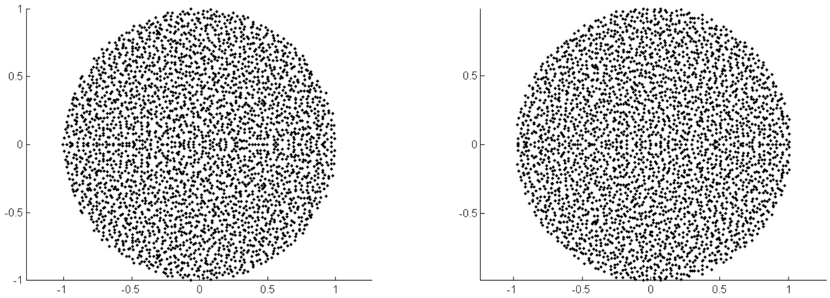

Figure 1.2: Eigenvalues of the matrix $n^{-1/2}\mathbf{X}$ for $n = 3000$ and $\rho = 0$. On the left, each entry is an iid Gaussian normal random variable. On the right, each entry is an iid Bernoulli random variable, taking the values $+1$ and -1 each with probability $1/2$.

matrices with independent entries but not identically distributed holds.

One can find applications of Wigner's semicircle law for matrices which satisfy conditions (1.1.3)– (1.1.7) in [25].

Ensembles of random matrices with independent entries. Let $X_{jk}, 1 \le j, k < \infty$, be an array of independent random variables with $\mathbb{E}X_{jk} = 0$. We consider the random matrix

$$\mathbf{X}_n = \{X_{jk}\}_{j,k=1}^n.$$

Denote by $\lambda_1, ..., \lambda_n$ the eigenvalues of the matrix $n^{-1/2}\mathbf{X}_n$ and define its spectral distribution function $\mathcal{F}^{\mathbf{X}_n}(x, y)$ by (1.1.2).

We say that the Circular law holds if $\mathcal{F}^{\mathbf{X}_n}(x, y)$ ($F^{\mathbf{X}_n}(x, y)$ respectively) converges to the distribution function $F(x, y)$ of the uniform distribution in the unit disc in $\mathbb{R}^2$. $F(x, y)$ is called the circular law. For matrices with independent identically distributed complex normal entries the Circular law was proved by Mehta, see [31]. He used the explicit expression of the joint density of the complex eigenvalues of the random matrix that was found by Ginibre [19]. Under some general conditions Girko proved Circular law in [20], but his proof is considered questionable in the literature. Recently, Edelman [14] proved convergence of $F^{\mathbf{X}_n}(x, y)$ to the circular law for real random matrices whose entries are real normal $N(0, 1)$. Assuming the existence of the $(4 + \varepsilon)$ moment and the existence of a density, Bai, see [4], proved almost sure convergence to the circular law. Under the assumption that $\mathbb{E}X_{11}^2 \log^{19+\varepsilon}(1 + |X_{11}|) < \infty$ Götze and Tikhomirov in [24] proved convergence of $F^{\mathbf{X}_n}(x, y)$ to $F(x, y)$. Almost sure convergence of $\mathcal{F}^{\mathbf{X}_n}(x, y)$ to the circular law under the assumption of a finite fourth, $(2 + \varepsilon)$ and finally of the second moment was established in [34] by Pan, Zhou and by

Tao, Vu in [40], [41] respectively. For a further discussion of the Circular Law see [6].

See Figure 1.2 for an illustration of the Circular law.

Ensembles of random matrices with correlated entries. Let us consider an array of random variables $X_{jk}, 1 \le j, k < \infty$, such that the pairs $(X_{jk}, X_{kj}), 1 \le j < k < \infty$, are independent random vectors with $\mathbb{E}X_{jk} = \mathbb{E}X_{kj} = 0, \mathbb{E}X_{jk}^2 = \mathbb{E}X_{kj}^2 = 1$ and $\mathbb{E}X_{jk}X_{kj} = \rho, |\rho| \le 1$. We also assume that $X_{jj}, 1 \le j < \infty$, are independent random variables, independent of $(X_{jk}, X_{kj}), 1 \le j < k < \infty$, and $\mathbb{E}X_{jj} = 0, \mathbb{E}X_{jj}^2 < \infty$. We consider the random matrix

$$\mathbf{X}_n = \{X_{jk}\}_{j,k=1}^n.$$

Denote by $\lambda_1, ..., \lambda_n$ the eigenvalues of the matrix $n^{-1/2}\mathbf{X}_n$ and define its spectral distribution function $\mathcal{F}^{\mathbf{X}_n}(x, y)$ by (1.1.2).

It is easy to see that this ensemble generalize previous ensembles. If $\rho = 1$ we have the ensemble of symmetric random matrices. If X_{ij} are i.i.d. then $\rho = 0$ and we get the ensemble of matrices with i.i.d. elements.

Define the density of uniformly distributed random variable on the ellipse

$$g(x, y) = \begin{cases} \frac{1}{\pi(1-\rho^2)}, & (x, y) \in \left\{u, v \in \mathbb{R} : \frac{u^2}{(1+\rho)^2} + \frac{v^2}{(1-\rho)^2} \le 1\right\} \\ 0, & \text{otherwise}, \end{cases}$$

and the corresponding distribution function

$$G(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du dv.$$

If all X_{ij} have finite fourth moment and densities then it was proved by Girko in [21] and [22] that $\mathcal{F}^{\mathbf{X}_n}$ converges to G. He called this result "Elliptic Law". But similarly to the case of the Circular law Girko's proof is considered questionable in the literature. Later the Elliptic law was proved for matrices with Gaussian entries in [39]. In this case one can write explicit formula for the density of eigenvalues of the matrix $n^{-1/2}\mathbf{X}_n$. For a discussion of the Elliptic law in the Gaussian case see also [17], [2, Chapter 18] and [29].

Figures 1.3 and 1.4 illustrate the Elliptic law for the two choices of the correlation between elements X_{12} and X_{21}, $\rho = 0.5$ and $\rho = -0.5$.

In this thesis we prove the Elliptic law under the assumption that all elements have a finite fourth moment only. Recently Nguyen and O'Rourke, [32], proved Elliptic law in general case assuming finite second moment only.

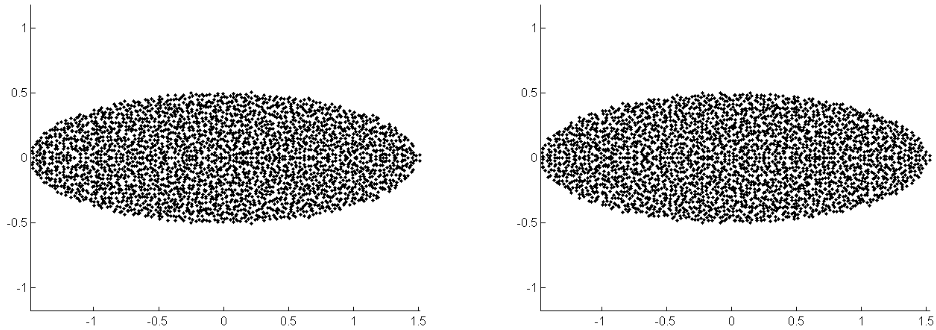

Figure 1.3: Eigenvalues of the matrix $n^{-1/2}\mathbf{X}_n$ for $n = 3000$ and $\rho = 0.5$. On the left, each entry is an iid Gaussian normal random variable. On the right, each entry is an iid Bernoulli random variable, taking the values $+1$ and -1 each with probability $1/2$.

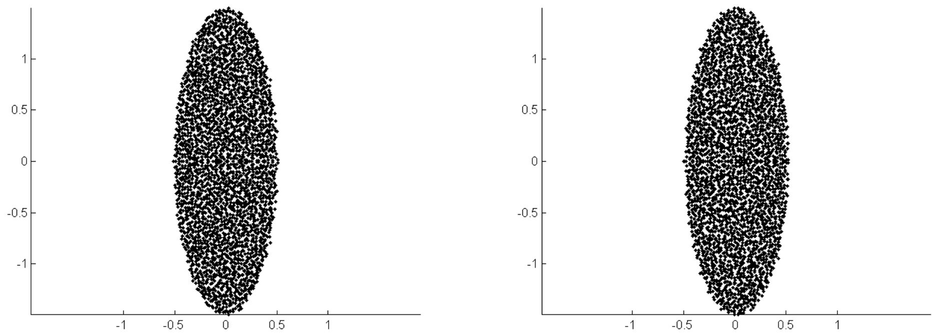

Figure 1.4: Eigenvalues of the matrix $n^{-1/2}\mathbf{X}_n$ for $n = 3000$ and $\rho = -0.5$. On the left, each entry is an iid Gaussian normal random variable. On the right, each entry is an iid Bernoulli random variable, taking the values $+1$ and -1 each with probability $1/2$.

A motivation for such models of random matrices is the the following (see [39]). The statistical properties of random matrices from this ensemble may be important in the understanding of the behavior of certain dynamical systems far from equilibrium. One example is the dynamics of neural networks. A simple dynamic model of neural network consists of n continues "scalar" degrees of freedom("neurons") obeying coupled nonlinear differential equations ("circuit equations"). The coupling between the neurons is given by a synaptic matrix $\mathbf{X}_n$ which, in general, is asymmetric and has a substantial degree of disorder. In this case, the eigenstates of the synaptic matrix play an important role in the dynamics particulary when the neuron nonlinearity is not too big.

1.1.3 Methods

To prove convergence of ESD to some limit we shall apply different methods: the moments method, the Stieltjes transforms method and the method of logarithmic potential. We briefly discuss the main ideas underlying these methods.

Moment method. The basic starting point is the observation that the moments of the ESD $\mathcal{F}^{\mathbf{X}_n}$ can be written as normalized traces of powers of $\mathbf{X}_n$:

$$\int\limits_{\mathbb{R}} x^k d\mathcal{F}^{\mathbf{X}_n} = \frac{1}{n}\operatorname{Tr}\left(\frac{1}{\sqrt{n}}\mathbf{X}\right)^k.$$

Taking expectation we get

$$\int\limits_{\mathbb{R}} x^k dF^{\mathbf{X}_n} = \frac{1}{n}\mathbb{E}\operatorname{Tr}\left(\frac{1}{\sqrt{n}}\mathbf{X}\right)^k.$$

This expression plays a fundamental role in RMT. By the moment convergence theorem the problem of showing that the expected ESD of a sequence of random matrices $\{\mathbf{X}_n\}$ tends to a limit reduces to showing that, for each fixed k, the sequence

$$\frac{1}{n}\mathbb{E}\operatorname{Tr}\left(\frac{1}{\sqrt{n}}\mathbf{X}\right)^k$$

tends to a limit β_k and then verifying the Carleman condition

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

The proof of the convergence of the ESD $\mathcal{F}^{\mathbf{X}_n}$ to a limit almost surely or in probability sense usually reduces to the estimation of the second or higher moments of

$$\frac{1}{n}\operatorname{Tr}\left(\frac{1}{\sqrt{n}}\mathbf{X}\right)^k.$$

We shall apply this method in Chapter 3 for symmetric matrices with Gaussian entries.

Stieltjes transform method. We now turn to the Stieltjes transform method which has turned out to be the most powerful and accurate tools in dealing with the ESD of the random matrix. By definition Stieltjes transform of the distribution function $G(x)$ is

$$S_G(z) = \int_{\mathbb{R}} \frac{1}{x-z} dG(x),$$

for all $z \in \mathbb{C}^+$. One has an inversion formula

$$G([a,b]) = \lim_{\varepsilon \to 0+} \frac{1}{\pi} \int_a^b \operatorname{Im} S_G(x + i\varepsilon) dx,$$

where a, b are continuity points of G and $a < b$. For the ESD of the random matrix $n^{-1/2}\mathbf{X}_n$ one has

$$S^{\mathbf{X}_n}(z) = \int_{\mathbb{R}} \frac{1}{x-z} d\mathcal{F}^{\mathbf{X}_n} = \frac{1}{n} \operatorname{Tr} \left(\frac{1}{\sqrt{n}} \mathbf{X}_n - z\mathbf{I} \right)^{-1}.$$

The quantity in the right hand side of previous formula is the trace of the resolvent of the matrix $n^{-1/2}\mathbf{X}_n - z\mathbf{I}$. By Theorem B.2.3 to prove convergence of the ESD to some limit $F(x)$ one should show convergence of the Stieltjes transforms to the corresponding limit and then show that this limit is the Stieltjes transform of $F(x)$. We will use this method in Chapters 2 and 3.

Method of logarithmic potential. It is well known that methods described above fail to deal with non-hermitian matrices, see for the discussion [4] or [6]. Girko in his paper [20] used the well known and popular in classical probability theory method of characteristic functions. Using V-transform he reduced the problem to the problem for Hermitian matrices $(n^{-1/2}\mathbf{X}_n - z\mathbf{I})^*(n^{-1/2}\mathbf{X}_n - z\mathbf{I})$. In this thesis we will use related method – the method of logarithmic potential.

Denote by μ_n the empirical spectral measure of the matrix $n^{-1/2}\mathbf{X}_n$ and recall the definition of the logarithmic potential (see Appendix B.3)

$$U_{\mu_n}(z) = -\int_{\mathbb{C}} \log|z-w| \mu_n(dw).$$

Let $s_1(n^{-1/2}\mathbf{X}_n - z\mathbf{I}) \geq s_2(n^{-1/2}\mathbf{X}_n - z\mathbf{I}) \geq ... \geq s_n(n^{-1/2}\mathbf{X}_n - z\mathbf{I})$ be the singular values of $n^{-1/2}\mathbf{X}_n - z\mathbf{I}$ and define the empirical spectral measure of singular values by

$$\nu_n(z, B) = \frac{1}{n} \#\{i \geq 1 : s_i(n^{-1/2}\mathbf{X}_n - z\mathbf{I}) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}),$$

We can rewrite the logarithmic potential of the measure μ_n via the logarithmic moments of the measure ν_n:

$$U_{\mu_n}(z) = -\int_{\mathbb{C}} \log|z-w|\mu_n(dw) = -\int_0^\infty \log x \nu_n(z,dx).$$

This allows us to consider the Hermitian matrix $(n^{-1/2}\mathbf{X}_n - z\mathbf{I})^*(n^{-1/2}\mathbf{X}_n - z\mathbf{I})$ instead of the asymmetric matrix $n^{-1/2}\mathbf{X}_n$. Now we can apply all previous methods to find the limiting measure ν_z for the sequence ν_n. Then using uniqueness properties of logarithmic potential one can show that μ_n converges to the unique limit μ and the logarithmic potential of the measure μ is equal to

$$U_\mu(z) = -\int_0^\infty \log x \nu_z(dx),$$

The main problem here is that $\log(\cdot)$ has two poles: at zero and on infinity. To overcome this difficulty we shall explore the behavior of the singular values of the matrix $n^{-1/2}\mathbf{X}_n - z\mathbf{I}$ and show uniform integrability of $\log(\cdot)$ with respect to the family ν_n, see Appendix B.3 for definition. The proof of the uniform integrability is based on the estimation of the least singular value of a square matrix. Recently, considerable progress has been achieved in this question. For a discussion see works of Rudelson, Vershynin [37],Vershynin [43], Götze, Tikhomirov [24] and Tao, Vu [40].

1.2 Universality in the strong law of large numbers

Let $\{X_i\}_{i\geq 1}$ – be a sequence of independent random variables, and denote $S_n = X_1 + ... + X_n$. We say that the strong law of large numbers holds if

$$\lim_{n\to\infty} \frac{S_n - \mathbb{E}S_n}{n} = 0 \text{ a.s.}$$

The strong law of large numbers was first proved by Borel for independent Bernoulli random variables.

One of the first strong laws of large numbers for general random variables was proved by Cantelli under the assumption $\mathbb{E}X_i^4 \leq C$ for all $i \geq 1$.

The most famous sufficient condition was established by Kolmogorov. He proved the strong law of large numbers assuming that the following condition on variances of X_i holds

$$\sum_{i=1}^n \frac{\mathbb{E}|X_i - \mathbb{E}X_i|^2}{n^2} < \infty. \tag{1.2.1}$$

See [38] for the proof.

Brunk and Prokhorov proved ([7] and [36]) that the sequence of arithmetic means $n^{-1}S_n$ converges almost surely to zero if X_n is a sequence of independent random variables with $\mathbb{E}X_i = 0$ and for some $\alpha \geq 1$

$$\sum_{i=1}^{n} \frac{\mathbb{E}|X_i|^{2\alpha}}{n^{\alpha+1}} < \infty, \tag{1.2.2}$$

For $\alpha = 1$ this result coincides with Kolmogorov's theorem. The Kolmogorov theorem and the Brunk–Prokhorov theorem were extended to the case of martingale differences, see [9], [10].

It is natural to try to extend the Kolmogorov and Brunk–Prokhorov theorems replacing the normalizing constants n to other positive quantities. In the case $\alpha = 1$ Loeve, [30], showed that the Kolmogorov theorem can be extended replacing n with positive numbers b_n, such that $b_n \leq b_{n+1}$ and $\lim_{n\to\infty} b_n = \infty$. In [16], it was demonstrated that for $\alpha > 1$ in the Brunk–Prokhorov theorem for martingale difference as normalizing constants one can take positive numbers b_n which satisfy the condition $b_n/b_{n+1} \leq (n/(n+1))^{\delta}, \delta > (\alpha-1)/(2\alpha)$. One should also instead of (1.2.2) assume that

$$\sum_{i=1}^{n} \frac{n^{\alpha-1}\mathbb{E}|X_i|^{2\alpha}}{b_n^{2\alpha}} < \infty. \tag{1.2.3}$$

This assertion is derived in [16, Theorem 3.1], which, as it is pertinent to note, is well known indeed (see, e.g., [9]).

Kruglov in [27] showed that in the Brunk–Prokhorov theorem one can take a sequence b_n such that the condition (1.2.2) holds and there exists a sequence $k_n, n \geq 1$ such that

$$\sup_{n\geq 1} k_{n+1}/k_n = d < \infty,\ 0 < b = \inf_{n\geq 1} b_{k_n}/b_{k_{n+1}} \leq \sup_{n\geq 1} b_{k_n}/b_{k_{n+1}} = c < 1.$$

In this thesis a new generalization of the Brunk-Prokhorov strong law of large numbers is given. We consider a martingale $\{Y_n, n \in \mathbb{N} = \{1, 2, \dots\}\}, Y_0 = 0$, relative to the filtration $\{\mathcal{F}_n, n \in \mathbb{N}\}$ and a sequence $b_n, n \in \mathbb{N}$ of unboundedly increasing positive numbers. We impose the conditions

$$\sum_{n=1}^{\infty} \frac{n^{\alpha-1}\mathbb{E}|Y_n - Y_{n-1}|^{2\alpha}}{b_n^{2\alpha}} < \infty, \tag{1.2.4}$$

$$\sum_{n=1}^{\infty} \frac{n^{\alpha-2}\sum_{k=1}^{n}\mathbb{E}|Y_k - Y_{k-1}|^{2\alpha}}{b_n^{2\alpha}} < \infty \tag{1.2.5}$$

for some $\alpha \geq 1$, and prove that

$$\lim_{n\to\infty} \frac{Y_n}{b_n} = 0 \text{ a.s. and } \lim_{n\to\infty} \mathbb{E}\left|\frac{\max_{1\leq k\leq n} Y_k}{b_n}\right|^{2\alpha} = 0.$$

In some cases, these conditions are automatically satisfied. In particular, they are satisfied under condition of the original Brunk-Prokhorov theorem.

For measurable separable martingales $\{Y_t, t \in \mathbb{R}_+\}$ with continuous parameter we can prove analogs of above theorems. In this case we can substitute the condition (1.2.1) by the following condition

$$\int_1^\infty \frac{d\mathbb{E}|Y_t|^{2\alpha}}{t^{2\alpha}} < \infty,$$

where $\alpha \geq 1$.

Now we turn our attention to independent and identically distributed random variables. If the random variables $X_1, X_2, ...$ are independent and identically distributed then it was proved by Kolmogorov, [38], that for the strong law of large numbers it is sufficient to assume that $\mathbb{E}|X_1| < \infty$. This result can be extended in the following way. Let $X_1, X_2, ...$ – be a sequence of independent identically distributed random variables, and assume that with probability one

$$\lim_{n\to\infty} \frac{S_n}{n} = C,$$

where C - some finite constant. Then $\mathbb{E}|X_1| < \infty$ and $C = \mathbb{E}X_1$. Hence for independent identically distributed random variables the condition $\mathbb{E}|X_1| < \infty$ is necessary and sufficient for convergence of S_n/n to some finite limit. One can also show that almost sure convergence of S_n/n can be replaced by convergence in average

$$\lim_{n\to\infty} \mathbb{E}\left|\frac{S_n}{n} - m\right| = 0.$$

Kolmogorov theorem due to Zygmund and Marcinkiewicz, see [30], can be extended in the following way. Let $X_1, X_2, ...$ – be a sequence of independent identically distributed random variables. If $\mathbb{E}|X_1|^\alpha < \infty$ for some $0 < \alpha < 1$, then

$$\lim_{n\to\infty} \frac{S_n}{n^{1/\alpha}} = 0 \text{ a.s}$$

If $\mathbb{E}|X_1|^\alpha < \infty$ for some $1 \leq \alpha \leq 2$, then

$$\lim_{n\to\infty} \frac{S_n - n\mathbb{E}X_1}{n^{1/\alpha}} = 0 \text{ a.s}$$

In this thesis we also prove analogs of Kolmogorov and Zygmund–Marcinkiewicz strong laws of large numbers for processes with independent homogeneous increments. Along with the convergence almost surely, we also prove the convergence in average .

The classical laws of large numbers are applied in particular in the Monte Carlo methods, e.g. to calculate high dimensional integrals. The proposed analogues of the strong law of large numbers can be used for the same purposes.

1.3 Structure of thesis

The structure of this thesis is as follows: In Chapter 2 we prove the Elliptic law for matrices with finite fourth moment. In Chapter 3 we consider ensembles of random matrices with martingale structure and prove that the empirical distribution function converges to Wigner's semicircle law. In Chapter 4 we establish the strong law of large number for some classes of random processes and give rather general sufficient conditions for convergence. All auxiliary results and necessary definitions are presented in Appendices A– C .

1.4 Notations

Throughout this thesis we assume that all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and we will write almost surely (a.s) instead of $\mathbb{P}$-almost surely. Let $\mathrm{Tr}(\mathbf{A})$ denote the trace of a matrix $\mathbf{A}$. For a vector $x = (x_1, ..., x_n)$ let $||x||_2 := (\sum_{i=1}^{n} x_i^2)^{1/2}$ and $||x||_3 := (\sum_{i=1}^{n} |x_i|^3)^{1/3}$. We denote the operator norm of the matrix $\mathbf{A}$ by $||\mathbf{A}|| := \sup_{||x||_2=1} ||\mathbf{A}x||_2$ and Hilbert–Schmidt norm by $||A||_{HS} := (\mathrm{Tr}(A^*A))^{1/2}$. By $[n]$ we mean the set $\{1, ..., n\}$ and let $\mathrm{supp}(x)$ denote the set of all non-zero coordinates of x. We will write $a \leq_m b$ if there is an absolute constant C depends on m only such that $a \leq Cb$.

Chapter 2

Elliptic law for random matrices

In this chapter we consider ensembles of random matrices $\mathbf{X}_n$ with independent vectors $(X_{ij}, X_{ji})_{i \neq j}$ of entries. Under the assumption of a finite fourth moment for the matrix entries it is proved that the empirical spectral distribution of eigenvalues converges in probability to the uniform distribution on an ellipse. The axes of the ellipse are determined by the correlation between X_{12} and X_{21}. This result is called Elliptic Law. Here the limit distribution is universal, that is it doesn't depend on the distribution of the matrix elements.

2.1 Main result

Let us consider real random matrix $\mathbf{X}_n(\omega) = \{X_{ij}(\omega)\}_{i,j=1}^n$ and assume that the following conditions (**C0**) hold
a) Pairs $(X_{ij}, X_{ji}), i \neq j$ are independent identically distributed (i.i.d.) random vectors;
b) $\mathbb{E}X_{12} = \mathbb{E}X_{21} = 0, \mathbb{E}X_{12}^2 = \mathbb{E}X_{21}^2 = 1$ and $\max(\mathbb{E}|X_{12}|^4, \mathbb{E}|X_{21}|^4) \leq M_4$;
c) $\mathbb{E}(X_{12}X_{21}) = \rho$, $|\rho| \leq 1$;
d) The diagonal entries X_{ii} are i.i.d. random variables, independent of off-diagonal entries, $\mathbb{E}X_{11} = 0$ and $\mathbb{E}X_{11}^2 < \infty$.

Denote by $\lambda_1, ..., \lambda_n$ the eigenvalues of the matrix $n^{-1/2}\mathbf{X}_n$ and define its empirical spectral measure by

$$\mu_n(B) = \frac{1}{n}\#\{1 \leq i \leq n : \lambda_i \in B\}, \quad B \in \mathcal{B}(\mathbb{C}).$$

Theorem 2.1.1 (Elliptic Law). *Let $\mathbf{X}_n$ satisfies the condition* (**C0**) *and* $|\rho| < 1$. *Then $\mu_n \to \mu$ in probability, and μ has the density g:*

$$g(x, y) = \begin{cases} \frac{1}{\pi(1-\rho^2)}, & (x, y) \in \left\{u, v \in \mathbb{R} : \frac{u^2}{(1+\rho)^2} + \frac{v^2}{(1-\rho)^2} \leq 1\right\} \\ 0, & \textit{otherwise.} \end{cases}$$

From now on we shall omit the index n in the notation for random matrices.

2.2 Gaussian case

Let the elements of the matrix $\mathbf{X}$ have Gaussian distribution with zero mean and correlations

$$\mathbb{E}X_{ij}^2 = 1 \text{ and } \mathbb{E}X_{ij}X_{ij} = \rho, \quad i \neq j, \quad |\rho| < 1.$$

The ensemble of such matrices can be specified by the probability measure

$$\mathbb{P}(dX) \sim \exp\left[-\frac{n}{2(1-\rho^2)}\operatorname{Tr}(XX^T - \rho X^2)\right].$$

It was proved that $\mu_n \xrightarrow{weak} \mu$, where μ has a density from Theorem 2.1.1, see [39]. We will use this result to prove Theorem 2.1.1 in the general case.

Remark 2.2.1. *This result can be generalized to the ensemble of Gaussian complex asymmetric matrices. In this case, the invariant measure is*

$$\mathbb{P}(dX) \sim \exp\left[-\frac{n}{1-|\rho|^2}\operatorname{Tr}(XX^T - 2\operatorname{Re}\rho X^2)\right]$$

and $\mathbb{E}|X_{ij}|^2 = 1, \mathbb{E}X_{ij}X_{ji} = |\rho|e^{2i\theta}$ *for* $i \neq j$. *Then the limit measure has a uniform density inside an ellipse which is centered at zero and has semiaxes* $1+|\rho|$ *in the direction* θ *and* $1-|\rho|$ *in the direction* $\theta + \pi/2$.

For a discussion of the Elliptic law in Gaussian case see also [17], [2, Chapter 18] and [29].

2.3 Proof of the main result

To prove Theorem 2.1.1 we shall use the method of the logarithmic potential and Lemma B.3.3.

Denote by $s_1(n^{-1/2}\mathbf{X} - z\mathbf{I}) \geq s_2(n^{-1/2}\mathbf{X} - z\mathbf{I}) \geq ... \geq s_n(n^{-1/2}\mathbf{X} - z\mathbf{I})$ the singular values of $n^{-1/2}\mathbf{X} - z\mathbf{I}$ and define the empirical spectral measure of singular values by

$$\nu_n(z, B) = \frac{1}{n}\#\{i \geq 1 : s_i(n^{-1/2}\mathbf{X} - z\mathbf{I}) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}),$$

We will omit the argument z in the notation of the measure $\nu_n(z, B)$ if it doesn't confuse.

Proof of Theorem 2.1.1. Due to Lemma B.3.3 our aim is to prove the convergence of ν_n to ν_z, uniform integrability of $\log(\cdot)$ with respect to $\{\nu_n\}_{n\geq 1}$ and show that ν_z determines the elliptic law.

From Theorem 2.5.1 we can conclude the uniform integrability of $\log(\cdot)$. The proof of Theorem 2.5.1 is based on Theorem 2.4.1 and some additional results.

In Theorem 2.6.1 it is proved that $\nu_n \xrightarrow{weak} \nu_z$ in probability, where ν_z is some probability measure, which doesn't depend on the distribution of the elements of the matrix $\mathbf{X}$.

If the matrix $\mathbf{X}$ has Gaussian elements we redenote μ_n by $\hat{\mu}_n$.

By Lemma B.3.3 there exists a probability measure $\hat{\mu}$ such that $\mu_n \xrightarrow{weak} \hat{\mu}$ in probability and $U_{\hat{\mu}}(z) = -\int_0^\infty \log x \nu_z(dx)$. But in the Gaussian case $\mu_n \xrightarrow{weak} \mu$ in probability and $U_\mu(z) = -\int_0^\infty \log x \nu_z(dx)$. We know that ν_z is the same for all matrices which satisfy the condition $(\mathbf{C0})$ and we have

$$U_{\hat{\mu}}(z) = -\int_0^\infty \log x \nu_z(dx) = U_\mu(z).$$

From unicity of the logarithmic potential we conclude that $\hat{\mu} = \mu$. □

Remark 2.3.1. *One can also use Theorem 2.7.2 and substitute the elements of the matrix $\mathbf{X}_n$ by Gaussian random variables, which satisfy the condition* $(\mathbf{C0})$.

2.4 Least singular value

Let $s_k(\mathbf{A})$ be the singular values of $\mathbf{A}$ arranged in the non-increasing order. From the properties of the largest and the smallest singular values it follows

$$s_1(\mathbf{A}) = ||A|| = \sup_{x:||x||_2=1} ||\mathbf{A}x||_2, \quad s_n(\mathbf{A}) = \inf_{x:||x||_2=1} ||\mathbf{A}x||_2.$$

To prove uniform integrability of $\log(\cdot)$ we need to estimate the probability of the event $\{s_n(\mathbf{A}) \le \varepsilon n^{-1/2}, ||\mathbf{X}|| \le K\sqrt{n}\}$, where $\mathbf{A} = \mathbf{X} - z\mathbf{I}$. We can assume that $\varepsilon n^{-1/2} \le K n^{1/2}$. If $|z| \ge 2K\sqrt{n}$ then the probability of the event is automatically zero. So we can consider the case when $|z| \le 2Kn^{1/2}$. We have $||\mathbf{A}|| \le ||\mathbf{X}|| + |z| \le 3Kn^{1/2}$. In this section we prove the following theorem

Theorem 2.4.1. *Let $\mathbf{A} = \mathbf{X} - z\mathbf{I}$, where $\mathbf{X}$ is $n \times n$ random matrix satisfying* $(\mathbf{C0})$. *Let $K > 1$. Then for every $\varepsilon > 0$ one has*

$$\mathbb{P}(s_n(\mathbf{A}) \le \varepsilon n^{-1/2}, ||\mathbf{A}|| \le 3K\sqrt{n}) \le C(\rho)\varepsilon^{1/8} + C_1(\rho)n^{-1/8},$$

where $C(\rho), C_1(\rho)$ are some constants which can depend only on ρ, K and M_4.

Remark 2.4.2. *Mark Rudelson and Roman Vershynin in [37] and Roman Vershynin in [43] found the bounds for the least singular value of matrices with independent entries and symmetric matrices respectively. In this section we will follow their ideas.*

2.4.1 The small ball probability via central limit theorem

Let $\mathcal{L}(Z, \varepsilon) = \sup_{v \in \mathbb{R}^d} \mathbb{P}(||Z - v||_2 < \varepsilon)$ be a Levy concentration function of a random variable Z taking values in $\mathbb{R}^d$.

The next statement gives the bound for Levy concentration function of a sum of independent random variables in $\mathbb{R}$.

Statement 2.4.3. *Let* $\{a_i\xi_i + b_i\eta_i\}_{i\ge 1}$ *be independent random variables,* $\mathbb{E}\xi_i = \mathbb{E}\eta_i = 0$, $\mathbb{E}\xi_i^2 = 1, \mathbb{E}\eta_i^2 = 1$, $\mathbb{E}\xi_i\eta_i = \rho$, $\max(\mathbb{E}\xi_i^4, \mathbb{E}\eta_i^4) \le M_4$, $\max\limits_{1\le i\le n} |a_i^{-1}b_i| = O(1)$. *We assume that* $c_1 n^{-1/2} \le |a_i| \le c_2 n^{-1/2}$, *where* c_1, c_2 *are some constants. Then*

$$\mathcal{L}\left(\sum_{i=1}^n (a_i\xi_i + b_i\eta_i), \varepsilon\right) \le \frac{C\varepsilon}{(1-\rho^2)^{1/2}} + \frac{C_1}{(1-\rho^2)^{3/2}n^{1/2}}.$$

Proof. It is easy to see that

$$\sigma^2 = \mathbb{E}(\sum_{i=1}^n Z_i)^2 = \sum_{i=1}^n |a_i|^2(1 + 2\rho a_i^{-1}b_i + (a_i^{-1}b_i)^2) \ge (1-\rho^2)||a||_2^2$$

and

$$\sum_{i=1}^n \mathbb{E}|a_i\xi_i + b_i\eta_i|^3 \le \sum_{i=1}^n |a_i|^3\mathbb{E}|\xi_i + a_i^{-1}b_i\eta_i|^3 \le C'||a||_3^3,$$

where we have used the fact $\max\limits_{1\le i\le n} |a_i^{-1}b_i| = O(1)$. By Central Limit Theorem A.1.1 for arbitrary $v \in \mathbb{R}$

$$\mathbb{P}\left(\left|\sum_{i=1}^n (a_i\xi_i + b_i\eta_i) - v\right| \le \varepsilon\right) \le \mathbb{P}\left(|g' - v| \le \varepsilon\right) + C'' \frac{\sum_{i=1}^n \mathbb{E}|a_i\xi_i + b_i\eta_i|^3}{\sigma^3},$$

where g' has gaussian distribution with zero mean and variance σ^2. The density of g' is uniformly bounded by $1/\sqrt{2\pi\sigma^2}$. We have

$$\mathbb{P}\left(|\sum_{i=1}^n (a_i\xi_i + b_i\eta_i) - v| \le \varepsilon\right) \le \frac{C\varepsilon}{(1-\rho^2)^{1/2}} + \frac{C_1}{(1-\rho^2)^{3/2}n^{1/2}}.$$

We can take maximum and conclude the statement. □

Remark 2.4.4. *Let us consider the case* $b_i = 0$ *for all* $i \ge 1$. *It is easy to show that*

$$\mathcal{L}\left(\sum_{i=1}^n a_i\xi_i, \varepsilon\right) \le C(\varepsilon + n^{-1/2}).$$

2.4.2 Decomposition of the sphere and invertibility

To prove Theorem 2.4.1, we shall partition the unit sphere S^{n-1} into the two sets of compressible and incompressible vectors, and show the invertibility of $\mathbf{A}$ on each set separately. See Appendix A.2 for the definition of compressible and incompressible vectors and their properties.

The following statement gives the bound for compressible vectors.

Lemma 2.4.5. *Let* $\mathbf{A}$ *be a matrix from Theorem 2.4.1 and let* $K > 1$*. There exist constants* $\delta, \tau, c \in (0,1)$ *that depend only on* K *and* M_4 *and such that the following holds. For every* $u \in \mathbb{R}^n$*, one has*

$$\mathbb{P}\left(\inf_{\frac{x}{||x||_2} \in Comp(\delta,\tau)} ||\mathbf{A}x - u||_2/||x||_2 \leq c_4\sqrt{n}, ||\mathbf{A}|| \leq 3K\sqrt{n}\right) \leq 2e^{-cn}. \qquad (2.4.1)$$

Proof. See [43, Statement 4.2]. The proof of this result for matrices which satisfy condition (**C0**) can be carried out by similar arguments. □

For incompressible vectors, we shall reduce the invertibility problem to a lower bound on the distance between a random vector and a random hyperplane. For this aim we recall Lemma 3.5 from [37]

Lemma 2.4.6. *Let* $\mathbf{A}$ *be a random matrix from theorem 2.4.1. Let* $A_1, ..., A_n$ *denote the column vectors of* $\mathbf{A}$*, and let* H_k *denote the span of all columns except the* k*-th. Then for every* $\delta, \tau \in (0,1)$ *and every* $\varepsilon > 0$*, one has*

$$\mathbb{P}(\inf_{x \in Incomp(\delta,\tau)} ||\mathbf{A}x||_2 < \varepsilon n^{-1}) \leq \frac{1}{\delta n}\sum_{k=1}^{n} \mathbb{P}(\mathrm{dist}(A_k, H_k) < \tau^{-1}\varepsilon). \qquad (2.4.2)$$

Lemma 2.4.6 reduces the invertibility problem to a lower bound on the distance between a random vector and a random hyperplane.

We decompose matrix $\mathbf{A} = \mathbf{X} - z\mathbf{I}$ into the blocks

$$\begin{pmatrix} a_{11} & V^T \\ U & \mathbf{B} \end{pmatrix} \qquad (2.4.3)$$

where $\mathbf{B}$ is $(n-1) \times (n-1)$ matrix and U, V random vectors with values in $\mathbb{R}^{n-1}$.

Let h be any unit vector orthogonal to $A_2, ..., A_n$. It follows that

$$0 = \begin{pmatrix} V^T \\ \mathbf{B} \end{pmatrix}^T h = h_1 V + \mathbf{B}^T g,$$

where $h = (h_1, g)$, and

$$g = -h_1 \mathbf{B}^{-T} V$$

From the definition of h

$$1 = ||h||_2^2 = |h_1|^2 + ||g||_2^2 = |h_1|^2 + |h_1|^2 ||\mathbf{B}^{-T}V||_2^2$$

Using this equations we estimate distance

$$\mathrm{dist}(A_1, H) \geq |(A_1, h)| = \frac{|a_{11} - (\mathbf{B}^{-T}V, U)|}{\sqrt{1 + ||\mathbf{B}^{-T}V||_2^2}}$$

It is easy to show that $||\mathbf{B}|| \le ||\mathbf{A}||$. Let the vector $e_1 \in S^{n-2}$ be such that $||\mathbf{B}|| = ||\mathbf{B}e_1||_2$. Then we can take the vector $e = (0, e_1)^T \in S^{n-1}$ and for this vector

$$||\mathbf{A}|| \ge ||\mathbf{A}e||_2 = ||(V^T e_1, \mathbf{B}e_1)^T||_2 \ge ||\mathbf{B}e_1||_2 = ||\mathbf{B}||.$$

The bound for right hand sand of (2.4.2) will follow from the following statement

Lemma 2.4.7. *Let matrix* $\mathbf{A}$ *be from Theorem 2.4.1. Then for all* $\varepsilon > 0$

$$\sup_{v\in\mathbb{R}} \mathbb{P}\left(\frac{|(\mathbf{B}^{-T}V, U) - v|}{\sqrt{1 + ||\mathbf{B}^{-T}V||_2^2}} \le \varepsilon, ||\mathbf{B}|| \le 3K\sqrt{n}\right) \le C(\rho)\varepsilon^{1/8} + C'(\rho)n^{-1/8}, \quad (2.4.4)$$

where $\mathbf{B}, U, V$ *are determined by* (2.4.3) *and* $C(\rho), C_1(\rho)$ *are some constants which can depend only on* ρ, K *and* M_4.

To get this bound we need several statements. First we introduce matrix and vector

$$\mathbf{Q} = \begin{pmatrix} \mathbf{O}_{n-1} & \mathbf{B}^{-T} \\ \mathbf{B}^{-1} & \mathbf{O}_{n-1} \end{pmatrix} \quad W = \begin{pmatrix} U \\ V \end{pmatrix}, \quad (2.4.5)$$

where $\mathbf{O}_{n-1}$ is $(n-1)\times(n-1)$ matrix with zero entries. The scalar product in (2.4.4) can be rewritten using definition of Q:

$$\sup_{v\in\mathbb{R}} \mathbb{P}\left(\frac{|(\mathbf{Q}W, W) - v|}{\sqrt{1 + ||\mathbf{B}^{-T}V||_2^2}} \le 2\varepsilon\right). \quad (2.4.6)$$

Introduce vectors

$$W' = \begin{pmatrix} U' \\ V' \end{pmatrix} \quad Z = \begin{pmatrix} U \\ V' \end{pmatrix}, \quad (2.4.7)$$

where U', U' are independent copies of U, V respectively. We need the following statement.

Statement 2.4.8.

$$\sup_{v\in\mathbb{R}} \mathbb{P}_W\left(|(\mathbf{Q}W, W) - v| \le 2\varepsilon\right) \le \mathbb{P}_{W,W'}\left(|(\mathbf{Q}\mathbf{P}_{J^c}(W - W'), \mathbf{P}_J W) - u| \le 2\varepsilon\right),$$

where u *doesn't depend on* $\mathbf{P}_J W = (\mathbf{P}_J U, \mathbf{P}_J V)^T$.

Proof. Let us fix v and denote

$$p := \mathbb{P}\left(|(\mathbf{Q}W, W) - v| \le 2\varepsilon\right).$$

We can decompose the set $[n]$ into union $[n] = J \cup J^c$. We can take $U_1 = \mathbf{P}_J U, U_2 = \mathbf{P}_{J^c} U, V_1 = \mathbf{P}_J V$ and $V_2 = \mathbf{P}_{J^c} V$. By Lemma A.1.2

$$\begin{aligned} p^2 &\le \mathbb{P}\left(|(\mathbf{Q}W, W) - v| \le 2\varepsilon, |(\mathbf{Q}Z, Z) - v| \le 2\varepsilon\right) \\ &\le \mathbb{P}\left(|(\mathbf{Q}W, W) - (\mathbf{Q}Z, Z)| \le 4\varepsilon\right). \end{aligned} \quad (2.4.8)$$

Let us rewrite $\mathbf{B}^{-T}$ in the block form

$$\mathbf{B}^{-T} = \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix}.$$

We get

$$\begin{aligned}(\mathbf{Q}W, W) &= (\mathbf{E}V_1, U_1) + (\mathbf{F}V_2, U_1) + (\mathbf{G}V_1, U_2) + (\mathbf{H}V_2, U_2)\\ &+ (\mathbf{E}^T U_1, V_1) + (\mathbf{G}^T U_2, V_1) + (\mathbf{F}^T U_1, V_2) + (\mathbf{H}^T U_2, V_2)\\ (\mathbf{Q}Z, Z) &= (\mathbf{E}V_1, U_1) + (\mathbf{F}V_2', U_1) + (\mathbf{G}V_1, U_2') + (\mathbf{H}V_2', U_2')\\ &+ (\mathbf{E}^T U_1, V_1) + (\mathbf{G}^T U_2', V_1) + (\mathbf{F}^T U_1, V_2') + (\mathbf{H}^T U_2', V_2')\end{aligned}$$

and

$$\begin{aligned}(\mathbf{Q}W, W) - (\mathbf{Q}Z, Z) &= 2(\mathbf{F}(V_2 - V_2'), U_1) + 2(\mathbf{G}^T(U_2 - U_2'), V_1)\\ &+ 2(\mathbf{H}V_2, V_2) - 2(\mathbf{H}V_2', V_2').\end{aligned} \tag{2.4.9}$$

The last two terms in (2.4.9) depend only on U_2, U_2', V_2, V_2' and we conclude that

$$p_1^2 \le \mathbb{P}\left(|(\mathbf{Q}P_{J^c}(W - W'), P_J W) - u| \le 2\varepsilon\right),$$

where $u = u(U_2, V_2, U_2', V_2', \mathbf{F}, \mathbf{G}, \mathbf{H})$. □

Statement 2.4.9. *For all $u \in \mathbb{R}^{n-1}$*

$$\mathbb{P}\left(\frac{\mathbf{B}^{-T}u}{||\mathbf{B}^{-T}u||_2} \in Comp(\delta, \tau) \text{ and } ||\mathbf{B}|| \le 3Kn^{1/2}\right) \le 2e^{-cn}.$$

Proof. Let $x = \mathbf{B}^{-T}u$. It is easy to see that

$$\left\{\frac{\mathbf{B}^{-T}u}{||\mathbf{B}^{-T}u||_2} \in Comp(\delta, \tau)\right\} \subseteq \left\{\exists x : \frac{x}{||x||_2} \in Comp(\delta, \tau) \text{ and } \mathbf{B}^T x = u\right\}$$

Replacing the matrix $\mathbf{A}$ with $\mathbf{B}^T$ one can easily check that the proof of Lemma 2.4.5 remains valid for $\mathbf{B}^T$ as well as for $\mathbf{A}$. □

Remark 2.4.10. *The Statement 2.4.9 holds for $\mathbf{B}^{-T}$ replaced with $\mathbf{B}^{-1}$.*

Statement 2.4.11. *Let $\mathbf{A}$ satisfies the condition ($\mathbf{C0}$) and $\mathbf{B}$ be the matrix from the decomposition (2.4.3). Assume that $||\mathbf{B}|| \le 3K\sqrt{n}$. Then with probability at least $1 - e^{-cn}$ matrix $\mathbf{B}$ has the following properties:*

a) $||\mathbf{B}^{-T}V||_2 \ge C$ with probability $1 - e^{-c'n}$ in W,

b) $||\mathbf{B}^{-T}V||_2 \le \varepsilon^{-1/2}||\mathbf{B}^{-T}||_{HS}$ with probability $1 - \varepsilon$ in V,

c) $||\mathbf{Q}W||_2 \ge \varepsilon||\mathbf{B}^{-T}||_{HS}$ with probability $1 - C'(\varepsilon + n^{-1/2})$ in W.

Proof. Let $\{e_k\}_{k=1}^n$ be a standard basis in $\mathbb{R}^n$. For all $1 \le k \le n$ define vectors by

$$x_k := \frac{\mathbf{B}^{-1}e_k}{||\mathbf{B}^{-1}e_k||}.$$

By Statement 2.4.9 vector x_k is incompressible with probability $1 - e^{-cn}$. We fix the matrix $\mathbf{B}$ with such property.
a) By the norm inequality $||U||_2 \le ||\mathbf{B}||_2||\mathbf{B}^{-T}U||_2$. We know that $||\mathbf{B}|| \le 3K\sqrt{n}$. By Lemma A.1.6 and Lemma A.1.7 $||U|| \ge \sqrt{n}$. Hence we have that $||\mathbf{B}^{-1}U|| \ge C$ with probability $1 - e^{-c'n}$.
b) By definition

$$||\mathbf{B}^{-T}V||_2^2 = \sum_{i=1}^n (\mathbf{B}^{-1}e_k, V)^2 = \sum_{i=1}^n ||\mathbf{B}^{-1}e_i||_2^2 (x_k, V)^2.$$

It is easy to see that $\mathbb{E}(V, x_k)^2 = 1$. So

$$\mathbb{E}||\mathbf{B}^{-T}V||_2^2 = \sum_{i=1}^n ||\mathbf{B}^{-1}e_i||_2^2 = ||\mathbf{B}^{-1}||_{HS}^2.$$

By the Markov inequality

$$\mathbb{P}(||\mathbf{B}^{-T}V||_2 \ge \varepsilon^{-1/2}||\mathbf{B}^{-1}||_{HS}) \le \varepsilon.$$

c) By Lemma A.1.3, Lemma A.2.3, Lemma A.1.5 and Remark 2.4.4 we get

$$\begin{aligned}
&\mathbb{P}(||\mathbf{Q}W||_2 \le \varepsilon||\mathbf{B}^{-1}||_{HS}) \le \mathbb{P}(||\mathbf{B}^{-T}V||_2 \le \varepsilon||B^{-1}||_{HS}) \\
&= \mathbb{P}(||\mathbf{B}^{-T}V||_2^2 \le \varepsilon||\mathbf{B}^{-1}||_{HS}^2) = \mathbb{P}(\sum_{i=1}^n ||\mathbf{B}^{-1}e_i||_2 (x_i, V)^2 \le \varepsilon^2||\mathbf{B}^{-1}||_{HS}^2) \\
&= \mathbb{P}(\sum_{i=1}^n p_i (x_i, V)^2 \le \varepsilon^2) \le 2\sum_{i=1}^n p_i \mathbb{P}((x_i, V) \le \sqrt{2}\varepsilon) \le C'(\varepsilon + n^{-1/2}).
\end{aligned}$$

□

Proof of Lemma 2.4.7. Let $\xi_1, ..., \xi_n$ be i.i.d. Bernoulli random variables with $\mathbb{E}\xi_i = c_0/2$. We define $J := \{i : \xi_i = 0\}$ and $\mathbb{E}_0 := \{|J^c| \le c_0 n\}$. From the large deviation inequality we may conclude that $\mathbb{P}(E_0) \ge 1 - 2\exp(-c_0^2 n/2)$. Introduce the event

$$E_1 := \{\varepsilon_0^{1/2}\sqrt{1 + ||\mathbf{B}^{-T}V||_2^2} \le ||\mathbf{B}^{-1}||_{HS} \le \varepsilon_0^{-1}||\mathbf{Q}\mathbf{P}_{J^c}(W - W')||_2\},$$

where ε_0 will be chosen later.

From Statement 2.4.11 we may conclude that

$$\mathbb{P}_{\mathbf{B},W,W',J}(E_1 \cup ||B|| \ge 3K\sqrt{n}) \ge 1 - C'(\varepsilon_0 + n^{-1/2}) - 2e^{-c'n}.$$

Consider the random vector

$$w_0 = \frac{1}{||\mathbf{QP}_{J^c}(W-W')||_2}\begin{pmatrix}\mathbf{B}^{-T}\mathbf{P}_{J^c}(V-V') \\ \mathbf{B}^{-1}\mathbf{P}_{J^c}(U-U')\end{pmatrix} = \begin{pmatrix} a \\ b\end{pmatrix}.$$

By Statement 2.4.9 it follows that the event $E_2 := \{a \in incomp(\delta, \tau)\}$ holds with probability

$$\mathbb{P}_{\mathbf{B}}(E_2 \cup ||\mathbf{B}|| \geq 3K\sqrt{n}|W, W', J) \geq 1 - 2\exp(-c''n).$$

Combining these probabilities we have

$$\begin{aligned}&\mathbb{P}_{\mathbf{B},W,W',J}(E_0,\ E_1,\ E_2 \cup ||\mathbf{B}|| \geq 3K\sqrt{n}) \\ &\geq 1 - 2e^{-c_0^2 n/2} - C'(\varepsilon_0 + n^{-1/2}) - 2e^{-c'n} - 2e^{-c''n} := 1 - p_0.\end{aligned}$$

We may fix J that satisfies $|J^c| \leq c_0$ and

$$\mathbb{P}_{\mathbf{B},W,W'}(E_1,\ E_2 \cup ||\mathbf{B}|| \geq 3K\sqrt{n}) \geq 1 - p_0.$$

By Fubini's theorem $\mathbf{B}$ has the following property with probability at least $1 - \sqrt{p_0}$

$$\mathbb{P}_{W,W'}(E_1,\ E_2 \cup ||\mathbf{B}|| \geq 3K\sqrt{n}|\mathbf{B}) \geq 1 - \sqrt{p_0}.$$

The event $\{||\mathbf{B}|| \geq 3K\sqrt{n}\}$ depends only on $\mathbf{B}$. We may conclude that the random matrix $\mathbf{B}$ has the following property with probability at least $1-\sqrt{p_0}$: either $||\mathbf{B}|| \geq 3K\sqrt{n}$, or

$$||\mathbf{B}|| \leq 3K\sqrt{n} \text{ and } \mathbb{P}_{W,W'}(E_1,\ E_2|\mathbf{B}) \geq 1 - \sqrt{p}_0 \tag{2.4.10}$$

The event we are interested in is

$$\Omega_0 := \left(\frac{|(\mathbf{Q}W, W) - u|}{\sqrt{1 + ||B^{-T}V||_2^2}} \leq 2\varepsilon\right).$$

We need to estimate the probability

$$\begin{aligned}\mathbb{P}_{\mathbf{B},W}(\Omega_0 \cap ||\mathbf{B}|| \leq 3K\sqrt{n}) &\leq \mathbb{P}_{\mathbf{B},W}(\Omega_0 \cap \text{ (2.4.10) holds}) \\ &+\mathbb{P}_{\mathbf{B},W}(||\mathbf{B}|| \leq 3K\sqrt{n} \cap \text{ (2.4.10) fails}).\end{aligned}$$

By the previous steps the last term is bounded by $\sqrt{p_0}$.

$$\mathbb{P}_{\mathbf{B},W}(\Omega_0 \cap ||\mathbf{B}|| \leq 3K\sqrt{n}) \leq \sup_{B \text{ satisfies (2.4.10)}} \mathbb{P}_W(\Omega_0|\mathbf{B}) + \sqrt{p_0}.$$

We can conclude that

$$\mathbb{P}_{\mathbf{B},W}(\Omega_0 \cap ||\mathbf{B}|| \leq 3K\sqrt{n}) \leq \sup_{B \text{ satisfies (2.4.10)}} \mathbb{P}_{W,W'}(\Omega_0,\ E_1|\mathbf{B}) + 2\sqrt{p_0}.$$

Let us fix $\mathbf{B}$ that satisfies (2.4.10) and denote $p_1 := \mathbb{P}_{W,W'}(\Omega_0,\ E_1|\mathbf{B})$. By Statement 2.4.8 and the first inequality in E_1 we get

$$p_1^2 \le \mathbb{P}_{W,W'}\left(\underbrace{|(\mathbf{Q}\mathbf{P}_{J^c}(W-W'), \mathbf{P}_J W) - v| \le \frac{\varepsilon}{\sqrt{\varepsilon_0}}||\mathbf{B}^{-1}||_{HS}}_{\Omega_1}\right)$$

and

$$\mathbb{P}_{W,W'}(\Omega_1) \le \mathbb{P}_{W,W'}(\Omega_1,\ E_1,\ E_2) + \sqrt{p_0}.$$

Further

$$p_1^2 \le \mathbb{P}_{W,W'}(|(w_0, \mathbf{P}_J W) - v| \le 2\varepsilon_0^{-3/2}\varepsilon,\ E_2) + \sqrt{p_0}.$$

By definition random vector w_0 is determined by the random vector $\mathbf{P}_{J^c}(W-W')$, which is independent of the random vector $\mathbf{P}_J W$. We fix $\mathbf{P}_{J^c}(W-W')$ and have

$$p_1^2 \le \sup_{\substack{w_0=(a,b)^T:\\ a\in Incomp(\delta,\tau)\\ w\in\mathbb{R}}} \mathbb{P}_{\mathbf{P}_J W}\left(|(w_0, \mathbf{P}_J W) - w| \le \varepsilon_0^{-3/2}\varepsilon\right) + \sqrt{p_0}.$$

Let us fix a vector w_0 and a number w. We can rewrite

$$(w_0, P_J W) = \sum_{i\in J}(a_i x_i + b_i y_i), \tag{2.4.11}$$

where $||a||_2^2 + ||b||_2^2 = 1$. From Lemma A.2.3 and Remark A.2.4 we know that at least $[2c_0 n]$ coordinates of vector $a \in Incomp(\delta,\tau)$ satisfy

$$\frac{\tau}{\sqrt{2n}} \le |a_k| \le \frac{1}{\sqrt{\delta n}},$$

where $\delta\tau^2/4 \le c_0 \le 1/4$. We denote the set of coordinates of a with this property by $\mathrm{spread}(a)$. By the construction of J we can conclude that $|\mathrm{spread}(a)| = [c_0 n]$. By Lemma A.1.5 we may reduce the sum (2.4.11) to the set $\mathrm{spread}(a)$. Now we will investigate the properties of $|b_i|$. Let us decompose the set $\mathrm{spread}(a)$ into the two sets $\mathrm{spread}(a) = I_1 \cup I_2$, where

a) $I_1 = \{i \in \mathrm{spread}(a) : |b_i| > Cn^{-1/2}\}$;

b) $I_2 = \{i \in \mathrm{spread}(a) : |b_i| \le Cn^{-1/2}\}$,

and C is some big constant. From $||b||_2^2 < 1$ it follows that $|I_1| \le \hat{c}_0 n$, where $c_0 \gg \hat{c}_0$ and $\hat{c}_0$ depends on C. For the second set I_2 we have $\max_{i\in I_2}|a_i^{-1}b_i| = O(1)$. By Lemma A.1.5 we get

$$\mathbb{P}(|\sum_{i\in\mathrm{spread}(a)}(a_i x_i + b_i y_i) - w| < 2\varepsilon_0^{-3/2}\varepsilon) \le \mathbb{P}(|\sum_{i\in I_2}(a_i x_i + b_i y_i) - w'| < 2\varepsilon_0^{-3/2}\varepsilon).$$

We can apply Statement 2.4.3

$$\mathbb{P}(|\sum_{i\in I_2}(a_i x_i + b_i y_i) - w'| < 2\varepsilon_0^{-3/2}\varepsilon) \le \frac{C_1\varepsilon_0^{-3/2}\varepsilon}{(1-\rho^2)^{1/2}} + C_2(1-\rho^2)^{-3/2}n^{-1/2}.$$

It follows that

$$\mathbb{P}_{\mathbf{B},W}(\Omega_0 \cap ||\mathbf{B}|| \le 3K\sqrt{n}) \le \left(\frac{C_1 \varepsilon_0^{-3/2}\varepsilon}{(1-\rho^2)^{1/2}} + C_2(1-\rho^2)^{-3/2}n^{-1/2}\right)^{1/2} + p_0^{1/4} + 2\sqrt{p_0}.$$

We take $\varepsilon_0 = \varepsilon^{1/2}$ and finally conclude

$$\mathbb{P}_{\mathbf{B},W}(\Omega_0 \cap ||\mathbf{B}|| \le 3K\sqrt{n}) \le C(\rho)\varepsilon^{1/8} + C'(\rho)n^{-1/8},$$

where $C(\rho), C'(\rho)$ are some constants which depend on ρ, K and M_4. □

Proof of Theorem 2.4.1. The result of the theorem follows from Lemmas 2.4.5–2.4.7. □

Remark 2.4.12. *It not very difficult to show that we can change matrix* $z\mathbf{I}$ *in Theorem 2.4.1 by an arbitrary non-random matrix* $\mathbf{M}$ *with* $||\mathbf{M}|| \le K\sqrt{n}$. *Results of the section 2.4.2 are based on Lemmas A.1.6 and A.1.7 which doesn't depend on shifts. It is easy to see that Statement 2.4.11 still holds true if we assume that* $\varepsilon < n^{-Q}$ *for some* $Q > 0$. *Then we can reformulate Theorem 2.4.1 in the following way: there exist some constants* $A, B > 0$ *such that*

$$\mathbb{P}(s_n(\mathbf{X}+\mathbf{M}) \le \varepsilon n^{-A}, ||\mathbf{X}+\mathbf{M}|| \le K\sqrt{n}) \le C(\rho)n^{-B}.$$

2.5 Uniform integrability of logarithm

In this section we prove the next result

Theorem 2.5.1. *Under the condition* (**C0**) $\log(\cdot)$ *is uniformly integrable in probability with respect to* $\{\nu_n\}_{n\ge1}$.

Before we need several lemmas about the behavior of singular values.

Lemma 2.5.2. *If the condition* (**C0**) *holds then there exists a constant* $K := K(\rho)$ *such that* $\mathbb{P}(s_1(\mathbf{X}) \ge K\sqrt{n}) = o(1)$.

Proof. Let us decompose the matrix $\mathbf{X}$ into the symmetric and skew-symmetric matrices:

$$\mathbf{X} = \frac{\mathbf{X}+\mathbf{X}^T}{2} + \frac{\mathbf{X}-\mathbf{X}^T}{2} = \mathbf{X}_1 + \mathbf{X}_2.$$

In [42, Theorem 2.3.23] it is proved that for some $K_1 > \sqrt{2(1+\rho)}$

$$\mathbb{P}(s_1(\mathbf{X}_1) \ge K_1\sqrt{n}) = o(1). \tag{2.5.1}$$

and for some $K_2 > \sqrt{2(1-\rho)}$

$$\mathbb{P}(s_1(\mathbf{X}_2) \ge K_2\sqrt{n}) = o(1) \tag{2.5.2}$$

Set $K = 2\max(K_1, K_2)$. From (2.5.1), (2.5.2) and inequality

$$s_1(\mathbf{X}) \le s_1(\mathbf{X}_1) + s_1(\mathbf{X}_2)$$

we may conclude the following bound

$$\mathbb{P}(s_1(\mathbf{X}) \ge K\sqrt{n}) \le \mathbb{P}\left(\left\{s_1(\mathbf{X}_1) \ge \frac{K\sqrt{n}}{2}\right\} \cup \left\{s_1(\mathbf{X}_2) \ge \frac{K\sqrt{n}}{2}\right\}\right)$$
$$\le \mathbb{P}\left(s_1(\mathbf{X}_1) \ge \frac{K\sqrt{n}}{2}\right) + \mathbb{P}\left(s_1(\mathbf{X}_2) \ge \frac{K\sqrt{n}}{2}\right) = o(1).$$

□

Lemma 2.5.3. *If the condition* (**C0**) *holds then there exist $c > 0$ and $0 < \gamma < 1$ such that a.s. for $n \gg 1$ and $n^{1-\gamma} \le i \le n-1$*

$$s_{n-i}(n^{-1/2}\mathbf{X} - z\mathbf{I}) \ge c\frac{i}{n}.$$

Proof. Set $s_i := s_i(n^{-1/2}\mathbf{X} - z\mathbf{I})$. Up to increasing γ, it is sufficient to prove the statement for all $2(n-1)^{1-\gamma} \le i \le n-1$ for some $\gamma \in (0,1)$ to be chosen later. We fix some $2(n-1)^{1-\gamma} \le i \le n-1$ and consider the matrix $\mathbf{A}'$ formed by the first $m := n - \lceil i/2 \rceil$ rows of $\sqrt{n}\mathbf{A}$. Let $s'_1 \ge ... \ge s'_m$ be the singular values of $\mathbf{A}'$. We get

$$n^{-1/2}s'_{n-i} \le s_{n-i}.$$

By R_i we denote the row of $\mathbf{A}'$ and $H_i = \operatorname{span}(R_j, j = 1, ..., m, j \ne i)$. By Lemma A.2.1 we obtain

$$s_1'^{-2} + ... + s'^{-2}_{n-\lceil i/2 \rceil} = \operatorname{dist}_1^{-2} + ... + \operatorname{dist}^{-2}_{n-\lceil i/2 \rceil}.$$

We have

$$\frac{i}{2n}s^{-2}_{n-i} \le \frac{i}{2}s'^{-2}_{n-i} \le \sum_{j=n-i}^{n-\lceil i/2 \rceil} s'^{-2}_j \le \sum_{j=1}^{n-\lceil i/2 \rceil} \operatorname{dist}_j^{-2}, \tag{2.5.3}$$

where $\operatorname{dist}_j := \operatorname{dist}(R_j, H_j)$. To estimate $\operatorname{dist}(R_j, H_j)$ we would like to apply Lemma A.1.8, but we can't do it directly, because R_j and H_j are not independent. Let's consider the case $j = 1$ only. To estimate the distance dist_1 we decompose the matrix $\mathbf{A}'$ into the blocks

$$\mathbf{A}' = \begin{pmatrix} a_{1,1} & Y \\ X & \mathbf{B} \end{pmatrix},$$

where $X \in \mathbb{R}^{m-1}, Y^T \in \mathbb{R}^{n-1}$ and $\mathbf{B}$ is an $m-1 \times n-1$ matrix formed by rows $B_1, ..., B_{m-1}$. We denote by $H_1' = \operatorname{span}(B_1, ..., B_{m-1})$. From the definition of the distance

$$\operatorname{dist}(R_1, H_1) = \inf_{v \in H_1} ||R_1 - v||_2 \ge \inf_{u \in H'} ||Y - u||_2 = \operatorname{dist}(Y, H_1')$$

and

$$\dim(H_1') \le \dim(H_1) \le n-1-i/2 \le n-1-(n-1)^{1-\gamma}.$$

Now the vector Y and the hyperplane H_1' are independent. Fixing realization of H_1', by Lemma A.1.8, with n, R, H replaced with $n-1, Y, H_1'$ respectively, we can obtain that

$$\mathbb{P}(\operatorname{dist}(Y, H_1') \le \frac{1}{2}\sqrt{n-1-\dim(H_1')}) \le \exp(-(n-1)^{\delta}).$$

Using this inequality it is easy to show that

$$\mathbb{P}\left(\bigcup_{n\gg 1}\bigcup_{i=\lceil 2(n-1)^{1-\gamma}\rceil}^{n-1}\bigcup_{j=1}^{n-\lceil i/2\rceil}\left\{\operatorname{dist}(R_j, H_j) \le \frac{1}{2}\sqrt{\frac{i}{2}}\right\}\right) < \infty.$$

Now by the Borel-Cantelli lemma and (2.5.3) we can conclude the statement of the lemma. □

Proof of Theorem 2.5.1. To prove Theorem 2.5.1 we need to show that there exist $p, q > 0$ such that

$$\lim_{t\to\infty}\overline{\lim_{n\to\infty}}\,\mathbb{P}\left(\int_0^{\infty} x^p \nu_n(dx) > t\right) = 0 \tag{2.5.4}$$

and

$$\lim_{t\to\infty}\overline{\lim_{n\to\infty}}\,\mathbb{P}\left(\int_0^{\infty} x^{-q} \nu_n(dx) > t\right) = 0. \tag{2.5.5}$$

By Lemma 2.5.2 there exists the set $\Omega_0 := \Omega_{0,n} = \{\omega \in \Omega : s_1(\mathbf{X}) \le K n^{1/2}\}$ such that

$$\mathbb{P}(\Omega_0) = 1 - o(1). \tag{2.5.6}$$

We conclude (2.5.4) from (2.5.6) for $p = 2$.

We denote $\Omega_1 := \Omega_{1,n} = \{\omega \in \Omega : s_{n-i} > \frac{ci}{n}, n^{1-\gamma} \le i \le n-1\}$. Let us consider the set $\Omega_2 := \Omega_{2,n} = \Omega_1 \cap \{\omega : s_n \ge n^{-B-1/2}\}$, where $B > 0$. We decompose probability from (2.5.5) into two terms

$$\mathbb{P}\left(\int_0^{\infty} x^{-q}\nu_n(dx) > t\right) = \mathbb{I}_1 + \mathbb{I}_2,$$

where

$$\mathbb{I}_1 := \mathbb{P}\left(\int_0^{\infty} x^{-q}\nu_n(dx) > t, \Omega_2\right),$$
$$\mathbb{I}_2 := \mathbb{P}\left(\int_0^{\infty} x^{-q}\nu_n(dx) > t, \Omega_2^c\right).$$

We can estimate $\mathbb{I}_2$ by

$$\mathbb{I}_2 \le \mathbb{P}(s_n(\mathbf{X} - \sqrt{n}z\mathbf{I}) \le n^{-A}, \Omega_0) + \mathbb{P}(\Omega_0^c) + \mathbb{P}(\Omega_1^c).$$

From Theorem 2.4.1 it follows that

$$\mathbb{P}(s_n(\mathbf{X} - \sqrt{n}z\mathbf{I}) \le n^{-B}, \Omega_0) \le C(\rho)n^{-1/8}. \tag{2.5.7}$$

By Lemma 2.5.3

$$\overline{\lim_{n\to\infty}} \mathbb{P}(\Omega_1^c) = 0. \tag{2.5.8}$$

From (2.5.6), (2.5.7) and (2.5.8) we conclude

$$\overline{\lim_{n\to\infty}} \mathbb{I}_2 = 0.$$

To prove (2.5.5) it remains to bound $\mathbb{I}_1$. From the Markov inequality

$$\mathbb{I}_1 \le \frac{1}{t}\mathbb{E}\left[\int_0^\infty x^{-q}\nu_n(dx)\mathbb{I}(\Omega_2)\right].$$

By the definition of Ω_2

$$\mathbb{E}\left[\int x^{-q}\nu_n(dx)\mathbb{I}(\Omega_2)\right] \le \frac{1}{n}\sum_{i=1}^{n-\lceil n^{1-\gamma}\rceil} s_i^{-q} + \frac{1}{n}\sum_{i=n-\lceil n^{1-\gamma}\rceil+1}^{n} s_i^{-q}$$
$$\le 2n^{q(B+1/2)-\gamma} + c^{-q}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{n}{i}\right)^q \le 2n^{q(B+1/2)-\gamma} + c^{-q}\int_0^1 s^{-q}ds.$$

If $0 < q < \min(1, \gamma/(B+1/2))$ then the last integral is finite. □

2.6 Convergence of singular values

Let $\mathcal{F}_n(x,z)$ be the empirical distribution function of the singular values $s_1 \ge ... \ge s_n$ of the matrix $n^{-1/2}\mathbf{X} - z\mathbf{I}$ which corresponds to the measure $\nu_n(z,\cdot)$.
In this section we prove the following theorem

Theorem 2.6.1. *Assume that the condition* (**C0**) *holds true. There exists a non-random distribution function $\mathcal{F}(x,z)$ such that for all continues and bounded functions $f(x)$, a.a. $z \in \mathbb{C}$ and all $\varepsilon > 0$*

$$\mathbb{P}\left(\left|\int_{\mathbb{R}} f(x)d\mathcal{F}_n(x,z) - \int_{\mathbb{R}} f(x)d\mathcal{F}(x,z)\right| > \varepsilon\right) \to 0 \text{ as } n \to \infty,$$

Proof. First we show that the family $\{\mathcal{F}(z,x)\}_{n\ge1}$ is tight. From the strong law of large numbers it follows that

$$\int_0^\infty x^2 d\mathcal{F}(x,z) \le \frac{1}{n^2}\sum_{i,j=1}^n X_{ij}^2 \to 1 \text{ as } n \to \infty.$$

Using this and the fact that $s_i(n^{-1/2}\mathbf{X} - z\mathbf{I}) \le s_i(n^{-1/2}\mathbf{X}) + |z|$ we conclude the tightness of $\{\mathcal{F}_n(z,x)\}_{n\ge1}$. If we show that $\mathcal{F}_n$ weakly converges in probability to some function $\mathcal{F}$, then $\mathcal{F}$ will be distribution function.

Introduce the following $2n \times 2n$ matrices

$$\mathbf{V} = \begin{pmatrix} \mathbf{O}_n & n^{-1/2}\mathbf{X} \\ n^{-1/2}\mathbf{X}^T & \mathbf{O}_n \end{pmatrix}, \quad \mathbf{J}(z) = \begin{pmatrix} \mathbf{O}_n & z\mathbf{I} \\ \overline{z}\mathbf{I} & \mathbf{O}_n \end{pmatrix} \tag{2.6.1}$$

where $\mathbf{O}_n$ denotes $n \times n$ matrix with zero entries. Consider the matrix

$$\mathbf{V}(z) := \mathbf{V} - \mathbf{J}(z). \tag{2.6.2}$$

It is well known that the eigenvalues of $\mathbf{V}(z)$ are the singular values of $n^{-1/2}\mathbf{X} - z\mathbf{I}$ with signs $\pm$.

It is easy to see that the empirical distribution function $F_n(x, z)$ of eigenvalues of the matrix $\mathbf{V}(z)$ can be written in the following way

$$F_n(x, z) = \frac{1}{2n}\sum_{i=1}^n \mathbb{I}\{s_i \le x\} + \frac{1}{2n}\sum_{i=1}^n \mathbb{I}\{-s_i \le x\}. \tag{2.6.3}$$

There is one to one correspondence between $\mathcal{F}_n(x, z)$ and $F_n(x, z)$

$$F_n(x, z) = \frac{1 + \operatorname{sgn}(x)\mathcal{F}_n(|x|, z)}{2}$$

Hence it is enough to show that there exists a non-random distribution function $F(x, z)$ such that for all continues and bounded functions $f(x)$, and a.a. $z \in \mathbb{C}$

$$\mathbb{P}\left(\left|\int_{\mathbb{R}} f(x)dF_n(x, z) - \int_{\mathbb{R}} f(x)dF(x, z)\right| > \varepsilon\right) \to 0 \text{ as } n \to \infty. \tag{2.6.4}$$

We denote the Stieltjes transforms of F_n and F by $S_n(x, z)$ and $S(x, z)$ respectively. Due to the relations between distribution functions and Stieltjes transforms, see Theorem B.2.3, (2.6.4) will follow from

$$\mathbb{P}(|S_n(\alpha, z) - S(\alpha, z)| > \varepsilon) \to 0 \text{ as } n \to \infty, \tag{2.6.5}$$

for a.a. $z \in \mathbb{C}$ and all $\alpha \in \mathbb{C}^+$.
Set

$$\mathbf{R}(\alpha, z) := (\mathbf{V}(z) - \alpha\mathbf{I}_{2n})^{-1}. \tag{2.6.6}$$

By definition $S_n(\alpha, z) = \frac{1}{2n}\operatorname{Tr}\mathbf{R}(\alpha, z)$. We introduce the following function

$$s_n(\alpha, z) := \mathbb{E}S_n(\alpha, z) = \frac{1}{2n}\sum_{i=1}^{2n}\mathbb{E}[\mathbf{R}(\alpha, z)]_{ii},$$

One can show that

$$s_n(\alpha, z) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\mathbf{R}(\alpha, z)]_{ii} = \frac{1}{n}\sum_{i=n+1}^{2n}\mathbb{E}[\mathbf{R}(\alpha, z)]_{ii}$$

We also denote $s(\alpha, z) := S(\alpha, z)$. By the Chebyshev inequality and Lemma 2.8.1 it is straightforward to check that

$$|s_n(\alpha, z) - s(\alpha, z)| \to 0 \text{ as } n \to \infty. \tag{2.6.7}$$

implies (2.6.5).

By the resolvent equality we may write

$$1 + \alpha s_n(\alpha, z) = \frac{1}{2n}\mathbb{E}\,\mathrm{Tr}(\mathbf{V}\mathbf{R}(\alpha, z)) - z t_n(\alpha, z) - \overline{z} u_n(\alpha, z).$$

Introduce the notation

$$\mathbb{A} := \frac{1}{2n}\mathbb{E}\,\mathrm{Tr}(\mathbf{V}\mathbf{R})$$

and represent $\mathbb{A}$ as follows

$$\mathbb{A} = \frac{1}{2}\mathbb{A}_1 + \frac{1}{2}\mathbb{A}_2,$$

where

$$\mathbb{A}_1 = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\mathbf{V}\mathbf{R}]_{ii}, \quad \mathbb{A}_2 = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\mathbf{V}\mathbf{R}]_{i+n,i+n}.$$

First we consider $\mathbb{A}_1$. By definition of the matrix $\mathbf{V}$, we have

$$\mathbb{A}_1 = \frac{1}{n^{3/2}}\sum_{j,k=1}^{n}\mathbb{E}X_{jk}R_{k+n,j}.$$

Note that

$$\frac{\partial \mathbf{R}}{\partial X_{jk}} = -\frac{1}{\sqrt{n}}\mathbf{R}[e_j e_{k+n}^T]\mathbf{R},$$

where we denote by e_i the column vector with 1 in position i and zeros in the other positions. Applying Lemma 2.8.3 we obtain

$$\mathbb{A}_1 = \mathbb{B}_1 + \mathbb{B}_2 + \mathbb{B}_3 + \mathbb{B}_4 + r_n(\alpha, z).$$

where

$$\mathbb{B}_1 = -\frac{1}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\mathbf{R}[e_j e_{k+n}^T]\mathbf{R}]_{k+n,j} = -\frac{1}{n^2}\sum_{j,k=1}^{n}\mathbb{E}(R_{k+n,j})^2$$

$$\mathbb{B}_2 = -\frac{1}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\mathbf{R}[e_{k+n} e_j^T]\mathbf{R}]_{k+n,j} = -\frac{1}{n^2}\sum_{j,k=1}^{n}\mathbb{E}R_{jj}R_{k+n,k+n}$$

$$\mathbb{B}_3 = -\frac{\rho}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\mathbf{R}[e_k e_{j+n}^T]\mathbf{R}]_{k+n,j} = -\frac{\rho}{n^2}\sum_{j,k=1}^{n}\mathbb{E}R_{k+n,k}R_{j+n,j}$$

$$\mathbb{B}_4 = -\frac{\rho}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\mathbf{R}[e_{j+n} e_k^T]\mathbf{R}]_{k+n,j} = -\frac{\rho}{n^2}\sum_{j,k=1}^{n}\mathbb{E}R_{kj}R_{k+n,j+n}.$$

Without loss of generality we shall assume from now on that $\mathbb{E}X_{11}^2 = 1$ because the impact of the diagonal is of order $O(n^{-1})$.

From $||\mathbf{R}||_{HS} \le \sqrt{n}||\mathbf{R}|| \le \sqrt{n}v^{-1}$ it follows

$$|\mathbb{B}_1| \le \frac{1}{n^2} \sum_{j,k=1}^{n} \mathbb{E}X_{jk}^2 \mathbb{E}(R_{k+n,j})^2 \le \frac{1}{nv^2}.$$

Similarly we get

$$|\mathbb{B}_4| \le \frac{1}{v^2 n}.$$

By Lemma 2.8.1 $\mathbb{B}_2 = -s_n^2(\alpha, z) + \varepsilon(\alpha, z)$. By Lemma 2.8.2 $\mathbb{B}_3 = -\rho t_n^2(\alpha, z) + \varepsilon(\alpha, z)$. We obtain that

$$\mathbb{A}_1 = -s_n^2(\alpha, z) - \rho t_n^2(\alpha, z) + \delta_n(\alpha, z).$$

Now we consider the term $\mathbb{A}_2$. By definition of the matrix $\mathbf{V}$, we have

$$\mathbb{A}_2 = \frac{1}{n^{3/2}} \sum_{j,k=1}^{n} \mathbb{E}X_{jk} R_{j,k+n}.$$

By Lemma 2.8.3 we may write expansion

$$\mathbb{A}_2 = \mathbb{C}_1 + \mathbb{C}_2 + \mathbb{C}_3 + \mathbb{C}_4 + r_n(\alpha, z). \tag{2.6.8}$$

where

$$\mathbb{C}_1 = -\frac{1}{n^2} \sum_{j,k=1}^{n} \mathbb{E}[\mathbf{R}[e_j e_{k+n}^T]\mathbf{R}]_{j,k+n} = -\frac{1}{n^2} \sum_{j,k=1}^{n} \mathbb{E}R_{jj} R_{k+n,k+n}$$

$$\mathbb{C}_2 = -\frac{1}{n^2} \sum_{j,k=1}^{n} \mathbb{E}[\mathbf{R}[e_{k+n} e_j^T]\mathbf{R}]_{j,k+n} = -\frac{1}{n^2} \sum_{j,k=1}^{n} \mathbb{E}(R_{j,k+n})^2$$

$$\mathbb{C}_3 = -\frac{\rho}{n^2} \sum_{j,k=1}^{n} \mathbb{E}[\mathbf{R}[e_k e_{j+n}^T]\mathbf{R}]_{j,k+n} = -\frac{\rho}{n^2} \sum_{j,k=1}^{n} \mathbb{E}R_{jk} R_{j+n,k+n}$$

$$\mathbb{C}_4 = -\frac{\rho}{n^2} \sum_{j,k=1}^{n} \mathbb{E}[\mathbf{R}[e_{j+n} e_k^T]\mathbf{R}]_{j,k+n} = -\frac{\rho}{n^2} \sum_{j,k=1}^{n} \mathbb{E}R_{j,j+n} R_{k,k+n}.$$

It is easy to show that

$$|\mathbb{C}_2| \le \frac{1}{v^2 n}, \quad |\mathbb{C}_3| \le \frac{1}{v^2 n}.$$

By Lemma 2.8.1 $\mathbb{C}_1 = -s_n^2(\alpha, z) + \varepsilon_n(\alpha, z)$. By Lemma 2.8.2 $\mathbb{C}_4 = -\rho u_n^2(\alpha, z) + \varepsilon_n(\alpha, z)$. We obtain that

$$\mathbb{A}_2 = -s_n^2(\alpha, z) - \rho u_n^2(\alpha, z) + \delta_n(\alpha, z).$$

Finally we get

$$\mathbb{A} = -s_n^2(\alpha, z) - \frac{\rho}{2} t_n^2(\alpha, z) - \frac{\rho}{2} u_n^2(\alpha, z) + \varepsilon_n(\alpha, z).$$

Now we will investigate the term $zt_n(\alpha, z)$ which we may represent as follows

$$\alpha t_n(\alpha, z) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}[\mathbf{V}(z)\mathbf{R}]_{j+n,j} = \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}[\mathbf{V}\mathbf{R}]_{j+n,j} - \overline{z} s_n(\alpha, z).$$

By definition of the matrix $\mathbf{V}$, we have

$$\alpha t_n(\alpha, z) = \frac{1}{n^{3/2}} \sum_{j,k=1}^{n} \mathbb{E} X_{jk} R_{j,k} - \overline{z} s_n(\alpha, z) =$$
$$\mathbb{D}_1 + \mathbb{D}_2 + \mathbb{D}_3 + \mathbb{D}_4 - \overline{z} s_n(\alpha, z) + r_n(\alpha, z),$$

where

$$\mathbb{D}_1 = -\frac{1}{n^2} \sum_{j,k=1}^{n} \mathbb{E}[\mathbf{R}[e_j e_{k+n}^T]\mathbf{R}]_{j,k} = -\frac{1}{n^2} \sum_{j,k=1}^{n} \mathbb{E} R_{j,j} R_{k+n,k}$$
$$\mathbb{D}_2 = -\frac{1}{n^2} \sum_{j,k=1}^{n} \mathbb{E}[\mathbf{R}[e_{k+n} e_j^T]\mathbf{R}]_{j,k} = -\frac{1}{n^2} \sum_{j,k=1}^{n} \mathbb{E} R_{j,k+n} R_{j,k}$$
$$\mathbb{D}_3 = -\frac{\rho}{n^2} \sum_{j,k=1}^{n} \mathbb{E}[\mathbf{R}[e_k e_{j+n}^T]\mathbf{R}]_{j,k} = -\frac{\rho}{n^2} \sum_{j,k=1}^{n} \mathbb{E} R_{j,k} R_{j+n,k}$$
$$\mathbb{D}_4 = -\frac{\rho}{n^2} \sum_{j,k=1}^{n} \mathbb{E}[\mathbf{R}[e_{j+n} e_k^T]\mathbf{R}]_{j,k} = -\frac{\rho}{n^2} \sum_{j,k=1}^{n} \mathbb{E} R_{j,j+n} R_{k,k}.$$

By the similar arguments as before we can prove that

$$|\mathbb{D}_2| \le \frac{1}{v^2 n}, \quad |\mathbb{D}_3| \le \frac{1}{v^2 n}$$

and $\mathbb{D}_1 = -s_n(\alpha, z) t_n(\alpha, z) + \varepsilon_n(\alpha, z)$, $\mathbb{D}_4 = -\rho s_n(\alpha, z) u_n(\alpha, z) + \varepsilon_n(\alpha, z)$. We obtain that

$$\alpha t_n(\alpha, z) = -s_n(\alpha, z) t_n(\alpha, z) - \rho s_n(\alpha, z) u_n(\alpha, z) - \overline{z} s_n(\alpha, z) + \delta_n(\alpha, z).$$

Similar we can prove that

$$\alpha u_n(\alpha, z) = -s_n(\alpha, z) u_n(\alpha, z) - \rho s_n(\alpha, z) t_n(\alpha, z) - z s_n(\alpha, z) + \delta_n(\alpha, z).$$

Finally we have the system of equations

$$1 + \alpha s_n(\alpha, z) + s_n^2(\alpha, z) = \tag{2.6.9}$$
$$- \frac{\rho}{2} t_n^2(\alpha, z) - \frac{z}{2} t_n(\alpha, z) - \frac{\rho}{2} u_n^2(\alpha, z) - \frac{\overline{z}}{2} u_n(\alpha, z) + \delta_n(\alpha, z)$$
$$\alpha t_n(\alpha, z) = \tag{2.6.10}$$
$$- s_n(\alpha, z) t_n(\alpha, z) - \rho s_n(\alpha, z) u_n(\alpha, z) - \overline{z} s_n(\alpha, z) + \delta_n(\alpha, z)$$
$$\alpha u_n(\alpha, z) = \tag{2.6.11}$$
$$- s_n(\alpha, z) u_n(\alpha, z) - \rho s_n(\alpha, z) t_n(\alpha, z) - z s_n(\alpha, z) + \delta_n(\alpha, z).$$

It follows from (2.6.10) and (2.6.11) that

$$(\alpha+s_n)(zt_n+\rho t_n^2) = -s_n(z\rho u_n+\overline{z}\rho t) - \rho^2 s_n t_n u_n - |z|^2 s_n + \delta_n(\alpha,z)$$
$$(\alpha+s_n)(\overline{z}u_n+\rho u_n^2) = -s_n(z\rho u_n+\overline{z}\rho t) - \rho^2 s_n t_n u_n - |z|^2 s_n + \delta_n(\alpha,z).$$

Hence, we can rewrite (2.6.9)

$$1+\alpha s_n(\alpha,z)+s_n^2(\alpha,z)+\rho^2 t_n^2(\alpha,z)+zt_n(\alpha,z)=\delta_n(\alpha,z). \tag{2.6.12}$$

From equations (2.6.10) and (2.6.11) we can write the equation for t_n

$$\left(\alpha+s_n-\frac{|\rho|^2 s_n^2}{\alpha+s_n}\right)t_n=\frac{\rho z s_n^2}{\alpha+s_n}-\overline{z}s_n+\delta_n(\alpha,z). \tag{2.6.13}$$

Let us denote

$$\Delta=\left(\alpha+s_n-\frac{|\rho|^2 s_n^2}{\alpha+s_n}\right).$$

After simple calculations we get

$$\begin{aligned}
&(\alpha+s_n)(zt_n+\rho t_n^2)=\\
&-s_n\left(\frac{2\rho^2|z|^2 s_n^2}{(\alpha+s_n)\Delta}-\frac{\overline{z}^2\rho s_n}{\Delta}-\frac{z^2\overline{\rho}s_n}{\Delta}\right)\\
&-|\rho|^2 s_n\left(\frac{\rho z s_n^2}{(\alpha+s)\Delta}-\frac{\overline{z}s_n}{\Delta}\right)\left(\frac{\rho\overline{z}s_n^2}{(\alpha+s)\Delta}-\frac{zs_n}{\Delta}\right)-|z|^2 s_n+\delta_n(\alpha,z).
\end{aligned}$$

Let us denote $y_n := s_n$ and $w_n := \alpha+(\rho t_n^2+zt_n)/y_n$. We can rewrite the equations (2.6.9), (2.6.10) and (2.6.11) in the following way

$$1+w_n y_n+y_n^2=\delta_n(\alpha,z) \tag{2.6.14}$$

$$w_n=\alpha+\frac{\rho t_n^2+zt_n}{y_n} \tag{2.6.15}$$

$$\begin{aligned}
&(\alpha+y_n)(zt_n+\rho t_n^2)=\\
&-y_n\left(\frac{2\rho^2|z|^2 y_n^2}{(\alpha+y_n)\Delta}-\frac{\overline{z}^2\rho y_n}{\Delta}-\frac{z^2\rho y_n}{\Delta}\right)-|z|^2 y_n\\
&-|\rho|^2 y_n\left(\frac{\rho z y_n^2}{(\alpha+y_n)\Delta}-\frac{\overline{z}y_n}{\Delta}\right)\left(\frac{\rho\overline{z}y_n^2}{(\alpha+y_n)\Delta}-\frac{zy_n}{\Delta}\right)+\delta_n(\alpha,z).
\end{aligned} \tag{2.6.16}$$

Remark 2.6.2. *If $\rho=0$ then we can rewrite (2.6.14), (2.6.15), and (2.6.16)*

$$\begin{aligned}
&1+w_n y_n+y_n^2=\delta_n(\alpha,z)\\
&w_n=\alpha+\frac{zt_n}{y_n}\\
&(w_n-\alpha)+(w_n-\alpha)^2 y_n-|z|^2 y_n=\delta_n(\alpha,z).
\end{aligned}$$

This equations determine the circular law, see [24].

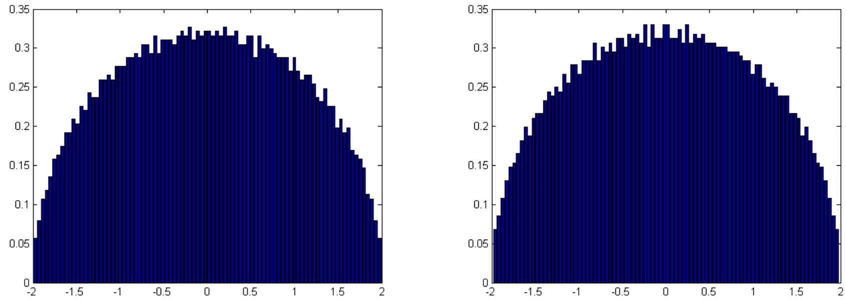

Figure 2.1: Empirical density of the eigenvalues of the matrix $\mathbf{V}$ for $n = 2000$. entries are Gaussian random variables. On the left $\rho = 0$ (Circular law case). On the right $\rho = 0.5$ (Elliptic law case).

We can see that the first equation (2.6.14) doesn't depend on ρ. Hence the first equation will be the same for all models of random matrices described in the introduction. On the Figure 2.1 we draw the distribution of eigenvalues of the matrix $\mathbf{V}$ for $\rho = 0$ (Circular law case) and $\rho = 0.5$ (Elliptic law case).

Now we prove the convergence of s_n to some limit s_0. Let $\alpha = u + iv, v > 0$. Using (2.6.12) we write

$$\alpha(s_n - s_m) = -(s_n - s_m)(s_n + s_m) - \rho^2(t_n - t_m)(t_n + t_m) - z(t_m - t_m) + \varepsilon_{n,m}.$$

By the triangle inequality and the fact that $|s_n| \leq v^{-1}$

$$|s_n - s_m| \leq \frac{2|s_n - s_m|}{v^2} + \frac{\rho^2|t_n - t_m||t_n + t_m|}{v} + \frac{|z||t_n - t_m|}{v} + \frac{|\varepsilon_{n,m}|}{v}. \tag{2.6.17}$$

From (2.6.13) it follows that

$$((\alpha + s_n)^2 - \rho^2 s_n^2)t_n = \rho z s_n^2 - \overline{z}\alpha s_n - \overline{z}s_n^2 + \varepsilon_n.$$

We denote $\Delta_n := ((\alpha + s_n)^2 - \rho^2 s_n^2)$. Again by the triangle inequality

$$\begin{aligned} |\Delta_m||t_n - t_m| &\leq |t_m||\Delta_n - \Delta_m| \\ &+ \frac{2|\rho||s_n - s_m| + 2|z||s_n - s_m|}{v} + |z||\alpha||s_n - s_m| + |\varepsilon_{n,m}|. \end{aligned} \tag{2.6.18}$$

We can find the lower bound for $|\Delta_m|$:

$$\begin{aligned} |\Delta_m| &= |\alpha + (1 - \rho)s_m||\alpha + (1 + \rho)s_m| \\ &\geq \operatorname{Im}(\alpha + (1 - \rho)s_m) \operatorname{Im}(\alpha + (1 + \rho)s_m) \geq v^2, \end{aligned} \tag{2.6.19}$$

where we have used the fact that $\operatorname{Im} s_m \geq 0$. From definition of Δ_n it is easy to see that

$$|\Delta_n - \Delta_m| \leq 2|\alpha||s_n - s_m| + \frac{2(1 + \rho^2)|s_n - s_m|}{v}. \tag{2.6.20}$$

We can take $|u| \le C$, then $|\alpha| \le v + C$. From (2.6.17),(2.6.18),(2.6.19) and (2.6.20) it follows that there exists constant C', which depends on ρ, C, z, such that

$$|s_n - s_m| \le \frac{C'}{v}|s_n - s_m| + |\varepsilon'_{n,m}(\alpha, z)|.$$

We can find v_0 such that

$$\frac{C'}{v} < 1 \quad \text{for all } v \ge v_0.$$

Since $\varepsilon'_{n,m}(\alpha, z)$ converges to zero uniformly for all $v \ge v_0, |u| \le C$ and s_n, s_m are locally bounded analytic functions in the upper half-plane we may conclude by Montel's Theorem (see [11, Theorem 2.9]) that there exists an analytic function s in the upper half-plane such that $\lim s_n = s$. Since s_n are Nevanlinna functions, (that is analytic functions mapping the upper half-plane into itself) s will be a Nevanlinna function too and there exists non-random distribution function $F(z, x)$ such that

$$s(\alpha, z) = \int \frac{dF(z, x)}{x - \alpha}.$$

The function s satisfies the equations (2.6.14), (2.6.15), and (2.6.16). □

2.7 Lindeberg's universality principe

It this section we will work with the random matrices $\mathbf{X}$ which satisfy the following conditions (**C1**):
a) Pairs $(X_{ij}, X_{ji}), i \ne j$ are independent random vectors;
b) $\mathbb{E}X_{ij} = \mathbb{E}X_{ji} = 0, \mathbb{E}X_{ij}^2 = \mathbb{E}X_{ji}^2 = 1$;
c) $\mathbb{E}(X_{ij}X_{ji}) = \rho$, $|\rho| \le 1$;
d) The diagonal entries X_{ii} are independent of off-diagonal entries, $\mathbb{E}X_{ii} = 0$ and $\mathbb{E}X_{ii}^2 < \infty$;
e) For all fixed $\tau > 0$ Lindeberg's condition holds

$$L_n(\tau) := \frac{1}{n^2}\sum_{i,j=1}^{n} \mathbb{E}|X_{ij}|^2 \mathbb{I}(|X_{ij}| \ge \tau\sqrt{n}) \to 0 \text{ as } n \to \infty. \tag{2.7.1}$$

Remark 2.7.1. *It is easy to see that the condition* (**C1**) *follows from the condition* (**C0**).

Let $\mathcal{F}_n(x, z)$ be the empirical distribution function of the singular values $s_1 \ge ... \ge s_n$ of the matrix $n^{-1/2}\mathbf{X} - z\mathbf{I}$ which corresponds to the measure $\nu_n(z, \cdot)$. Similar we define the function $\mathcal{G}_n(x, z)$ if the matrix $\mathbf{X}$ satisfies (**C1**) and has Gaussian elements.

We prove the following theorem.

Theorem 2.7.2. *Under the condition* (**C1**) *for all continues and bounded functions* $f(x)$*, a.a.* $z \in \mathbb{C}$ *and all* $\varepsilon > 0$

$$\mathbb{P}\left(\left|\int_{\mathbb{R}} f(x) d\mathcal{F}_n(x,z) - \int_{\mathbb{R}} f(x) d\mathcal{G}_n(x,z)\right| > \varepsilon\right) \to 0 \text{ as } n \to \infty.$$

Let us introduce the matrices $\mathbf{V}, \mathbf{J}(z), \mathbf{V}(z)$ by formulas (2.6.1), (2.6.2) and the empirical distribution function $F_n(x,z)$ of the matrix $\mathbf{V}(z)$ by the formula (2.6.3). Similarly we define $G_n(x,z)$. Due to one to one correspondence between $\mathcal{F}_n(x,z)$ and $F_n(x,z)$ it is enough to show that for all continues and bounded functions $f(x)$, and a.a. $z \in \mathbb{C}$

$$\mathbb{P}\left(\left|\int_{\mathbb{R}} f(x) dF_n(x,z) - \int_{\mathbb{R}} f(x) dG_n(x,z)\right| > \varepsilon\right) \to 0 \text{ as } n \to \infty. \tag{2.7.2}$$

We denote the Stieltjes transforms of F_n and G_n by $S_n(x,z)$ and $\hat{S}_n(x,z)$ respectively. Due to the relations between distribution functions and Stieltjes transforms, (2.7.2) will follow from

$$\mathbb{P}(|S_n(\alpha,z) - \hat{S}_n(\alpha,z)| > \varepsilon) \to 0 \text{ as } n \to \infty, \tag{2.7.3}$$

for a.a. $z \in \mathbb{C}$ and all $\alpha \in \mathbb{C}^+$.
Set

$$\mathbf{R}(\alpha,z) := (\mathbf{V}(z) - \alpha \mathbf{I}_{2n})^{-1}.$$

By definition $S_n(\alpha,z) = \frac{1}{2n} \operatorname{Tr} \mathbf{R}(\alpha,z)$. We introduce the following function

$$s_n(\alpha,z) := \mathbb{E} S_n(\alpha,z) = \frac{1}{2n} \sum_{i=1}^{2n} \mathbb{E}[\mathbf{R}(\alpha,z)]_{ii},$$

Similarly we can define $\hat{s}_n(\alpha,z)$. One can show that

$$s_n(\alpha,z) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\mathbf{R}(\alpha,z)]_{ii} = \frac{1}{n} \sum_{i=n+1}^{2n} \mathbb{E}[\mathbf{R}(\alpha,z)]_{ii}$$

By the Chebyshev inequality and Lemma 2.8.1 it is straightforward to check that

$$|s_n(\alpha,z) - \hat{s}_n(\alpha,z)| \to 0 \text{ as } n \to \infty. \tag{2.7.4}$$

implies (2.7.3).

We divide the proof of (2.7.4) into the two subsections 2.7.1 and 2.7.2.

Note that we can substitute τ in (2.7.1) by a decreasing sequence τ_n tending to zero such that:

$$L_n(\tau_n) \to 0 \text{ as } n \to \infty, \tag{2.7.5}$$

and $\lim_{n\to\infty} \tau_n \sqrt{n} = \infty$.

2.7.1 Truncation

In this section we truncate and centralize the elements of the matrix $\mathbf{X}$. We define the matrices $\mathbf{X}^{(c)} = [X_{ij}\mathbb{I}(|X_{ij}| \le \tau_n\sqrt{n})]_{i,j=1}^n$ and $\mathbf{V}^{(c)}$ replacing $\mathbf{X}$ by $\mathbf{X}^{(c)}$ in (2.6.1). Denote the empirical distribution function of eigenvalues of $\mathbf{V}^{(c)}(z)$ by $F_n^{(c)}(x,z)$. Due to [4, Theorem A.43] the uniform distance between the empirical distribution functions $F_n(x,z)$ and $F_n^{(c)}(x,z)$ can be estimated by

$$\sup_x |F_n(x,z) - F_n^{(c)}(x,z)| \le \frac{1}{2n}\operatorname{Rank}(\mathbf{V}(z) - \mathbf{V}^{(c)}(z)).$$

The right hand side can be bounded by

$$\min\left(1, \frac{1}{n}\sum_{i\le j}\mathbb{I}(|X_{ij}| \ge \tau_n\sqrt{n})\right).$$

Denote $\xi_n := \frac{1}{n}\sum_{i\le j}\mathbb{I}(|X_{ij}| \ge \tau_n\sqrt{n})$. It is easy to see that

$$\mathbb{E}\xi_n \le \frac{1}{\tau_n^2 n^2}\sum_{i\le j}\mathbb{E}X_{ij}^2\mathbb{I}(|X_{ij}| \ge \tau_n\sqrt{n}) \to 0 \tag{2.7.6}$$

and

$$\mathbb{E}(\xi_n - \mathbb{E}\xi_n)^2 \le \frac{1}{n^3\tau_n^2}\sum_{i\le j}\mathbb{E}X_{ij}^2\mathbb{I}(|X_{ij}| \ge \tau_n\sqrt{n}) = o\left(\frac{1}{n}\right). \tag{2.7.7}$$

By the Bernstein's inequality

$$\mathbb{P}(|\xi_n - \mathbb{E}\xi_n| \ge \varepsilon) \le 2\exp\left(-\frac{\varepsilon^2 n}{n\mathbb{E}(\xi_n - \mathbb{E}\xi_n)^2 + \varepsilon}\right).$$

By (2.7.6), (2.7.7) and the Borel-Cantelli Lemma we conclude that a.s.

$$\sup_x |F_n(x,z) - F_n^{(c)}(x,z)| \to 0 \text{ as } n \to \infty.$$

Now we centralize the entries of $\mathbf{X}^{(c)}$. Define the matrices $\overline{\mathbf{X}} = \left[X_{ij}^{(c)} - \mathbb{E}X_{ij}^{(c)}\right]_{i,j=1}^n$ and $\overline{\mathbf{V}}$ replacing $\mathbf{X}$ by $\overline{\mathbf{X}}$ in (2.6.1). Denote the empirical distribution function of eigenvalues of $\overline{\mathbf{V}}(z)$ by $\overline{F}_n(x,z)$.

Let $L(F,G)$ be the Levy distance between the empirical distribution functions of eigenvalues of matrices $\mathbf{A}$ and $\mathbf{B}$. If $\mathbf{A}$ and $\mathbf{B}$ are normal matrices, then it is proved in [4, Corollary A.41] that

$$L^3(F,G) \le \frac{1}{n}\operatorname{Tr}[(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^*], \tag{2.7.8}$$

Using (2.7.8) we can write

$$L^3(F_n^{(c)}(x,z), \overline{F}_n(x,z)) \le \frac{1}{2n}\operatorname{Tr}[(\mathbf{V}^{(c)}(z) - \overline{\mathbf{V}}(z))(\mathbf{V}^{(c)}(z) - \overline{\mathbf{V}}(z))^*]$$
$$\le \frac{1}{n^2}\sum_{i\le j}|\mathbb{E}X_{ij}\mathbb{I}(|X_{ij}| \le \tau\sqrt{n})|^2 \le \frac{1}{n^2}\sum_{i\le j}\mathbb{E}X_{ij}^2\mathbb{I}(|X_{ij}| \ge \tau\sqrt{n})| \to 0.$$

In what follows assume from now that $|X_{ij}| \le \tau_n \sqrt{n}$, $\tau_n \to 0$ as $n \to \infty$. We also have that $\mathbb{E} X_{ij} = \mathbb{E} X_{ji} = 0$. One may also check that

$$\frac{1}{n^2} \sum_{i,j=1}^{n} |\mathbb{E} X_{ij}^2 - 1| \le 2 L_n(\tau_n). \tag{2.7.9}$$

2.7.2 Universality of the spectrum of singular values

In this section we prove that one can substitute the matrix $\mathbf{X}$ which satisfies (**C0**) by the matrix $\mathbf{Y}$ with Gaussian elements. Define the matrix $\mathbf{Z}(\varphi) = \mathbf{X} \cos\varphi + \mathbf{Y} \sin\varphi$ and introduce the following $2n \times 2n$ matrix

$$\mathbf{V}(\varphi) = \begin{pmatrix} \mathbf{O}_n & n^{-1/2}\mathbf{Z}(\varphi) \\ n^{-1/2}\mathbf{Z}^T(\varphi) & \mathbf{O}_n \end{pmatrix},$$

where $\mathbf{O}_n$ denotes $n \times n$ matrix with zero entries. The matrix $\mathbf{V}(0)$ corresponds to $\mathbf{V}$ from (2.6.1) and $\mathbf{V}(\pi/2)$ is

$$\mathbf{V}(\pi/2) = \begin{pmatrix} \mathbf{O}_n & n^{-1/2}\mathbf{Y} \\ n^{-1/2}\mathbf{Y}^T & \mathbf{O}_n \end{pmatrix},$$

Consider the matrix

$$\mathbf{V}(z,\varphi) := \mathbf{V}(\varphi) - \mathbf{J}(z).$$

Set

$$\mathbf{R}(\alpha, z, \varphi) := (\mathbf{V}(z,\varphi) - \alpha \mathbf{I}_{2n})^{-1}.$$

Introduce the following function

$$s_n(\alpha, z, \varphi) = \frac{1}{2n} \sum_{i=1}^{2n} \mathbb{E}[\mathbf{R}(\alpha, z, \varphi)]_{ii},$$

Note that $s_n(\alpha, z, 0)$ and $s_n(\alpha, z, \pi/2)$ are Stieltjes transforms of $\mathbf{V}(z, 0)$ and $\mathbf{V}(z, \pi/2)$ respectively.
Obviously we have

$$s_n(\alpha, z, \frac{\pi}{2}) - s_n(\alpha, z, 0) = \int_0^{\frac{\pi}{2}} \frac{\partial s_n(\alpha, z, \varphi)}{\partial \varphi} d\varphi. \tag{2.7.10}$$

To simplify the arguments we will omit arguments in the notations of matrices and Stieltjes transforms. We have

$$\frac{\partial \mathbf{V}}{\partial \varphi} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial Z_{ij}}{\partial \varphi} (e_i e_{j+n}^T + e_{j+n} e_i^T),$$

where we denote by e_i the column vector with 1 in the position i and zeros in the other positions. We may rewrite the integrand in (2.7.10) in the following way

$$\begin{aligned}\frac{\partial s_n}{\partial\varphi} &= -\frac{1}{2n}\operatorname{Tr}\mathbf{R}\frac{\partial\mathbf{V}}{\partial\varphi}\mathbf{R} \\ &= -\frac{1}{2n^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{Tr}\mathbf{R}\frac{\partial Z_{ij}}{\partial\varphi}e_i e_{j+n}^T\mathbf{R} \\ &- \frac{1}{2n^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{Tr}\mathbf{R}\frac{\partial Z_{ij}}{\partial\varphi}e_{j+n}e_i^T\mathbf{R} \\ &= \frac{1}{2n^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial Z_{ij}}{\partial\varphi}u_{ij} + \frac{1}{2n^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial Z_{ij}}{\partial\varphi}v_{ij},\end{aligned} \tag{2.7.11}$$

where $u_{ij} = -[\mathbf{R}^2]_{j+n,i}$ and $v_{ij} = -[\mathbf{R}^2]_{i,j+n}$. We estimate only the first term in (2.7.11), which we denote by

$$\mathbb{A} = \frac{1}{2n^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial Z_{ij}}{\partial\varphi}u_{ij}.$$

For all $1 \le i \le j \le n$ introduce the random variables

$$\xi_{ij} := Z_{ij}, \quad \hat{\xi}_{ij} := \frac{\partial Z_{ij}}{\partial\varphi} = -\sin\varphi X_{ij} + \cos\varphi Y_{ij},$$

Using Taylor's formula one may write

$$\begin{aligned}u_{ij}(\xi_{ij},\xi_{ji}) &= u_{ij}(0,0) + \xi_{ij}\frac{\partial u_{ij}}{\partial\xi_{ij}}(0,0) + \xi_{ji}\frac{\partial u_{ij}}{\partial\xi_{ji}}(0,0) \\ &+ \mathbb{E}_\theta(1-\theta)\xi_{ij}^2\frac{\partial^2 u_{ij}}{\partial\xi_{ij}^2}(\theta\xi_{ij},\theta\xi_{ji}) \\ &+ 2\mathbb{E}_\theta(1-\theta)\xi_{ij}\xi_{ji}\frac{\partial^2 u_{ij}}{\partial\xi_{ij}\partial\xi_{ji}}(\theta\xi_{ij},\theta\xi_{ji}) \\ &+ \mathbb{E}_\theta(1-\theta)\xi_{ji}^2\frac{\partial^2 u_{ij}}{\partial\xi_{ji}^2}(\theta\xi_{ij},\theta\xi_{ji}),\end{aligned}$$

where θ has a uniform distribution on $[0,1]$ and is independent of (ξ_{ij},ξ_{ji}). Multiplying both sides of the last equation by $\hat{\xi}_{ij}$ and taking mathematical expectation on both sides we have

$$\begin{aligned}
\mathbb{E}\hat{\xi}_{ij}u_{ij}(\xi_{ij},\xi_{ji}) &= \mathbb{E}\hat{\xi}_{ij}u_{ij}(0,0)\\
&+ \mathbb{E}\hat{\xi}_{ij}\xi_{ij}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,0) + \mathbb{E}\hat{\xi}_{ij}\xi_{ji}\frac{\partial u_{ij}}{\partial \xi_{ji}}(0,0)\\
&+ \mathbb{E}(1-\theta)\hat{\xi}_{ij}\xi_{ij}^2\frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2}(\theta\xi_{ij},\theta\xi_{ji})\\
&+ 2\mathbb{E}(1-\theta)\hat{\xi}_{ij}\xi_{ij}\xi_{ji}\frac{\partial^2 u_{ij}}{\partial \xi_{ij}\partial \xi_{ji}}(\theta\xi_{ij},\theta\xi_{ji})\\
&+ \mathbb{E}(1-\theta)\hat{\xi}_{ij}\xi_{ji}^2\frac{\partial^2 u_{ij}}{\partial \xi_{ji}^2}(\theta\xi_{ij},\theta\xi_{ji}).
\end{aligned}$$

It is easy to check that $\mathbb{E}\hat{\xi}_{ij}u_{ij}(0,0) = 0$.

Set

$$\begin{aligned}
A_{ij}^1 &:= \mathbb{E}\hat{\xi}_{ij}\xi_{ij}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,0), \quad A_{ij}^2 := \mathbb{E}\hat{\xi}_{ij}\xi_{ji}\frac{\partial u_{ij}}{\partial \xi_{ji}}(0,0),\\
A_{ij}^3 &:= \mathbb{E}(1-\theta)\hat{\xi}_{ij}\xi_{ij}^2\frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2}(\theta\xi_{ij},\theta\xi_{ji}),\\
A_{ij}^4 &:= 2\mathbb{E}(1-\theta)\hat{\xi}_{ij}\xi_{ij}\xi_{ji}\frac{\partial^2 u_{ij}}{\partial \xi_{ij}\partial \xi_{ji}}(\theta\xi_{ij},\theta\xi_{ji}),\\
A_{ij}^5 &:= \mathbb{E}(1-\theta)\hat{\xi}_{ij}\xi_{ji}^2\frac{\partial^2 u_{ij}}{\partial \xi_{ji}^2}(\theta\xi_{ij},\theta\xi_{ji}).
\end{aligned}$$

We can rewrite the term $\mathbb{A}$ in the following way $\mathbb{A} = \mathbb{A}_1 + ... + \mathbb{A}_5$, where

$$\mathbb{A}_k = \frac{1}{2n^{3/2}}\sum_{i,j=1}^n A_{ij}^k,$$

and $k = 1, ..., 5$.

We first estimate the term $\mathbb{A}_1$. The bound for the term $\mathbb{A}_2$ may be obtained in the similar way. It is easy to see that

$$\hat{\xi}_{ij}\xi_{ij} = -\frac{1}{2}\sin 2\varphi X_{ij}^2 + \cos^2\varphi X_{ij}Y_{ij} - \sin^2\varphi X_{ij}Y_{ij} + \frac{1}{2}\sin 2\varphi Y_{ij}^2.$$

The random variable Y_{ij} and the vector (X_{ij}, X_{ji}) are independent. Using this fact we conclude that

$$\mathbb{E}X_{ij}Y_{ij}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,0) = \mathbb{E}Y_{ij}\mathbb{E}X_{ij}\mathbb{E}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,0) = 0, \tag{2.7.12}$$

$$\mathbb{E}Y_{ij}^2\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,0) = \mathbb{E}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,0), \tag{2.7.13}$$

$$\mathbb{E}X_{ij}^2\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,0) = \mathbb{E}X_{ij}^2\mathbb{E}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,0). \tag{2.7.14}$$

A direct calculation shows that the derivative of $u_{ij} = -[\mathbf{R}^2]_{ji}$ is equal to

$$\begin{aligned}\frac{\partial u_{ij}}{\partial \xi_{ij}} &= \left[\mathbf{R}^2 \frac{\partial \mathbf{Z}}{\partial \xi_{ij}} \mathbf{R}\right]_{j+n,i} + \left[\mathbf{R} \frac{\partial \mathbf{Z}}{\partial \xi_{ij}} \mathbf{R}^2\right]_{j+n,i} \\ &= \frac{1}{\sqrt{n}}[\mathbf{R}^2 e_i e_{j+n}^T \mathbf{R}]_{j+n,i} + \frac{1}{\sqrt{n}}[\mathbf{R}^2 e_{j+n} e_i^T \mathbf{R}]_{j+n,i} \\ &+ \frac{1}{\sqrt{n}}[\mathbf{R} e_i e_{j+n}^T \mathbf{R}^2]_{j+n,i} + \frac{1}{\sqrt{n}}[\mathbf{R} e_{j+n} e_i^T \mathbf{R}^2]_{j+n,i} \\ &= \frac{1}{\sqrt{n}}[\mathbf{R}^2]_{j+n,i}[\mathbf{R}]_{j+n,i} + \frac{1}{\sqrt{n}}[\mathbf{R}^2]_{j+n,j+n}[\mathbf{R}]_{ii} \\ &+ \frac{1}{\sqrt{n}}[\mathbf{R}]_{j+n,i}[\mathbf{R}^2]_{j+n,i} + \frac{1}{\sqrt{n}}[\mathbf{R}]_{j+n,j+n}[\mathbf{R}^2]_{ii}.\end{aligned}$$

Using the obvious bound for the spectral norm of the matrix resolvent $||\mathbf{R}|| \le v^{-1}$ we get

$$\left|\frac{\partial u_{ij}}{\partial \xi_{ij}}\right| \le \frac{C}{\sqrt{n} v^3}. \tag{2.7.15}$$

From (2.7.12)–(2.7.15) and (2.7.9) we deduce

$$|\mathbb{A}_1| \le \frac{C}{v^3 n^2} \sum_{i,j=1}^n |\mathbb{E} X_{ij}^2 - 1| \le \frac{C L_n(\tau_n)}{v^3}.$$

Now we estimate the term $\mathbb{A}_3$. For the terms $\mathbb{A}_k, k = 4, 5$ it is straightforward to check that the same bounds hold. By the direct calculation on may show that the second derivative of $u_{ij} = -[\mathbf{R}^2]_{j+n,i}$ is equal to

$$\begin{aligned}\frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2} &= -\left[\mathbf{R}^2 \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R} \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R}\right]_{j+n,i} - \left[\mathbf{R} \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R}^2 \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R}\right]_{j+n,i} \\ &- \left[\mathbf{R} \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R} \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R}^2\right]_{j+n,i} = \mathbb{T}_1 + \mathbb{T}_2 + \mathbb{T}_3.\end{aligned}$$

Let's expand the term $\mathbb{T}_1$

$$\mathbb{T}_1 = -\left[\mathbf{R}^2 \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R} \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R}\right]_{j+n,i} = \mathbb{T}_{11} + \mathbb{T}_{12} + \mathbb{T}_{13} + \mathbb{T}_{14}, \tag{2.7.16}$$

where we denote

$$\begin{aligned}\mathbb{T}_{11} &= -\frac{1}{n}[\mathbf{R}^2]_{j+n,i}[\mathbf{R}]_{j+n,i}[\mathbf{R}]_{j+n,i}, \quad \mathbb{T}_{12} = -\frac{1}{n}[\mathbf{R}^2]_{j+n,i}[\mathbf{R}]_{j+n,j+n}[\mathbf{R}]_{ii}, \\ \mathbb{T}_{13} &= -\frac{1}{n}[\mathbf{R}^2]_{j+n,j+n}[\mathbf{R}]_{ii}[\mathbf{R}]_{j+n,i}, \quad \mathbb{T}_{14} = -\frac{1}{n}[\mathbf{R}^2]_{j+n,j+n}[\mathbf{R}]_{i,j+n}[\mathbf{R}]_{ii}.\end{aligned}$$

Using again the bound $||\mathbf{R}|| \le v^{-1}$ we can show that

$$\max(|\mathbb{T}_{11}|, |\mathbb{T}_{12}|, |\mathbb{T}_{13}|, |\mathbb{T}_{14}|) \le \frac{C}{n v^4}.$$

From the expansion (2.7.16) and the bounds of $\mathbb{T}_{1i}, i = 1, 2, 3, 4$ we conclude that

$$|\mathbb{T}_1| \le \frac{C}{nv^4}.$$

Repeating the above arguments one can show that

$$\max(|\mathbb{T}_2|, |\mathbb{T}_3|) \le \frac{C}{nv^4}.$$

Finally we have

$$\left| \frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2}(\theta\xi_{ij}, \theta\xi_{ji}) \right| \le \frac{C}{nv^4}.$$

Using the fact that $|\xi_{ij}| \le \tau_n \sqrt{n}$ we deduce the bound

$$|\mathbb{A}_3| \le \frac{C\tau_n}{v^4}.$$

From the bounds of $\mathbb{A}_k, k = 1, ..., 5$, it immediately follows that

$$|s_n(\alpha, z, \frac{\pi}{2}) - s_n(\alpha, z, 0)| \le \frac{C\tau_n}{v^4} + \frac{CL_n(\tau_n)}{v^3}.$$

We may turn τ_n to zero and conclude the statement of Theorem 2.7.2.

2.8 Some technical lemmas

Lemma 2.8.1. *Under the condition* (**C0**) *for* $\alpha = u + iv, v > 0$

$$\mathbb{E} \left| \frac{1}{n} \sum_{i=1}^n R_{ii}(\alpha, z) - \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n R_{ii}(\alpha, z) \right) \right|^2 \le \frac{C}{nv^2}.$$

Proof. To prove this lemma we will use Girko's method. Let $\mathbf{X}^{(j)}$ be the matrix $\mathbf{X}$ with the j-th row and column removed. Define the matrices $\mathbf{V}^{(j)}$ and $\mathbf{V}^{(j)}(z)$ as in (2.6.1) and $\mathbf{R}^{(j)}$ by (2.6.6). It is easy to see that

$$\operatorname{Rank}(\mathbf{V}(z) - \mathbf{V}^{(j)}(z)) = \operatorname{Rank}(\mathbf{VJ} - \mathbf{V}^{(j)}\mathbf{J}) \le 4.$$

Then

$$\frac{1}{n} |\operatorname{Tr}(\mathbf{V}(z) - \alpha\mathbf{I})^{-1} - \operatorname{Tr}(\mathbf{V}^{(j)}(z) - \alpha\mathbf{I})^{-1}| \le \frac{\operatorname{Rank}(\mathbf{V}(z) - \mathbf{V}^{(j)}(z))}{nv} \le \frac{4}{nv}. \quad (2.8.1)$$

We introduce the family of σ-algebras $\mathcal{F}_i = \sigma\{X_{j,k}, j, k > i\}$ and conditional mathematical expectation $\mathbb{E}_i = \mathbb{E}(\cdot|\mathcal{F}_i)$ with respect to this σ-algebras. We can write

$$\frac{1}{n} \operatorname{Tr} \mathbf{R} - \frac{1}{n} \mathbb{E} \operatorname{Tr} \mathbf{R} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_i \operatorname{Tr} \mathbf{R} - \mathbb{E}_{i-1} \operatorname{Tr} \mathbf{R} = \sum_{i=1}^n \gamma_i.$$

The sequence $(\gamma_i, \mathcal{F}_i)_{i\geq 1}$ is a martingale difference. By (2.8.1)

$$|\gamma_i| = \frac{1}{n}|\mathbb{E}_i(\operatorname{Tr}\mathbf{R} - \operatorname{Tr}\mathbf{R}^{(i)}) - \mathbb{E}_{i-1}(\operatorname{Tr}\mathbf{R} - \operatorname{Tr}\mathbf{R}^{(i)})| \\ \leq |\mathbb{E}_i(\operatorname{Tr}\mathbf{R} - \operatorname{Tr}\mathbf{R}^{(i)})| + |\mathbb{E}_{i-1}(\operatorname{Tr}\mathbf{R} - \operatorname{Tr}\mathbf{R}^{(i)})| \leq \frac{C}{vn}. \tag{2.8.2}$$

From the Burkholder inequality for martingale difference (see [38])

$$\mathbb{E}\left|\sum_{i=1}^{n}\gamma_i\right|^2 \leq K_2\mathbb{E}\left(\sum_{i=1}^{n}|\gamma_i|^2\right)$$

and (2.8.2) it follows

$$\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}R_{ii}(\alpha,z) - \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}R_{ii}(\alpha,z)\right)\right|^2 \leq K_2\mathbb{E}(\sum_{i=1}^{n}|\gamma_i|^2) \leq K_2\frac{C}{nv^2}.$$

□

Lemma 2.8.2. *Under the condition* (**C0**) *for* $\alpha = u + iv, v > 0$

$$\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^{n}R_{i,i+n}(\alpha,z) - \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}R_{i,i+n}(\alpha,z)\right)\right|^2 \leq \frac{C}{nv^4}.$$

Proof. As in Lemma 2.8.1 we introduce the matrices $\mathbf{V}^{(j)}$ and $\mathbf{R}^{(j)}$. We have

$$\mathbf{V}\mathbf{J} = \mathbf{V}^{(j)}\mathbf{J} + e_je_j^T\mathbf{V}\mathbf{J} + \mathbf{V}\mathbf{J}e_je_j^T + e_{j+n}e_{j+n}^T\mathbf{V}\mathbf{J} + \mathbf{V}\mathbf{J}e_{j+n}e_{j+n}^T$$

By the resolvent equality $\mathbf{R} - \mathbf{R}^{(j)} = -\mathbf{R}^{(j)}(\mathbf{V}(z) - \mathbf{V}^{(j)}(z))\mathbf{R}$

$$\frac{1}{n}\sum_{k=1}^{n}(\mathbf{R}_{k,k+n} - \mathbf{R}^{(j)}_{k,k+n}) = \\ = \frac{1}{n}\sum_{k=1}^{n}[\mathbf{R}^{(j)}(e_je_j^T\mathbf{V}\mathbf{J} + e_{j+n}e_{j+n}^T\mathbf{V}\mathbf{J} + \mathbf{V}\mathbf{J}e_je_j^T + \mathbf{V}\mathbf{J}e_{j+n}e_{j+n}^T)\mathbf{R}]_{k,k+n} \\ = \mathbb{T}_1 + \mathbb{T}_2 + \mathbb{T}_3 + \mathbb{T}_4.$$

Let us consider the first term. The arguments for other terms are similar.

$$\sum_{k=1}^{n}[\mathbf{R}^{(j)}e_je_j^T\mathbf{V}\mathbf{J}\mathbf{R}]_{k,k+n} = \operatorname{Tr}\mathbf{R}^{(j)}e_je_{j+n}^T\mathbf{V}\mathbf{J}\mathbf{R} = \sum_{i=1}^{2n}[\mathbf{R}^{(j)}\mathbf{R}]_{ij}[e_je_{j+n}^T\mathbf{V}\mathbf{J}]_{ji}.$$

From $\max(||\mathbf{R}^{(j)}||, ||\mathbf{R}||) \leq v^{-1}$ and the Hölder inequality it follows that

$$\mathbb{E}\left|\sum_{k=1}^{n}[\mathbf{R}^{(j)}e_je_j^T\mathbf{V}\mathbf{J}\mathbf{R}]_{k,k+n}\right|^2 \leq \frac{C}{v^4}.$$

By the similar arguments as in Lemma 2.8.1 we can conclude the statement of the Lemma. □

Lemma 2.8.3. *Under the condition* (**C0**) *for* $\alpha = u + iv, v > 0$

$$\frac{1}{n^{3/2}} \sum_{j,k=1}^{n} \mathbb{E} X_{jk} R_{k+n,j} =$$
$$= \frac{1}{n^2} \sum_{j,k=1}^{n} \mathbb{E} \left[\frac{\partial \mathbf{R}}{\partial X_{jk}} \right]_{k+n,j} + \frac{\rho}{n^2} \sum_{j,k=1}^{n} \mathbb{E} \left[\frac{\partial \mathbf{R}}{\partial X_{kj}} \right]_{k+n,j} + r_n(\alpha, z),$$

where

$$|r_n(\alpha, z)| \le \frac{C}{\sqrt{n} v^3}.$$

Proof. By Taylor's formula

$$\mathbb{E} X f(X,Y) = f(0,0)\mathbb{E}X + f'_x(0,0)\mathbb{E}X^2 + f'_y(0,0)\mathbb{E}XY \quad (2.8.3)$$
$$+ \mathbb{E}(1-\theta)[X^3 f''_{xx}(\theta X, \theta Y) + 2X^2 Y f''_{xy}(\theta X, \theta Y) + XY^2 f''_{yy}(\theta X, \theta Y)]$$

and

$$\mathbb{E} f'_x(X,Y) = f'_x(0,0) + \mathbb{E}(1-\theta)[X f''_{xx}(\theta X, \theta Y) + Y f''_{xy}(\theta X, \theta Y)] \quad (2.8.4)$$
$$\mathbb{E} f'_y(X,Y) = f'_y(0,0) + \mathbb{E}(1-\theta)[X f''_{xy}(\theta X, \theta Y) + Y f''_{yy}(\theta X, \theta Y)],$$

where θ has a uniform distribution on $[0,1]$. From (2.8.3) and (2.8.4) for $j \ne k$

$$\left| \mathbb{E} X_{jk} R_{k+n,j} - \mathbb{E} \left[\frac{\partial \mathbf{R}}{\partial X_{jk}} \right]_{k+n,j} - \rho \mathbb{E} \left[\frac{\partial \mathbf{R}}{\partial X_{kj}} \right]_{k+n,j} \right| \le$$
$$\le (|X_{jk}|^3 + |X_{jk}|) \left| \left[\frac{\partial^2 \mathbf{R}}{\partial X_{jk}^2} (\theta X_{jk}, \theta X_{kj}) \right]_{k+n,j} \right|$$
$$+ (|X_{kj}|^2 |X_{jk}| + |X_{kj}|) \left| \left[\frac{\partial^2 \mathbf{R}}{\partial X_{kj}^2} (\theta X_{jk}, \theta X_{kj}) \right]_{k+n,j} \right|$$
$$+ (2|X_{jk}|^2 |X_{kj}| + |X_{jk}| + |X_{kj}|) \left| \left[\frac{\partial^2 \mathbf{R}}{\partial X_{jk} \partial X_{kj}} (\theta X_{jk}, \theta X_{kj}) \right]_{k+n,j} \right|.$$

Let us consider the first term in the sum. The bounds for the second and third terms can be obtained by the similar arguments. We have

$$\frac{\partial^2 \mathbf{R}}{\partial X_{jk}^2} = \frac{1}{n} \mathbf{R}(e_j e_{n+k}^T + e_{n+k} e_j^T) \mathbf{R} (e_j e_{n+k}^T + e_{n+k} e_j^T) \mathbf{R} \quad = \mathbb{P}_1 + \mathbb{P}_2 + \mathbb{P}_3 + \mathbb{P}_4,$$

where

$$\mathbb{P}_1 = \frac{1}{n} \mathbf{R} e_j e_{n+k}^T \mathbf{R} e_j e_{n+k}^T \mathbf{R}$$
$$\mathbb{P}_2 = \frac{1}{n} \mathbf{R} e_j e_{n+k}^T \mathbf{R} e_{n+k} e_j^T \mathbf{R}$$
$$\mathbb{P}_3 = \frac{1}{n} \mathbf{R} e_{n+k} e_j^T \mathbf{R} e_j e_{n+k}^T \mathbf{R}$$
$$\mathbb{P}_4 = \frac{1}{n} \mathbf{R} e_{n+k} e_j^T \mathbf{R} e_{n+k} e_j^T \mathbf{R}.$$

From $|\mathbf{R}_{i,j}| \leq v^{-1}$ it follows that

$$\frac{1}{n^{5/2}} \sum_{j,k=1}^{n} \mathbb{E}|X_{jk}|^{\alpha} |[\mathbb{P}_i]_{n+k,j}| \leq \frac{C}{\sqrt{n} v^3}$$

for $\alpha = 1, 3$ and $i = 1, ..., 4$.

For $j = k$

$$\frac{1}{n^2} \left| \sum_{j=1}^{n} \mathbb{E} \left[\frac{\partial \mathbf{R}}{\partial X_{jj}} \right]_{j+n,j} \right| = \frac{1}{n^2} \sum_{j=1}^{n} (\mathbb{E} R_{j+n,j}^2 + \mathbb{E}|R_{j,j} R_{j+n,j+n}|) \leq \frac{C}{n v^2}.$$

Hence we can add this term to the sum

$$\frac{\rho}{n^2} \sum_{\substack{j,k=1 \\ j \neq k}}^{n} \mathbb{E} \left[\frac{\partial \mathbf{R}}{\partial X_{kj}} \right]_{k+n,j}.$$

$\square$

CHAPTER 3

Semicircle law for a class of random matrices with dependent entries

In this chapter we study ensembles of random symmetric matrices and consider symmetric matrices $\mathbf{X}_n = \{X_{ij}\}_{i,j=1}^n$ with a random field type dependence, such that $\mathbb{E}X_{ij} = 0$, $\mathbb{E}X_{ij}^2 = \sigma_{ij}^2$, where σ_{ij} may be different numbers. Assuming that the average of the normalized sums of variances in each row converges to one and Lindeberg condition holds true we prove that the empirical spectral distribution of eigenvalues converges to Wigner's semicircle law.

3.1 Introduction

Let $X_{jk}, 1 \le j \le k < \infty$, be triangular array of random variables with $\mathbb{E}X_{jk} = 0$ and $\mathbb{E}X_{jk}^2 = \sigma_{jk}^2$, and let $X_{jk} = X_{kj}$ for $1 \le j < k < \infty$. We consider the random matrix

$$\mathbf{X}_n = \{X_{jk}\}_{j,k=1}^n.$$

Denote by $\lambda_1 \le ... \le \lambda_n$ eigenvalues of matrix $n^{-1/2}\mathbf{X}_n$ and define its spectral distribution function by

$$\mathcal{F}^{\mathbf{X}_n}(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}(\lambda_i \le x),$$

where $\mathbb{I}(B)$ denotes the indicator of an event B. We set $F^{\mathbf{X}_n}(x) := \mathbb{E}\mathcal{F}^{\mathbf{X}_n}(x)$. Let $g(x)$ and $G(x)$ denote the density and the distribution function of the standard semicircle law

$$g(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\mathbb{I}(|x| \le 2), \quad G(x) = \int_{-\infty}^x g(u)du.$$

Introduce the σ-algebras

$$\mathfrak{F}^{(i,j)} := \sigma\{X_{kl} : 1 \le k \le l \le n, (k,l) \ne (i,j)\}, 1 \le i \le j \le n.$$

For any $\tau > 0$ we introduce Lindeberg's ratio for random matrices as

$$L_n(\tau) := \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}|X_{ij}|^2\mathbb{I}(|X_{ij}| \ge \tau\sqrt{n}).$$

We assume that the following conditions hold

$$\mathbb{E}(X_{ij}|\mathfrak{F}^{(i,j)}) = 0; \tag{3.1.1}$$

$$\frac{1}{n^2}\sum_{i,j=1}^{n} \mathbb{E}|\mathbb{E}(X_{ij}^2|\mathfrak{F}^{(i,j)}) - \sigma_{ij}^2| \to 0 \text{ as } n \to \infty; \tag{3.1.2}$$

$$\text{for any fixed } \tau > 0 \quad L_n(\tau) \to 0 \text{ as } n \to \infty. \tag{3.1.3}$$

Furthermore, we will use condition (3.1.3) not only for the matrix $\mathbf{X}_n$, but for other matrices as well, replacing X_{ij} in the definition of Lindeberg's ratio by corresponding elements.

For all $1 \le i \le n$ let $B_i^2 := \frac{1}{n}\sum_{j=1}^n \sigma_{ij}^2$. We need to impose additional conditions on the variances σ_{ij}^2 given by

$$\frac{1}{n}\sum_{i=1}^{n} |B_i^2 - 1| \to 0 \text{ as } n \to \infty; \tag{3.1.4}$$

$$\max_{1 \le i \le n} B_i \le C, \tag{3.1.5}$$

where C is some absolute constant.

Remark 3.1.1. *It is easy to see that the conditions* (3.1.4) *and* (3.1.5) *follow from the following condition*

$$\max_{1 \le i \le n} \left|B_i^2 - 1\right| \to 0 \text{ as } n \to \infty. \tag{3.1.6}$$

The main result of the paper is the following theorem

Theorem 3.1.2. *Let $\mathbf{X}_n$ satisfy conditions* (3.1.1)–(3.1.5). *Then*

$$\sup_x |F^{\mathbf{X}_n}(x) - G(x)| \to 0 \text{ as } n \to \infty.$$

Let us fix i, j. It is easy to see that for all $(k,l) \ne (i,j)$

$$\mathbb{E}X_{ij}X_{kl} = \mathbb{E}\mathbb{E}(X_{ij}X_{kl}|\mathfrak{F}^{(i,j)})) = \mathbb{E}X_{kl}\mathbb{E}(X_{ij}|\mathfrak{F}^{(i,j)}) = 0.$$

Hence the elements of the matrix $\mathbf{X}_n$ are uncorrelated. If we additionally assume that the elements of the matrix $\mathbf{X}_n$ are independent random variables then conditions (3.1.1) and (3.1.2) are automatically satisfied. The following Theorem 3.1.2 follows immediately in the case when the matrix $\mathbf{X}_n$ has independent entries.

Theorem 3.1.3. *Assume that the elements X_{ij} of the matrix $\mathbf{X}_n$ are independent for all $1 \le i \le j \le n$ and $\mathbb{E}X_{ij} = 0$, $\mathbb{E}X_{ij}^2 = \sigma_{ij}^2$. Assume that $\mathbf{X}_n$ satisfies conditions* (3.1.3)–(3.1.5). *Then*

$$\sup_x |F^{\mathbf{X}_n}(x) - G(x)| \to 0 \text{ as } n \to \infty.$$

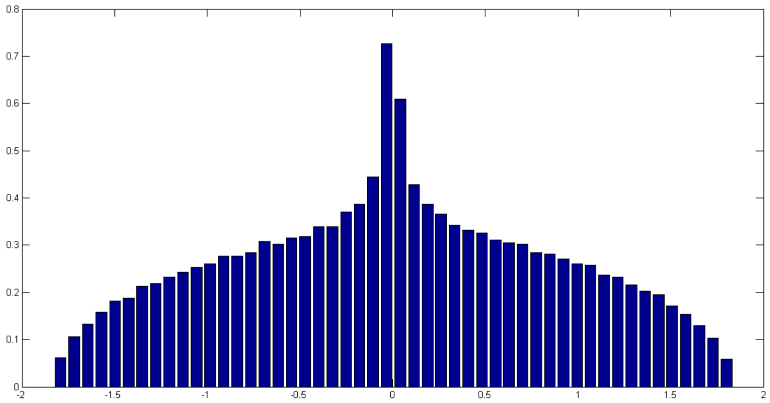

Figure 3.1: Spectrum of matrix $\mathbf{X}_n$.

The following example illustrates that without condition (3.1.4) convergence to Wigner's semicircle law doesn't hold.

Example 3.1.4. Let $\mathbf{X}_n$ denote a block matrix

$$\mathbf{X}_n = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{pmatrix},$$

where $\mathbf{A}$ is $m \times m$ symmetric random matrix with Gaussian elements with zero mean and unit variance, $\mathbf{B}$ is $m \times (n-m)$ random matrix with i.i.d. Gaussian elements with zero mean and unit variance. Furthermore, let $\mathbf{D}$ be a $(n-m) \times (n-m)$ diagonal matrix with Gaussian random variables on the diagonal with zero mean and unit variance. If we set $m := n/2$ then it is not difficult to check that condition (3.1.4) doesn't hold. We simulated the spectrum of the matrix $\mathbf{X}_n$ and illustrated a limiting distribution on Figure 3.1.

Remark 3.1.5. *We conjecture that Theorem 3.1.2 (Theorem 3.1.3 respectively) holds without assumption* (3.1.5)*.*

Define the Levy distance between the distribution functions F_1 and F_2 by

$$L(F_1, F_2) = \inf\{\varepsilon > 0 : F_1(x - \varepsilon) - \varepsilon \le F_2(x) \le F_1(x + \varepsilon) + \varepsilon\}.$$

The following theorem formulates Lindeberg's universality scheme for random matrices.

Theorem 3.1.6. *Let $\mathbf{X}_n, \mathbf{Y}_n$ denote independent symmetric random matrices with $\mathbb{E}X_{ij} = \mathbb{E}Y_{ij} = 0$ and $\mathbb{E}X_{ij}^2 = \mathbb{E}Y_{ij}^2 = \sigma_{ij}^2$. Suppose that the matrix $\mathbf{X}_n$ satisfies conditions* (3.1.1)–(3.1.4)*, and the matrix $\mathbf{Y}_n$ has independent Gaussian elements. Additionally assume that for the matrix $\mathbf{Y}_n$ conditions* (3.1.3) *and* (3.1.4) *hold. Then*

$$L(F^{\mathbf{X}_n}(x), F^{\mathbf{Y}_n}(x)) \to 0 \text{ as } n \to \infty.$$

In view of Theorem 3.1.6 to prove Theorem 3.1.2 it remains to show convergence to semicircle law in the Gaussian case.

Theorem 3.1.7. *Assume that the entries Y_{ij} of the matrix $\mathbf{Y}_n$ are independent for all $1 \le i \le j \le n$ and have Gaussian distribution with $\mathbb{E}Y_{ij} = 0$, $\mathbb{E}Y_{ij}^2 = \sigma_{ij}^2$. Assume that conditions (3.1.3)–(3.1.5) are satisfied. Then*

$$\sup_x |F^{\mathbf{Y}_n}(x) - G(x)| \to 0 \text{ as } n \to \infty.$$

Remark 3.1.8. *For related ensembles of random covariance matrices it is well known that spectral distribution function of eigenvalues converges to the Marchenko–Pastur law. In this case Götze and Tikhomirov in [23] received similar results to [25]. Recently Adamczak, [1], proved the Marchenko–Pastur law for matrices with martingale structure. He assumed that the matrix elements have moments of all orders and imposed conditions similar to (3.1.4). Another class of random matrices with dependent entries was considered in [33] by O'Rourke.*

From now on we shall omit the index n in the notation for random matrices.

3.2 Proof of Theorem 3.1.6

We denote the Stieltjes transforms of $F^{\mathbf{X}}$ and $F^{\mathbf{Y}}$ by $S^{\mathbf{X}}(z)$ and $S^{\mathbf{Y}}(z)$ respectively. Due to the relations between distribution functions and Stieltjes transforms, the statement of Theorem 3.1.6 will follow from

$$|S^{\mathbf{X}}(z) - S^{\mathbf{Y}}(z)| \to 0 \text{ as } n \to \infty. \tag{3.2.1}$$

Set

$$\mathbf{R}^{\mathbf{X}}(z) := \left(\frac{1}{\sqrt{n}}\mathbf{X} - z\mathbf{I}\right)^{-1} \text{ and } \mathbf{R}^{\mathbf{Y}}(z) := \left(\frac{1}{\sqrt{n}}\mathbf{Y} - z\mathbf{I}\right)^{-1}.$$

By definition

$$S^{\mathbf{X}}(z) = \frac{1}{n}\operatorname{Tr}\mathbb{E}\mathbf{R}^{\mathbf{X}}(z) \text{ and } S^{\mathbf{Y}}(z) = \frac{1}{n}\operatorname{Tr}\mathbb{E}\mathbf{R}^{\mathbf{Y}}(z).$$

We divide the proof of (3.2.1) into the two subsections 3.2.1 and 3.2.2.

Note that we can substitute τ in (3.1.3) by a decreasing sequence τ_n tending to zero such that

$$L_n(\tau_n) \to 0 \text{ as } n \to \infty. \tag{3.2.2}$$

and $\lim_{n\to\infty} \tau_n \sqrt{n} = \infty$.

3.2.1 Truncation of random variables

In this section we truncate the elements of the matrices $\mathbf{X}$ and $\mathbf{Y}$. Let us omit the indices $\mathbf{X}$ and $\mathbf{Y}$ in the notations of the resolvent and the Stieltjes transforms.

Consider some symmetric $n \times n$ matrix $\mathbf{D}$. Put $\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{D}$. Let

$$\tilde{\mathbf{R}} = \left(\frac{1}{\sqrt{n}}\tilde{\mathbf{X}} - z\mathbf{I}\right)^{-1}.$$

Lemma 3.2.1.

$$|\operatorname{Tr}\mathbf{R} - \operatorname{Tr}\tilde{\mathbf{R}}| \le \frac{1}{v^2}(\operatorname{Tr}\mathbf{D}^2)^{\frac{1}{2}}.$$

Proof. By the resolvent equation

$$\mathbf{R} = \tilde{\mathbf{R}} - \frac{1}{\sqrt{n}}\mathbf{R}\mathbf{D}\tilde{\mathbf{R}}. \tag{3.2.3}$$

For resolvent matrices we have, for $z = u + iv, v > 0$,

$$\max\{||\mathbf{R}||, ||\tilde{\mathbf{R}}||\} \le \frac{1}{v}. \tag{3.2.4}$$

Using (3.2.3) and (3.2.4) it is easy to show that

$$|\operatorname{Tr}\mathbf{R} - \operatorname{Tr}\tilde{\mathbf{R}}| = \frac{1}{\sqrt{n}}|\operatorname{Tr}\mathbf{R}\mathbf{D}\tilde{\mathbf{R}}| \le \frac{1}{v^2}(\operatorname{Tr}\mathbf{D}^2)^{\frac{1}{2}}.$$

□

We split the matrix entries as $X = \hat{X} + \check{X}$, where $\hat{X} := X\mathbb{I}(|X| < \tau_n\sqrt{n})$ and $\check{X} := X\mathbb{I}(|X| \ge \tau_n\sqrt{n})$. Define the matrix $\hat{\mathbf{X}} = \{\hat{X}_{ij}\}_{i,j=1}^n$. Let

$$\hat{\mathbf{R}}(z) := \left(\frac{1}{\sqrt{n}}\hat{\mathbf{X}} - z\mathbf{I}\right)^{-1} \text{ and } \hat{S}(z) = \frac{1}{n}\mathbb{E}\operatorname{Tr}\hat{\mathbf{R}}(z).$$

By Lemma 3.2.1

$$|S(z) - \hat{S}(z)| \le \frac{1}{v^2}\left(\frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}X_{ij}^2\mathbb{I}(|X_{ij}| \ge \tau_n\sqrt{n})\right)^{1/2} = v^{-2}L_n^{\frac{1}{2}}(\tau_n).$$

From (3.2.2) we conclude that

$$|S(z) - \hat{S}(z)| \to 0 \text{ as } n \to \infty.$$

Introduce the centralized random variables $\overline{X}_{ij} = \hat{X}_{ij} - \mathbb{E}(\hat{X}_{ij}|\mathfrak{F}^{(i,j)})$ and the matrix $\overline{\mathbf{X}} = \{\overline{X}_{ij}\}_{i,j=1}^n$. Let

$$\overline{\mathbf{R}}(z) := \left(\frac{1}{\sqrt{n}}\overline{\mathbf{X}} - z\mathbf{I}\right)^{-1} \text{ and } \overline{S}(z) = \frac{1}{n}\mathbb{E}\operatorname{Tr}\overline{\mathbf{R}}(z).$$

Again by Lemma 3.2.1

$$|\hat{S}(z)-\overline{S}(z)| \le \frac{1}{v^2}\left(\frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}X_{ij}^2\mathbb{I}(|X_{ij}|\ge \tau_n\sqrt{n})\right)^{1/2} = v^{-2}L_n^{\frac{1}{2}}(\tau_n).$$

In view of (3.2.2) the right hand side tends to zero as $n\to\infty$.

Now we show that (3.1.2) will hold if we replace $\mathbf{X}$ by $\overline{\mathbf{X}}$. For all $1\le i\le j\le n$

$$\begin{aligned}&\mathbb{E}(\overline{X}_{ij}^2|\mathfrak{F}^{(i,j)})-\sigma_{ij}^2\\ &=\mathbb{E}(\overline{X}_{ij}^2|\mathfrak{F}^{(i,j)})-\mathbb{E}(\hat{X}_{ij}^2|\mathfrak{F}^{(i,j)})-\mathbb{E}(\check{X}_{ij}^2|\mathfrak{F}^{(i,j)})+\mathbb{E}(X_{ij}^2|\mathfrak{F}^{(i,j)})-\sigma_{ij}^2.\end{aligned}$$

By the triangle inequality and (3.1.2), (3.2.2)

$$\begin{aligned}&\frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}|\mathbb{E}(\overline{X}_{ij}^2|\mathfrak{F}^{(i,j)})-\sigma_{ij}^2| \\ &\le \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}|\mathbb{E}(X_{ij}^2|\mathfrak{F}^{(i,j)})-\sigma_{ij}^2| + 2L_n(\tau_n)\to 0 \text{ as } n\to\infty.\end{aligned} \tag{3.2.5}$$

It is also not very difficult to check that condition (3.1.4) holds true for the matrix $\mathbf{X}$ replaced by $\overline{\mathbf{X}}$.
Similarly, one may truncate the elements of the matrix $\mathbf{Y}$ and consider the matrix $\overline{\mathbf{Y}}$ with the entries $Y_{ij}\mathbb{I}(|Y_{ij}|\le\tau_n\sqrt{n})$. Then one may check that

$$\frac{1}{n^2}\sum_{i,j=1}^n |\mathbb{E}\overline{Y}_{ij}^2-\sigma_{ij}^2|\to 0 \text{ as } n\to\infty. \tag{3.2.6}$$

In what follows assume from now on that $|X_{ij}|\le\tau_n\sqrt{n}$ and $|Y_{ij}|\le\tau_n\sqrt{n}$. We shall write $\mathbf{X},\mathbf{Y}$ instead of $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ respectively.
In what follows assume from now on that $|X_{ij}|\le\tau_n\sqrt{n}$ and $|Y_{ij}|\le\tau_n\sqrt{n}$.

3.2.2 Universality of the spectrum of eigenvalues

To prove (3.2.1) we will use a method introduced in [5]. Define the matrix $\mathbf{Z}:=\mathbf{Z}(\varphi):=\mathbf{X}\cos\varphi+\mathbf{Y}\sin\varphi$. It is easy to see that $\mathbf{Z}(0)=\mathbf{X}$ and $\mathbf{Z}(\pi/2)=\mathbf{Y}$. Set $\mathbf{W}:=\mathbf{W}(\varphi):=n^{-1/2}\mathbf{Z}$ and

$$\mathbf{R}(z,\varphi):=(\mathbf{W}-z\mathbf{I})^{-1}.$$

Introduce the Stieltjes transform

$$S(z,\varphi):=\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\mathbf{R}(z,\varphi)]_{ii}.$$

Note that $S(z,0)$ and $S(z,\pi/2)$ are the Stieltjes transforms $S^{\mathbf{X}}(z)$ and $S^{\mathbf{Y}}(z)$ respectively.

Obviously we have

$$S(z,\frac{\pi}{2}) - S(z,0) = \int_0^{\frac{\pi}{2}} \frac{\partial S(z,\varphi)}{\partial \varphi} d\varphi. \tag{3.2.7}$$

To simplify the arguments we will omit arguments in the notations of matrices and Stieltjes transforms. We have

$$\frac{\partial \mathbf{W}}{\partial \varphi} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial Z_{ij}}{\partial \varphi} e_i e_j^T,$$

where we denote by e_i the column vector with 1 in position i and zeros in the other positions. We may rewrite the integrand in (3.2.7) in the following way

$$\begin{aligned}\frac{\partial S}{\partial \varphi} &= -\frac{1}{n}\mathbb{E}\operatorname{Tr}\mathbf{R}\frac{\partial \mathbf{W}}{\partial \varphi}\mathbf{R} \\ &= -\frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\operatorname{Tr}\mathbf{R}\frac{\partial Z_{ij}}{\partial \varphi} e_i e_j^T \mathbf{R} \\ &= \frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\frac{\partial Z_{ij}}{\partial \varphi} u_{ij},\end{aligned} \tag{3.2.8}$$

where $u_{ij} = -[\mathbf{R}^2]_{ji}$.
For all $1 \le i \le j \le n$ introduce the random variables

$$\xi_{ij} := Z_{ij}, \quad \hat{\xi}_{ij} := \frac{\partial Z_{ij}}{\partial \varphi} = -\sin\varphi X_{ij} + \cos\varphi Y_{ij},$$

and the sets of random variables

$$\xi^{(ij)} := \{\xi_{kl} : 1 \le k \le l \le n, (k,l) \ne (i,j)\}.$$

Using Taylor's formula one may write

$$u_{ij}(\xi_{ij},\xi^{(ij)}) = u_{ij}(0,\xi^{(ij)}) + \xi_{ij}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,\xi^{(ij)}) + \mathbb{E}_\theta \theta(1-\theta)\xi_{ij}^2 \frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2}(\theta\xi_{ij},\xi^{(ij)}),$$

where θ has a uniform distribution on $[0,1]$ and is independent of $(\xi_{ij},\xi^{(ij)})$. Multiplying both sides of the last equation by $\hat{\xi}_{ij}$ and taking mathematical expectation on both sides we have

$$\begin{aligned}\mathbb{E}\hat{\xi}_{ij}u_{ij}(\xi_{ij},\xi^{(ij)}) &= \mathbb{E}\hat{\xi}_{ij}u_{ij}(0,\xi^{(ij)}) + \mathbb{E}\hat{\xi}_{ij}\xi_{ij}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,\xi^{(ij)}) \\ &+ \mathbb{E}\theta(1-\theta)\hat{\xi}_{ij}\xi_{ij}^2\frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2}(\theta\xi_{ij},\xi^{(ij)}).\end{aligned} \tag{3.2.9}$$

By independence of Y_{ij} and $\xi^{(ij)}$ we get

$$\mathbb{E}Y_{ij}u_{ij}(0,\xi^{(ij)}) = \mathbb{E}Y_{ij}\mathbb{E}u_{ij}(0,\xi^{(ij)}) = 0. \tag{3.2.10}$$

By the properties of conditional expectation and condition (3.1.1)

$$\mathbb{E}X_{ij}u_{ij}(0,\xi^{(ij)}) = \mathbb{E}u_{ij}(0,\xi^{(ij)})\mathbb{E}(X_{ij}|\mathfrak{F}^{(i,j)}) = 0. \tag{3.2.11}$$

By (3.2.9), (3.2.10) and (3.2.11) we can rewrite (3.2.8) in the following way

$$\begin{aligned}\frac{\partial S}{\partial \varphi} &= \frac{1}{n^2}\sum_{i,j=1}^{n}\mathbb{E}\hat{\xi}_{ij}\xi_{ij}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,\xi^{(ij)}) + \frac{1}{n^2}\sum_{i,j=1}^{n}\mathbb{E}\theta(1-\theta)\hat{\xi}_{ij}\xi_{ij}^2\frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2}(\theta\xi_{ij},\xi^{(ij)})\\ &= \mathbb{A}_1 + \mathbb{A}_2.\end{aligned}$$

It is easy to see that

$$\hat{\xi}_{ij}\xi_{ij} = -\frac{1}{2}\sin 2\varphi X_{ij}^2 + \cos^2\varphi X_{ij}Y_{ij} - \sin^2\varphi X_{ij}Y_{ij} + \frac{1}{2}\sin 2\varphi Y_{ij}^2.$$

The random variables Y_{ij} are independent of X_{ij} and $\xi^{(ij)}$. Using this fact we conclude that

$$\mathbb{E}X_{ij}Y_{ij}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,\xi^{(ij)}) = \mathbb{E}Y_{ij}\mathbb{E}X_{ij}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,\xi^{(ij)}) = 0, \tag{3.2.12}$$

$$\mathbb{E}Y_{ij}^2\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,\xi^{(ij)}) = \sigma_{ij}^2\mathbb{E}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,\xi^{(ij)}). \tag{3.2.13}$$

By the properties of conditional mathematical expectation we get

$$\mathbb{E}X_{ij}^2\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,\xi^{(ij)}) = \mathbb{E}\frac{\partial u_{ij}}{\partial \xi_{ij}}(0,\xi^{(ij)})\mathbb{E}(X_{ij}^2|\mathfrak{F}^{(i,j)}). \tag{3.2.14}$$

A direct calculation shows that the derivative of $u_{ij} = -[\mathbf{R}^2]_{ji}$ is equal to

$$\begin{aligned}\frac{\partial u_{ij}}{\partial \xi_{ij}} &= \left[\mathbf{R}^2\frac{\partial \mathbf{Z}}{\partial \xi_{ij}}\mathbf{R}\right]_{ji} + \left[\mathbf{R}\frac{\partial \mathbf{Z}}{\partial \xi_{ij}}\mathbf{R}^2\right]_{ji}\\ &= \frac{1}{\sqrt{n}}[\mathbf{R}^2 e_i e_j^T\mathbf{R}]_{ji} + \frac{1}{\sqrt{n}}[\mathbf{R}^2 e_j e_i^T\mathbf{R}]_{ji} + \frac{1}{\sqrt{n}}[\mathbf{R}e_i e_j^T\mathbf{R}^2]_{ji} + \frac{1}{\sqrt{n}}[\mathbf{R}e_j e_i^T\mathbf{R}^2]_{ji}\\ &= \frac{1}{\sqrt{n}}[\mathbf{R}^2]_{ji}[\mathbf{R}]_{ji} + \frac{1}{\sqrt{n}}[\mathbf{R}^2]_{jj}[\mathbf{R}]_{ii} + \frac{1}{\sqrt{n}}[\mathbf{R}]_{ji}[\mathbf{R}^2]_{ji} + \frac{1}{\sqrt{n}}[\mathbf{R}]_{jj}[\mathbf{R}^2]_{ii}.\end{aligned}$$

Using the obvious bound for the spectral norm of the matrix resolvent $||\mathbf{R}|| \le v^{-1}$ we get

$$\left|\frac{\partial u_{ij}}{\partial \xi_{ij}}\right| \le \frac{C}{\sqrt{n}v^3}. \tag{3.2.15}$$

From (3.2.12)–(3.2.15) and (3.2.5)–(3.2.6) we deduce

$$|\mathbb{A}_1| \le \frac{C}{n^2 v^3}\sum_{i,j=1}^{n}\mathbb{E}|\mathbb{E}(X_{ij}^2|\mathfrak{F}^{(i,j)}) - \sigma_{ij}^2| \to 0 \text{ as } n \to \infty. \tag{3.2.16}$$

It remains to estimate $\mathbb{A}_2$. We calculate the second derivative of u_{ij}

$$\frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2} = -\left[\mathbf{R}^2 \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R} \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R}\right]_{ji} - \left[\mathbf{R} \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R}^2 \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R}\right]_{ji}$$
$$- \left[\mathbf{R} \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R} \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R}^2\right]_{ji} = \mathbb{T}_1 + \mathbb{T}_2 + \mathbb{T}_3.$$

Let's expand the term $\mathbb{T}_1$

$$\mathbb{T}_1 = -\left[\mathbf{R}^2 \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R} \frac{\partial \mathbf{V}}{\partial \xi_{ij}} \mathbf{R}\right]_{ji} = \mathbb{T}_{11} + \mathbb{T}_{12} + \mathbb{T}_{13} + \mathbb{T}_{14}, \tag{3.2.17}$$

where we denote

$$\mathbb{T}_{11} = -\frac{1}{n}[\mathbf{R}^2]_{ji}[\mathbf{R}]_{ji}[\mathbf{R}]_{ji}, \quad \mathbb{T}_{12} = -\frac{1}{n}[\mathbf{R}^2]_{ji}[\mathbf{R}]_{jj}[\mathbf{R}]_{ii},$$
$$\mathbb{T}_{13} = -\frac{1}{n}[\mathbf{R}^2]_{jj}[\mathbf{R}]_{ii}[\mathbf{R}]_{ji}, \quad \mathbb{T}_{14} = -\frac{1}{n}[\mathbf{R}^2]_{jj}[\mathbf{R}]_{ij}[\mathbf{R}]_{ii}.$$

Using again the bound $||\mathbf{R}|| \le v^{-1}$ we can show that

$$\max(|\mathbb{T}_{11}|, |\mathbb{T}_{12}|, |\mathbb{T}_{13}|, |\mathbb{T}_{14}|) \le \frac{C}{nv^4}.$$

From the expansion (3.2.17) and the bounds of $\mathbb{T}_{1i}, i = 1, 2, 3, 4$ we conclude that

$$|\mathbb{T}_1| \le \frac{C}{nv^4}.$$

Repeating the above arguments one can show that

$$\max(|\mathbb{T}_2|, |\mathbb{T}_3|) \le \frac{C}{nv^4}.$$

Finally we have

$$\left| \frac{\partial^2 u_{ij}}{\partial \xi_{ij}^2} (\theta \xi_{ij}, \xi^{(ij)}) \right| \le \frac{C}{nv^4}.$$

Using the assumption $|\xi_{ij}| \le \tau_n \sqrt{n}$ and the condition (3.1.4) we deduce the bound

$$|\mathbb{A}_2| \le \frac{C\tau_n}{v^4}. \tag{3.2.18}$$

We may turn τ_n to zero and conclude the statement of Theorem 3.1.6 from (3.2.7), (3.2.8), (3.2.16) and (3.2.18).

3.3 Proof of Theorem 3.1.7

We prove the theorem using the moment method. It is easy to see that the moments of $F^{\mathbf{Y}}(x)$ can be rewritten as normalized traces of powers of $\mathbf{Y}$:

$$\int_{\mathbb{R}} x^k dF^{\mathbf{Y}}(x) = \mathbb{E} \frac{1}{n} \operatorname{Tr} \left(\frac{1}{\sqrt{n}} \mathbf{Y} \right)^k.$$

It is sufficient to prove that

$$\mathbb{E}\frac{1}{n}\operatorname{Tr}\left(\frac{1}{\sqrt{n}}\mathbf{Y}\right)^k = \int_{\mathbb{R}} x^k dG(x) + o_k(1).$$

for $k \geq 1$, where $o_k(1)$ tends to zero as $n \to \infty$ for any fixed k.
It is well known that the moments of semicircle law are given by the Catalan numbers

$$\beta_k = \int_{\mathbb{R}} x^k dG(x) = \begin{cases} \frac{1}{m+1}\binom{2m}{m}, & k = 2m \\ 0, & k = 2m+1. \end{cases}$$

Furthermore we shall use the notations and the definitions from [4]. A graph is a triple (E, V, F), where E is the set of edges, V is the set of vertices, and F is a function, $F : E \to V \times V$. Let $\boldsymbol{i} = (i_1, ..., i_k)$ be a vector taking values in $\{1, ..., n\}^k$. For a vector $\boldsymbol{i}$ we define a Γ-graph as follows. Draw a horizontal line and plot the numbers $i_1, ..., i_l$ on it. Consider the distinct numbers as vertices, and draw k edges e_j from i_j to $i_{j+1}, j = 1, ..., k$, using $i_{k+1} = i_1$ by convention. Denote the number of distinct i_j's by t. Such a graph is called a $\Gamma(k, t)$-graph.

Two $\Gamma(k, t)$-graphs are said to be isomorphic if they can be converted each other by a permutation of $(1, ..., n)$. By this definition, all Γ-graphs are classified into isomorphism classes. We shall call the $\Gamma(k, t)$-graph canonical if it has the following properties:
1) Its vertex set is $\{1, ..., t\}$;
2) Its edge set is $\{e_1, ..., e_k\}$;
3) There is a function g from $\{1, ..., k\}$ onto $\{1, ..., t\}$ satisfying $g(1) = 1$ and $g(i) \leq \max\{g(1), ..., g(i-1)\} + 1$ for $1 < i \leq k$;
4) $F(e_i) = (g(i), g(i+1))$, for $i = 1, ..., k$, with the convention $g(k+1) = g(1) = 1$.

It is easy to see that each isomorphism class contains one and only one canonical Γ-graph that is associated with a function g, and a general graph in this class can be defined by $F(e_j) = (i_{g(j)}, i_{g(j+1)})$. It is easy to see that each isomorphism class contains $n(n-1)...(n-t+1)$ $\Gamma(k, t)$-graphs.

We shall classify all canonical graphs into three categories. Category 1 consists of all canonical $\Gamma(k, t)$-graphs with the property that each edge is coincident with exactly one other edge of opposite direction and the graph of noncoincident edges forms a tree. It is easy to see if k is odd then there are no graphs in category 1. If k is even, i.e. $k = 2m$, say, we denote a $\Gamma(k, t)$-graph by $\Gamma_1(2m)$. Category 2 consists of all canonical graphs that have at least one edge with odd multiplicity. We shall denote this category by $\Gamma_2(k, t)$. Finally, category 3 consists of all other canonical graphs, which we denote by $\Gamma_3(k, t)$.

It is known, see [4, Lemma 2.4], that the number of $\Gamma_1(2m)$-graphs is equal to $\frac{1}{m+1}\binom{2m}{m}$.

We expand the traces of powers of $\mathbf{Y}$ in a sum

$$\mathrm{Tr}\left(\frac{1}{\sqrt{n}}\mathbf{Y}\right)^k = \frac{1}{n^{k/2}} \sum_{i_1,i_2,\ldots,i_k} Y_{i_1 i_2} Y_{i_2 i_3} \ldots Y_{i_k i_1}, \tag{3.3.1}$$

where the summation is taken over all sequences $\boldsymbol{i} = (i_1, \ldots, i_k) \in \{1, \ldots, n\}$.

For each vector $\boldsymbol{i}$ we construct a graph $G(\boldsymbol{i})$ as above. We denote by $Y(\boldsymbol{i}) = Y(G(\boldsymbol{i}))$.

Then we may split the moments of $F^{\mathbf{Y}}(x)$ into three terms

$$\mathbb{E}\frac{1}{n}\mathrm{Tr}\left(\frac{1}{\sqrt{n}}\mathbf{Y}\right)^k = \frac{1}{n^{k/2+1}} \sum_{\boldsymbol{i}} \mathbb{E} Y_{i_1 i_2} Y_{i_2 i_3} \ldots Y_{i_k i_1} = S_1 + S_2 + S_3,$$

where

$$S_j = \frac{1}{n^{k/2+1}} \sum_{\Gamma(k,t) \in C_j} \sum_{G(\boldsymbol{i}) \in \Gamma(k,t)} \mathbb{E}[Y(G(\boldsymbol{i}))],$$

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

From the independence of Y_{ij} and $\mathbb{E}Y_{ij}^{2s-1} = 0, s \geq 1$, it follows that $S_2 = 0$.

For the graphs from categories C_1 and C_3 we introduce further notations. Let us consider the $\Gamma(k,t)$-graph $G(\boldsymbol{i})$. Without loss of generality we assume that $i_l, l = 1, \ldots, t$ are distinct coordinates of the vector $\boldsymbol{i}$ and define a vector $\boldsymbol{i}_t = (i_1, \ldots, i_t)$. We also set $G(\boldsymbol{i}_t) := G(\boldsymbol{i})$. Let $\tilde{\boldsymbol{i}}_t = (i_1, \ldots, i_{q-1}, i_{q+1}, \ldots, i_t)$ and $\hat{\boldsymbol{i}}_t = (i_1, \ldots, i_{p-1}, i_{p+1}, \ldots, i_{q-1}, i_{q+1}, \ldots, i_t)$ be vectors derived from $\boldsymbol{i}_t$ by deleting the elements in the position q and p, q respectively. We denote the graph without the vertex i_q and all edges linked to it by $G(\tilde{\boldsymbol{i}}_t)$. If the vertex i_q is incident to a loop we denote by $G'(\boldsymbol{i}_t)$ the graph with this loop removed.

Now we will estimate the term S_3. For a graph from category C_3 we know that k has to be even, i.e. $k = 2m$, say. We illustrate the example of a $\Gamma_3(k,t)$-graph in Figure 3.2. This graph corresponds to the term $Y(G(\boldsymbol{i}_3)) = Y_{i_1 i_1}^2 Y_{i_1 i_2}^2 Y_{i_2 i_3}^4 Y_{i_3 i_3}^2$. We mention that $\mathbb{E}Y_{i_p i_q}^{2s} \leq_s \sigma_{i_p i_q}^{2s}$. Hence we may rewrite the terms which correspond to the graphs from category C_3 via variances.

In each graph from category C_3 there is at least one vertex incident to a loop with multiplicity greater or equal to two or an edge with multiplicity greater then two. It is possible as well that both cases occur.

Suppose that there is a vertex, let's say i_1, which is incident to a loop with multiplicity $s \geq 1$. It remains to consider the remaining $2(m-s)$ edges. We will consequently delete edges and vertices from the graph using the following algorithm:

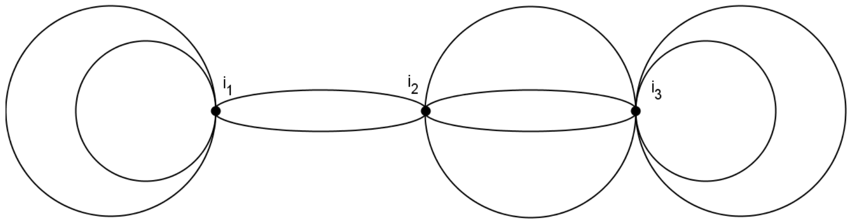

Figure 3.2: Graph $\Gamma_3(10,3)$.

1. If the number of distinct vertices is equal to one we should go to step (3). Otherwise, we take a vertex, let's say $i_q, q \neq 1$, such that there are no new edges starting from it. We also take a vertex i_p which is connected to i_q by an edge of this graph. There are three possibilities:

   (a) There is a loop incident to vertex i_q with multiplicity $2a, a \geq 1$. In this case we estimate

   $$\frac{1}{n^{m+1}} \sum_{G(\boldsymbol{i}) \in \Gamma(2m,t)} \mathbb{E}[Y(G(\boldsymbol{i}))] \leq_a \frac{1}{n^{m+1}} \sum_{\boldsymbol{i}_t} \mathbb{E}[Y(G'(\boldsymbol{i}_t))] \sigma_{i_q i_q}^{2a}.$$

   Applying the inequality $n^{-1}\sigma_{i_q i_q}^2 \leq B_{i_q}^2 \leq C^2$, a times we delete all loops incident to this vertex;

   (b) There is no loop incident to i_q, but the multiplicity of the edge from i_p to i_q is equal to 2. In this case we estimate

   $$\frac{1}{n^{m+1}} \sum_{G(\boldsymbol{i}) \in \Gamma(2m,t)} \mathbb{E}[Y(G(\boldsymbol{i}))] \leq \frac{1}{n^{m+1}} \sum_{\tilde{\boldsymbol{i}}_t} \mathbb{E}[Y(G(\tilde{\boldsymbol{i}}_t))] \sum_{i_q=1}^{n} \sigma_{i_p i_q}^2.$$

   We may delete the vertex i_q and the two coinciding edges from i_p to i_q using condition (3.1.5);

   (c) There is no loop incident to i_q, but the multiplicity of the edge from i_p to i_q is equal to $2b, b > 1$. In this case we estimate

   $$\frac{1}{n^{m+1}} \sum_{G(\boldsymbol{i}) \in \Gamma(2m,t)} \mathbb{E}[Y(G(\boldsymbol{i}))] \leq_b \frac{1}{n^{m+1}} \sum_{\tilde{\boldsymbol{i}}_t} \mathbb{E}[Y(G(\tilde{\boldsymbol{i}}_t))] \sum_{i_q=1}^{n} \sigma_{i_p i_q}^{2b}.$$

   Here we may use the inequality $n^{-1}\sigma_{i_p i_q}^2 \leq B_{i_p}^2 \leq C^2$, $b-1$ times and consequently delete all coinciding edges except two. Then we may apply (b);

2. go to step (1);

3. in this step one may use the bound

$$\frac{1}{n^{s+1}}\sum_{i_1=1}^{n}\mathbb{E}Y_{i_1i_1}^{2s} \le_s \frac{1}{n^{s+1}}\sum_{i_1=1}^{n}\sigma_{i_1i_1}^{2s} \le \frac{C^{2(s-1)}}{n^2}\sum_{i_1=1}^{n}\sigma_{i_1i_1}^{2}$$
$$\le C^{2(s-1)}\tau_n^2 + \frac{C^{2(s-1)}}{n^2}\sum_{i_1=1}^{n}\mathbb{E}Y_{i_1i_1}^2\mathbb{I}(|Y_{i_1i_1}| \ge \tau_n\sqrt{n})$$
$$\le C^{2(s-1)}\tau_n^2 + C^{2(s-1)}L_n(\tau_n) = o_s(1),$$

where we have used the inequality $n^{-1}\sigma_{i_1i_1}^2 \le B_{i_1}^2 \le C^2$.

Applying this algorithm we get the bound

$$\frac{1}{n^{m+1}}\sum_{G(\boldsymbol{i})\in\Gamma(2m,t)}\mathbb{E}[Y(G(\boldsymbol{i}))] \le_m C^{2(m-1)}(\tau_n^2 + L_n(\tau_n)) = o_m(1).$$

If there are no loops, but just an edge with multiplicity greater then two, then we can apply the above procedure again and use in the step (3) the following bound for $s \ge 2$

$$\frac{1}{n^{s+1}}\sum_{\substack{i_1,i_2=1\\ i_1\ne i_2}}^{n}\mathbb{E}Y_{i_1i_2}^{2s} \le_s \frac{1}{n^{s+1}}\sum_{i_1,i_2=1}^{n}\sigma_{i_1i_2}^{2s} \le \frac{C^{2(s-2)}}{n^3}\sum_{i_1,i_2=1}^{n}\sigma_{i_1i_2}^{4}$$
$$\le C^{2(s-2)}\frac{\tau_n^2}{n^2}\sum_{i_1,i_2=1}^{n}\sigma_{i_1i_2}^2 + \frac{C^{2(s-2)}}{n^3}\sum_{i_1,i_2=1}^{n}\sigma_{i_1i_2}^2\mathbb{E}Y_{i_1i_2}^2\mathbb{I}(|Y_{i_1i_2}| \ge \tau_n\sqrt{n})$$
$$\le C^{2(s-1)}\tau_n^2 + C^{2(s-1)}L_n(\tau_n) = o_s(1),$$

where we have used the inequality $n^{-1}\sigma_{i_1i_2}^2 \le B_{i_1}^2 \le C^2$ and (3.1.5).

As an example we recommend to check this algorithm for the graph in Figure 3.2.

It is easy to see that the number of different canonical graphs in C_3 is of order $O_m(1)$. Finally for the term S_3 we get

$$S_3 = o_m(1).$$

It remains to consider the term S_1. For a graph from category C_1 we know that k has to be even, i.e. $k = 2m$, say. In the category C_1 using the notations of $\boldsymbol{i}_t, \tilde{\boldsymbol{i}}_t, \hat{\boldsymbol{i}}_t, \boldsymbol{j}_t, \tilde{\boldsymbol{j}}_t$ and $\hat{\boldsymbol{j}}_t$ we take $t = m+1$.
We illustrate on the left part of Figure 3.3 an example of the tree of noncoincident edges of a $\Gamma_1(2m)$-graph for $m = 5$. The term corresponding to this tree is $Y(G(\boldsymbol{i}_6)) = Y_{i_1i_2}^2 Y_{i_2i_3}^2 Y_{i_2i_4}^2 Y_{i_1i_5}^2 Y_{i_5i_6}^2$.

We denote by $\sigma^2(\boldsymbol{i}_{m+1}) = \sigma^2(G(\boldsymbol{i}_{m+1}))$ the product of m numbers $\sigma_{i_si_t}^2$, where $i_s, i_t, s < t$ are vertices of the graph $G(\boldsymbol{i}_{m+1})$ connected by edges of this graph. In

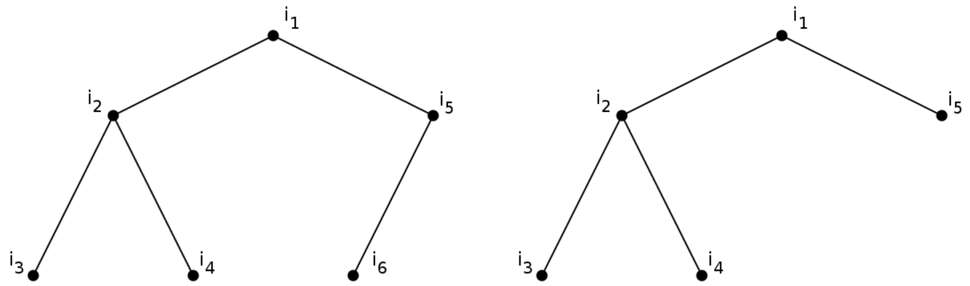

Figure 3.3: On the left, the tree of noncoincident edges of a $\Gamma_1(10)$-graph is shown. On the right, the tree of noncoincident edges of a $\Gamma_1(10)$-graph with deleted leaf i_6 is shown .

our example, $\sigma^2(\boldsymbol{i}_{m+1}) = \sigma^2(\boldsymbol{i}_6) = \sigma^2_{i_1 i_2}\sigma^2_{i_2 i_3}\sigma^2_{i_2 i_4}\sigma^2_{i_1 i_5}\sigma^2_{i_5 i_6}$.

If $m = 1$ then $\sigma^2(\boldsymbol{i}_2) = \sigma^2_{i_1 i_2}$ and

$$\frac{1}{n^2}\sum_{\substack{i_1,i_2=1\\ i_1\neq i_2}}^{n} \sigma^2_{i_1 i_2} = \frac{1}{n}\sum_{i_1=1}^{n}\left[\frac{1}{n}\sum_{i_2=1}^{n}\sigma^2_{i_1 i_2} - 1\right] + 1 + o(1), \tag{3.3.2}$$

where we have used $n^{-2}\sum_{i_1=1}^{n}\sigma^2_{i_1 i_1} = o(1)$. By (3.1.4) the first term is of order $o(1)$. The number of canonical graphs in C_1 for $m = 1$ is equal to 1. We conclude for $m = 1$ that

$$S_1 = n^{-2}\sum_{\Gamma_1(2)}\sum_{\substack{i_1,i_2=1\\ i_1\neq i_2}}^{n}\sigma^2_{i_1 i_2} = 1 + o(1),$$

Now we assume that $m > 1$. Then we can find a leaf in the tree, let's say i_q, and a vertex i_p, which is connected to i_q by an edge of this tree. We have $\sigma^2(\boldsymbol{i}_{m+1}) = \sigma^2(\tilde{\boldsymbol{i}}_{m+1})\cdot\sigma^2_{i_p i_q}$, where $\sigma^2(\tilde{\boldsymbol{i}}_{m+1}) = \sigma^2(G(\tilde{\boldsymbol{i}}_{m+1}))$.

In our example we can take the leaf i_6. On the right part of Figure 3.3 we have drawn the tree with deleted leaf i_6. We have $\sigma^2_{i_p i_q} = \sigma^2_{i_5 i_6}$ and $\sigma^2(\tilde{\boldsymbol{i}_6}) = \sigma^2_{i_1 i_2}\sigma^2_{i_2 i_3}\sigma^2_{i_2 i_4}\sigma^2_{i_1 i_5}$.

It is easy to see that

$$\frac{1}{n^{m+1}}\sum_{\boldsymbol{i}_{m+1}}\sigma^2(\boldsymbol{i}_{m+1}) = \frac{1}{n^{m+1}}\sum_{\tilde{\boldsymbol{i}}_{m+1}}\sigma^2(\tilde{\boldsymbol{i}}_{m+1})\sum_{i_q=1}^{n}\sigma^2_{i_p i_q} + o_m(1)$$

$$= \frac{1}{n^m}\sum_{\tilde{\boldsymbol{i}}_{m+1}}\sigma^2(\tilde{\boldsymbol{i}}_{m+1})\left[\frac{1}{n}\sum_{i_q=1}^{n}\sigma^2_{i_p i_q} - 1\right] \tag{3.3.3}$$

$$+ \frac{1}{n^m}\sum_{\tilde{\boldsymbol{i}}_{m+1}}\sigma^2(\tilde{\boldsymbol{i}}_{m+1}) \tag{3.3.4}$$

$$+ o_m(1),$$

where we have added some graphs from category C_3 and used the similar bounds as for the term S_3. Now we will show that the term (3.3.3) is of order $o_m(1)$. Note that

$$\frac{1}{n^m}\sum_{\tilde{\boldsymbol{i}}_{m+1}}\sigma^2(\tilde{\boldsymbol{i}}_{m+1})\left|\frac{1}{n}\sum_{i_q=1}^{n}\sigma^2_{i_p i_q} - 1\right| \tag{3.3.5}$$

$$= \frac{1}{n}\sum_{i_p=1}^{n}\left|\frac{1}{n}\sum_{i_q=1}^{n}\sigma^2_{i_p i_q} - 1\right|\frac{1}{n^{m-1}}\sum_{\hat{\boldsymbol{i}}_{m+1}}\sigma^2(\tilde{\boldsymbol{i}}_{m+1}).$$

We can sequentially delete leafs from the tree and using (3.1.5) write the bound

$$\frac{1}{n^{m-1}}\sum_{\hat{\boldsymbol{i}}_{m+1}}\sigma^2(\tilde{\boldsymbol{i}}_{m+1}) \le C^{2(m-1)}. \tag{3.3.6}$$

By (3.3.6) and (3.1.4) we have shown that (3.3.3) is of order $o_m(1)$. For the second term (3.3.4), i.e.

$$\frac{1}{n^m}\sum_{\tilde{\boldsymbol{i}}_{m+1}}\sigma^2(\tilde{\boldsymbol{i}}_{m+1})$$

we can repeat the above procedure and stop if we arrive at only two vertices in the tree. It the last step we can use the result (3.3.2). Finally we get

$$S_1 = \frac{1}{n^{m+1}}\sum_{\Gamma_1(2m)}\sum_{\boldsymbol{i}_{m+1}}\sigma^2(\boldsymbol{i}_{m+1}) = \frac{1}{m+1}\binom{2m}{m} + o_m(1),$$

which proves Theorem 3.1.7.

CHAPTER 4

Strong law of large numbers for random processes

In this chapter for martingales with continuous parameter we provide sufficient conditions for the strong law of large numbers and prove analogs of the Kolmogorov, Zygmund–Marcinkiewicz, and Brunk–Prokhrov strong laws of large numbers. A new generalization of the the Brunk–Prokhorov strong law of large numbers is given for martingales with discrete times. Along with the almost surely convergence, we also prove the convergence in average .

4.1 Extension of the Brunk–Prokhorov theorem

In works [27] and [16] generalizations of the Brunk–Prokhorov theorem are given. What follows below is a new generalization of the Brunk-Prokhorov theorem. We can take arbitrary positive numbers as normalizing constants, if they form an unboundedly increasing sequence. Generality of normalizing constants is achieved by imposing an additional condition on the random variables. In some cases, this condition is automatically satisfied. In particular, it is satisfied under condition of the original Brunk–Prokhorov theorem.

Theorem 4.1.1. *Let $\{Y_n, n \in \mathbb{N} = \{1,2,\dots\}\}$ be a martingale relative to filtration $\{\mathcal{F}_n, n \in \mathbb{N}\}$ and $b_n, n \in \mathbb{N}$ be a sequence of unboundedly increasing positive numbers. Assume $Y_0 = 0$. If*

$$\sum_{n=1}^{\infty} \frac{n^{\alpha-1}\mathbb{E}|Y_n - Y_{n-1}|^{2\alpha}}{b_n^{2\alpha}} < \infty, \tag{4.1.1}$$

$$\sum_{n=1}^{\infty} \frac{n^{\alpha-2}\sum_{k=1}^{n}\mathbb{E}|Y_k - Y_{k-1}|^{2\alpha}}{b_n^{2\alpha}} < \infty \tag{4.1.2}$$

for some $\alpha \geq 1$, then

$$\lim_{n\to\infty} \frac{Y_n}{b_n} = 0 \text{ a.s. and } \lim_{n\to\infty} \mathbb{E}\left|\frac{\max_{1\leq k\leq n} Y_k}{b_n}\right|^{2\alpha} = 0. \tag{4.1.3}$$

Proof. If $\alpha = 1$, then condition (4.1.2) follows from condition (4.1.1). From condition (4.1.1) and the Kronecker lemma [30, p. 252] it follows that $\lim_{n\to\infty} \mathbb{E}|Y_n/b_n|^2 = 0$. This and the Doob inequality for moments

$$\mathbb{E}(\max_{1\leq k\leq n} |Y_k|)^2 \leq 4\mathbb{E}|Y_n|^2$$

imply the second statement of (4.1.3). Textbook [30, p. 407] contains proof that condition (4.1.1) implies the first condition of (4.1.3). Below, we assume that $\alpha > 1$. Using the generalization of the Doob inequality in [38, p. 687], we obtain for any $\varepsilon > 0$

$$\begin{aligned} \varepsilon^{2\alpha} P\{\sup_{k\geq n} \frac{|Y_k|}{b_k} > \varepsilon\} &= \lim_{m\to\infty} \varepsilon^{2\alpha} P\{\sup_{n\leq k\leq m} \frac{|Y_k|}{b_k} > \varepsilon\} \\ &\leq \frac{\mathbb{E}|Y_n|^{2\alpha}}{b_n^{2\alpha}} + \sum_{k=n+1}^{\infty} \frac{\mathbb{E}|Y_k|^{2\alpha} - E|Y_{k-1}|^{2\alpha}}{b_k^{2\alpha}}. \end{aligned} \tag{4.1.4}$$

Let us prove the second statement of (4.1.3). Thanks to the Doob inequality for moments, it is sufficient to show that $\lim_{n\to\infty} \mathbb{E}|Y_n/b_n|^{2\alpha} = 0$. According to the Burkholder inequality [10, p. 396], there exists a constant C_α, such that

$$\mathbb{E}|Y_n|^{2\alpha} \leq C_\alpha n^{\alpha-1} \sum_{k=1}^{n} \mathbb{E}|Y_k - Y_{k-1}|^{2\alpha}.$$

This and the Holder inequality imply that

$$\mathbb{E}|Y_n|^{2\alpha} \leq C_\alpha n^{\alpha-1} \sum_{k=1}^{n} \mathbb{E}|Y_k - Y_{k-1}|^{2\alpha}.$$

It is sufficient to prove that

$$\lim_{n\to\infty} \frac{n^{\alpha-1}}{b_n^{2\alpha}} \sum_{k=1}^{n} \mathbb{E}|Y_k - Y_{k-1}|^{2\alpha} = 0. \tag{4.1.5}$$

According to the Kronecker lemma, it follows from condition (4.1.2) that

$$\lim_{n\to\infty} \frac{1}{b_n^{2\alpha}} \sum_{k=1}^{n} k^{\alpha-1} \mathbb{E}|Y_k - Y_{k-1}|^{2\alpha} = 0. \tag{4.1.6}$$

We denote by A_n the sum in (4.1.6) and $c_0 = 0$,

$$c_k = \mathbb{E}|Y_1 - Y_0|^{2\alpha} + \cdots + \mathbb{E}|Y_k - Y_{k-1}|^{2\alpha}.$$

Using the Abel transformation, we can write the sum in the following form:

$$A_n = \sum_{k=1}^{n} k^{\alpha-1} \mathbb{E}|Y_k - Y_{k-1}|^{2\alpha} = n^{\alpha-1} c_n + \sum_{k=1}^{n-1} (k^{\alpha-1} - (k+1)^{\alpha-1}) c_k.$$

Thanks to the inequality $(k+1)^{\alpha-1} - k^{\alpha-1} \leq 2\alpha k^{\alpha-2}$ the following estimation is valid:

$$A_n \geq n^{\alpha-1} c_n - 2\alpha \sum_{k=1}^{n-1} k^{\alpha-2} c_k.$$

We denote the latter sum by B_n. According to the Kronecker lemma, condition (4.1.2) implies that $\lim_{n\to\infty} B_n/b_n^{2\alpha} = 0$. From this and (4.1.6) it follows that

$$0 = \lim_{n\to\infty} A_n/b_n^{2\alpha} \geq \limsup_{n\to\infty} n^{\alpha-1}c_n/b_n^{2\alpha} - \lim_{n\to\infty} 2\alpha B_n/b_n^{2\alpha} = \limsup_{n\to\infty} n^{\alpha-1}c_n/b_n^{2\alpha}.$$

Statement (4.1.5) is proved.

Let us prove that

$$\lim_{n\to\infty} \sum_{k=n+1}^{\infty} \frac{\mathbb{E}|Y_k|^{2\alpha} - \mathbb{E}|Y_{k-1}|^{2\alpha}}{b_k^{2\alpha}} = 0. \tag{4.1.7}$$

The series can be estimated as follows:

$$\sum_{k=n+1}^{\infty} \frac{\mathbb{E}|Y_k|^{2\alpha} - \mathbb{E}|Y_{k-1}|^{2\alpha}}{b_k^{2\alpha}} = -\frac{\mathbb{E}|Y_n|^{2\alpha}}{b_{n+1}^{2\alpha}} + \sum_{k=n+1}^{\infty} \Big(\frac{1}{b_k^{2\alpha}} - \frac{1}{b_{k+1}^{2\alpha}}\Big)\mathbb{E}|Y_k|^{2\alpha}$$

$$\leq C_\alpha \sum_{k=n+1}^{\infty} \Big(\frac{1}{b_k^{2\alpha}} - \frac{1}{b_{k+1}^{2\alpha}}\Big) k^{\alpha-1} \sum_{j=1}^{k} \mathbb{E}|Y_j - Y_{j-1}|^{2\alpha}$$

$$= C_\alpha \frac{(n+1)^{\alpha-1}}{b_{n+1}^{2\alpha}} \sum_{k=1}^{n+1} \mathbb{E}|Y_k - Y_{k-1}|^{2\alpha} + C_\alpha \sum_{k=n+2}^{\infty} \frac{k^{\alpha-1}\mathbb{E}|Y_k - Y_{k-1}|^{2\alpha}}{b_k^{2\alpha}}$$

$$+ C_\alpha \sum_{k=n+2}^{\infty} \frac{(k^{\alpha-1} - (k-1)^{\alpha-1})}{b_k^{2\alpha}} \sum_{j=1}^{k-1} \mathbb{E}|Y_j - Y_{j-1}|^{2\alpha}.$$

Let us recall, that $k^{\alpha-1} - (k-1)^{\alpha-1} \leq 2\alpha k^{\alpha-2}$. Thanks to these estimations, conditions (4.1.1), (4.1.2) and statement (4.1.5), (4.1.7) are fulfilled.

The second statement in (4.1.3), (4.1.4) and (4.1.7) implies that sequence $\sup_{k\geq n} |Y_k|/b_n$ converges by probability to zero as $n \to \infty$. It decreases monotonically and thus converges almost everywhere to zero. □

Remark 4.1.2. *In work [16] it was shown that condition (4.1.2) follows from condition (4.1.1), if for some $\delta > (\alpha-1)/(2\alpha)$ the ratio b_n/n^δ increases with an increase of $n \in \mathbb{N}$. In particular, condition (4.1.2) follows from condition (4.1.1) at $b_n = n$ for all $n \in \mathbb{N}$. The theorem was proved by Chow with this choice of normalizing constant [10, p. 397].*

4.2 Strong law of large numbers for martingales with continuous parameter

Theorem 4.2.1. *Let $\{Y_t, t \in \mathbb{R}_+\}$ be a measurable separable martingale relative to filtration $\{\mathcal{F}_t, t \in \mathbb{R}_+\}$ and $f(t), t \geq 0$, be an unboundedly increasing positive function. If*

$$\int_1^\infty \frac{d\mathbb{E}|Y_t|^\alpha}{f^\alpha(t)} < \infty \tag{4.2.1}$$

for some $\alpha \geq 1$*, then*

$$\lim_{t\to\infty} \frac{\sup\limits_{0\leq s\leq t} Y_s}{f(t)} = 0 \ a.s. \tag{4.2.2}$$

Proof. We may assume that function $f(t), t \geq 1$, is right continues. In the opposite case, it can be replaced by function $f(t+0) = \lim\limits_{s\downarrow t} f(s), t \geq 1$. It is easy to see that it is right continuous and almost everywhere coincides with f for the Lebesgue measure. Note that the function $g(t) = \inf\{s : f(s) > t\}, t \geq 1$, is right continuous, unboundedly increases, and satisfies the inequalities $2^{n+1} \geq f(g(2^n)) \geq 2^n$ for all $n \in \mathbb{N} = \{1, 2, \dots\}$, beginning with some number n_0. below, we will assume that $n_0 = 1$. By the Doob inequality [13, p. 285] for moments, we obtain

$$\varepsilon^\alpha \sum_{n=1}^{\infty} \mathbb{P} \sup_{g(2^n)\leq t\leq g(2^{n+1})} \frac{|Y_t|}{f(t)} > \varepsilon \leq \sum_{n=1}^{\infty} \frac{\mathbb{E}|Y_{g(2^{n+1})}|^\alpha}{f^\alpha(g(2^n))}. \tag{4.2.3}$$

The series on the right is convergent. This is due to (4.2.1) and the following relations:

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}|Y_{g(2^{n+1})}|^\alpha}{2^{\alpha n}} = \sum_{n=1}^{\infty} \frac{1}{2^{\alpha n}} \Big(\sum_{k=1}^{n} (\mathbb{E}|Y_{g(2^{k+1})}|^\alpha - \mathbb{E}|Y_{g(2^k)}|^\alpha) + \mathbb{E}|Y_{g(2)}|^\alpha \Big)$$

$$= C + \sum_{k=1}^{\infty} \int\limits_{g(2^k)}^{g(2^{k+1})} d\mathbb{E}|Y_t|^\alpha \sum_{n=k}^{\infty} \frac{1}{2^{\alpha n}} = C + C_1 \sum_{k=1}^{\infty} \frac{1}{2^{\alpha(k+2)}} \int\limits_{g(2^k)}^{g(2^{k+1})} d\mathbb{E}|Y_t|^\alpha$$

$$\leq C + C_1 \sum_{k=1}^{\infty} \int\limits_{g(2^k)}^{g(2^{k+1})} \frac{d\mathbb{E}|Y_t|^\alpha}{f^\alpha(t)} \leq C + C_1 \int\limits_{1}^{\infty} \frac{d\mathbb{E}|Y_t|^\alpha}{f^\alpha(t)} < \infty,$$

where C, C_1 – are some positive constants.

Because of the convergence of the series of the left in (4.2.3) by the Borel-Cantelli lemma, the sequence

$$\left\{ \sup_{g(2^n)\leq t\leq g(2^{n+1})} |Y_t|/f(t) \right\}_{n\geq 1}$$

converges to zero almost surely. Consequently,

$$\lim_{t\to\infty} \frac{Y_t}{f(t)} = 0 \text{ a.s.} \tag{4.2.4}$$

We denote by Ω' the set of elementary events, for which (4.2.4) is true. For all $\omega \in \Omega'$ and $\varepsilon > 0$ there are $s(\omega, \varepsilon) > 0$, such that

$$|Y_s|/f(s) < \varepsilon \text{ for all } s \geq s(\omega, \varepsilon),$$

and, consequently, for any $t > s(\omega, \varepsilon)$ inequality

$$\sup_{0\leq s\leq t} |Y_s(\omega)|/f(t) \leq \sup_{0\leq s\leq s(\omega,\varepsilon)} |Y_s(\omega)|/f(t) + \varepsilon$$

is fulfilled. This implies (4.2.2), since number $\varepsilon > 0$ can be chosen arbitrary small and $P(\Omega') = 1$. $\square$

4.3 Analogues of the Kolmogorov and Prokhorov-Chow theorems for martingales

The following theorem presents the analogue of the Kolmogorov and Prokhorov-Chow theorems for martingales.

Theorem 4.3.1. *Let $\{Y_t, t \in \mathbb{R}_+\}$ be a measurable separable martingale relative to some filtration $\{\mathcal{F}_t, t \in \mathbb{R}_+\}$. If*

$$\int_1^\infty \frac{d\mathbb{E}|Y_t|^{2\alpha}}{t^{2\alpha}} < \infty \tag{4.3.1}$$

for some $\alpha \geq 1$, then

$$\lim_{t\to\infty} \frac{\sup_{0\le s\le t} Y_s}{t} = 0 \ a.s. \ and \ \lim_{t\to\infty} \mathbb{E}\left(\frac{\sup_{0\le s\le t} |Y_s|}{t}\right)^{2\alpha} = 0. \tag{4.3.2}$$

Proof. The first statement in (4.3.2) follows from Theorem 4.2.1. By inequalities

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}|Y_{n+1}|^{2\alpha} - \mathbb{E}|Y_n|^{2\alpha}}{(n+1)^{2\alpha}} < \int_n^{n+1} \frac{d\mathbb{E}|Y_t|^{2\alpha}}{t^{2\alpha}} < \sum_{n=1}^{\infty} \frac{\mathbb{E}|Y_{n+1}|^{2\alpha} - \mathbb{E}|Y_n|^{2\alpha}}{n^{2\alpha}}$$

we obtain that condition (4.3.1) is equal to the following condition

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}|Y_{n+1}|^{2\alpha} - \mathbb{E}|Y_n|^{2\alpha}}{n^{2\alpha}} < \infty.$$

Hence, by Kronecker lemma,

$$\lim_{n\to\infty} \mathbb{E}\left|\frac{Y_n}{n}\right|^{2\alpha} = \lim_{n\to\infty} \frac{1}{n^{2\alpha}} \sum_{k=1}^{n} (\mathbb{E}|Y_k|^{2\alpha} - E|Y_{k-1}|^{2\alpha}) = 0.$$

For any $t \geq 1$ exists $n_t \in \mathbb{N}$, such that $n_t \leq t < n_t + 1$.The second statement in (4.3.2) follows from inequalities

$$\mathbb{E}(\sup_{0\le s\le t} |Y_s|/t)^{2\alpha} \le (2\alpha/(2\alpha-1)^{2\alpha}\mathbb{E}|Y_t|^{2\alpha}/t^{2\alpha} \le (2\alpha/(2\alpha-1)^{2\alpha}\mathbb{E}|Y_{n_t+1}|^{2\alpha}/n_t^{2\alpha}$$

The theorem is proved $\square$

Remark 4.3.2. *Theorems 4.2.1 and 4.3.1 are valid for the difference $Y_t = X_t - \mathbb{E}X_t$, where $\{X_t, t \in \mathbb{R}_+\}$ – is a measurable separable random process with independent increments. It is enough to say that under the conditions of these theorems, difference $X_t - \mathbb{E}X_t, t \in \mathbb{R}_+$, is a martingale relatively the natural filtration.*

4.4 The strong law of large numbers for homogeneous random processes with independent increments

It is natural to assume that the known strong laws of large numbers for sums of independent random variables have their analogues for homogeneous random processes with independent increments.

Theorem 4.4.1. *(i). Let $\{X_t, t \in \mathbb{R}_+\}$ is a separable homogeneous random process with independent increments. Assume that $\mathbb{E}X_1 = 0$, if $\mathbb{E}|X_1|^\alpha < \infty$ and $\alpha \geq 1$. If $\mathbb{E}|X_1|^\alpha < \infty$ for some $\alpha \in (0,2)$, then*

$$\lim_{t\to\infty} \frac{\sup\limits_{0\leq s\leq t} X_t}{t^{1/\alpha}} = 0 \text{ a.s. and } \lim_{t\to\infty} \mathbb{E}\left(\frac{\sup\limits_{0\leq s\leq t} |X_s|}{t^{1/\alpha}}\right)^\alpha = 0. \tag{4.4.1}$$

(ii). If for some constants $\alpha \in (0,2)$ and $c \in \mathbb{R} = (-\infty, \infty)$

$$\lim_{t\to\infty} \frac{X_t - ct}{t^{1/\alpha}} = 0 \text{ a.s.,}$$

then $\mathbb{E}|X_1|^\alpha < \infty$ and $c = \mathbb{E}X_1$ for $\alpha \geq 1$.

Proof. (i). We can assume that $X_0 = 0$. Otherwise, instead of X_t we can take $X_t - X_0$. Random variables X_n is the sum $X_n = \sum\limits_{k=1}^{n}(X_k - X_{k-1})$ of independent identically distributed random variables. By the Chatterjee theorem [8], we have

$$\lim_{n\to\infty} \frac{X_n}{n^{1/\alpha}} = 0 \text{ a.s. and } \lim_{n\to\infty} \mathbb{E}\left(\frac{X_n}{n^{1/\alpha}}\right)^\alpha = 0. \tag{4.4.2}$$

Using the reasoning of the proof of Theorem 4.2.1, we can be sure that the first statement of (4.4.2) implies the first statement of (4.4.1). To prove the second statement of (4.4.1) we need the Doob inequality [13, p. 303], whereby

$$\mathbb{E}(\max_{s\in S_n} |X_s|)^\alpha \leq E(\max_{0\leq k<2^n} |\sum_{j=1}^{k}(X_{j2^{-n}t} - X_{(j-1)2^{-n}t})|)^\alpha \leq 8\mathbb{E}|X_t|^\alpha$$

for any $n \in \mathbb{N}, t > 0, \alpha \geq 1$, where $S_n = \{k2^{-n}t : k = 0, ..., 2^n - 1\}$. Hence, due to the separability of random process $\{X_t, t \geq 0\}$ it follows that

$$\mathbb{E}(\sup_{0\leq s\leq t} |X_s|)^\alpha \leq 8\mathbb{E}|X_t|^\alpha.$$

By this inequality and the second statement of (4.4.2) we obtain the second statement of (4.4.1) for $\alpha \in [1,2)$.

It remains for us to prove the second statement of (4.4.1) for $\alpha \in (0,1)$. By the separability and homogeneity of random process with independent increments

$\{X_t, t \geq 0\}$, functions $Y_n = \sup\limits_{n \leq t \leq n+1} |X_t - X_n|, n \in \mathbb{N}$, are independent identically distributed random variables. We denote by $[s]$ the integer part of $s > 0$. For any $t > 1$ and $s \in (0, t]$ we have

$$|X_s| \leq |X_{[s]}| + Y_{[s]} \leq \sum_{k=1}^{[s]} |X_k - X_{k-1}| + Y_{[s]} \leq 2 \sum_{k=1}^{[t]} Y_k$$

and consequently

$$\mathbb{E}(\sup_{0 \leq s \leq t} |X_s|)^\alpha \leq 2^\alpha \sum_{k=1}^{[t]} \mathbb{E} Y_k^\alpha = 2^\alpha [t] \mathbb{E} Y_1^\alpha.$$

This implies the second statement of (4.4.1) for $\alpha \in (0, 1)$, if $\mathbb{E} Y_1^\alpha < \infty$.

Let us prove that $\mathbb{E} Y_1^\alpha < \infty$. We first show that all medians m_s of random variable X_s are limited. We assume the opposite: that, e.g., $\lim\limits_{n \to \infty} m_{s_n} = \infty$ for some sequence $s_n \in [0, 1], n \in \mathbb{N}$. We may assume that sequence $\{s_n\}_{n \geq 1}$ converges to some number $s \in [0, 1]$. By homogeneity, random process $\{X_t, t \geq 0\}$ is stochastically continuous. The distribution functions of random variables $X_{s_n}, n \in \mathbb{N}$, thus converge to the distribution function of random variable X_s. By Theorem 1.1.1 in work [28], the following inequalities are valid:

$$l_s \leq \liminf_{n \to \infty} l_{s,n} \leq \limsup_{n \to \infty} r_{s,n} \leq r_s,$$

where $l_s, l_{s,n}, r_s, r_{s,n}$ – the minimum and maximum medians of random variables X_s and X_{s_n}. We arrive at a contradiction, because both l_s and r_s are finite, and, consequently, $d = \sup\limits_{0 \leq s \leq 1} |m_s| < \infty$. By the symmetrization inequality in [30, p. 261] we have

$$P(\sup_{0 \leq s \leq 1} |X_s - m_s| \geq y) \leq 4P(|X_1| \geq y).$$

Using integration by parts, we obtain the inequality

$$\mathbb{E}(\sup_{0 \leq s \leq 1} |X_s - m_s|)^\alpha \leq 4\mathbb{E}|X_1|^\alpha.$$

Thus

$$Y_1 = \sup_{0 \leq s \leq 1} |X_s| \leq \sup_{0 \leq s \leq 1} |X_s - m_s| + d,$$

then

$$EY_1^\alpha \leq 4\mathbb{E}|X_1|^\alpha + d^\alpha < \infty.$$

(ii). We continue to assume that $X_0 = 0$. Random variable X_n is the sum $X_n = \sum\limits_{k=1}^{n} (X_k - X_{k-1})$ of independent identically distributed random variables. With assumption $\lim\limits_{n \to \infty} (X_n - cn)/n^{1/\alpha} = 0$ a.s. Hence, due to the Kolmogorov theorem

for $\alpha = 1$ and the Zygmund–Marcinkiewicz theorem for some $\alpha \in (0, 2), \alpha \neq 1$, we get $\mathbb{E}|X_2 - X_1|^\alpha < \infty$ and $c = \mathbb{E}(X_2 - X_1)$ for $\alpha \in [1, 2)$. It remains to be notes that the random variables X_1 and $X_2 - X_1$ are identically distributed and, consequently, $\mathbb{E}|X_1|^\alpha = \mathbb{E}|X_2 - X_1|^\alpha$ and $c = \mathbb{E}X_1$. The theorem is proved. □

Appendix A

Some results from probability and linear algebra

A.1 Probability theory

Theorem A.1.1 (Central Limit Theorem). *Let $Z_1, ..., Z_n$ be independent random variables with $\mathbb{E}Z_i = 0$ and finite third moment, and let $\sigma^2 = \sum_{i=1}^n \mathbb{E}|Z_i|^2$. Consider a standard normal variable g. The for every $t > 0$:*

$$\left|\mathbb{P}\left(\frac{1}{\sigma}\sum_{i=1}^n Z_i \le t\right) - \mathbb{P}\left(g \le t\right)\right| \le C\sigma^{-3}\sum_{i=1}^n \mathbb{E}|Z_i|^3,$$

where C is an absolute constant.

Lemma A.1.2. *Let event $E(X, Y)$ depends on independent random vectors X and Y then*

$$\mathbb{P}(E(X, Y)) \le (\mathbb{P}(E(X, Y), E(X, Y')))^{1/2},$$

where Y' is an independent copy of Y.

Proof. See in [12]. □

Lemma A.1.3. *Let $Z_1, ..., Z_n$ be a sequence of random variables and $p_1, ..., p_n$ be non-negative real numbers such that*

$$\sum_{i=1}^n p_i = 1,$$

then for every $\varepsilon > 0$

$$\mathbb{P}(\sum_{i=1}^n p_i Z_i \le \varepsilon) \le 2\sum_{i=1}^n p_i \mathbb{P}(Z_i \le 2\varepsilon).$$

Proof. See in [43]. □

We recall definition of Levy concentration function

Definition A.1.4. *Levy concentration function of random variable Z with values from $\mathbb{R}^d$ is a function*

$$\mathcal{L}(Z, \varepsilon) = \sup_{v \in \mathbb{R}^d} \mathbb{P}(||Z - v||_2 < \varepsilon).$$

Lemma A.1.5. *Let* $S_J = \sum_{i \in J} \xi_i$, *where* $J \subset [n]$, *and* $I \subset J$ *then*

$$\mathcal{L}(S_J, \varepsilon) \leq \mathcal{L}(S_I, \varepsilon).$$

Proof. Let us fix arbitrary v. From independence of ξ_i we conclude

$$\mathbb{P}(|S_J - v| \leq \varepsilon) \leq \mathbb{E}\mathbb{P}(|S_I + S_{J/I} - v| \leq \varepsilon | \{\xi_i\}_{i \in I}) \leq \sup_{u \in \mathbb{R}} \mathbb{P}(|S_I - u| \leq \varepsilon).$$

□

Lemma A.1.6. *Let* Z *be a random variable with* $\mathbb{E}Z^2 \geq 1$ *and with finite fourth moment, and put* $M_4^4 := \mathbb{E}(Z - \mathbb{E}Z)^4$. *Then for every* $\varepsilon \in (0, 1)$ *there exists* $p = p(M_4, \varepsilon)$ *such that*

$$\mathcal{L}(Z, \varepsilon) \leq p.$$

Proof. See in [37]. □

Lemma A.1.7. *Let* $X = (X_1, ..., X_n)$ *be a random vector in* $\mathbb{R}^n$ *with independent coordinates* X_k.
1. Suppose there exist numbers $\varepsilon_0 \geq 0$ *and* $L \geq 0$ *such that*

$$\mathcal{L}(X_k, \varepsilon) \leq L\varepsilon \quad \text{for all } \varepsilon \geq \varepsilon_0 \text{ and all } k.$$

Then

$$\mathcal{L}(X, \varepsilon) \leq (CL\varepsilon)^n \quad \text{for all } \varepsilon \geq \varepsilon_0,$$

where C *is an absolute constant.*
2. Suppose there exist numbers $\varepsilon > 0$ *and* $p \in (0, 1)$ *such that*

$$\mathcal{L}(X_k, \varepsilon) \leq L\varepsilon \quad \text{for all } k.$$

Then there exist numbers $\varepsilon_1 = \varepsilon_1(\varepsilon, p) > 0$ *and* $p_1 = p_1(\varepsilon, p) \in (0, 1)$ *such that*

$$\mathcal{L}(X, \varepsilon) \leq (p_1)^n.$$

Proof. See [43, Lemma 3.4]. □

Lemma A.1.8. *There exist* $\gamma > 0$ *and* $\delta > 0$ *such that for all* $n \gg 1$ *and* $1 \leq i \leq n$, *any deterministic vector* $v \in \mathbb{C}$ *and any subspace* H *of* $\mathbb{C}^n$ *with* $1 \leq \dim(H) \leq n - n^{1-\gamma}$, *we have, denoting* $R := (X_1, ..., X_n) + v$,

$$\mathbb{P}(\text{dist}(R, H) \leq \frac{1}{2}\sqrt{n - \dim(H)}) \leq \exp(-n^{\delta}).$$

Proof. See [41, Statement 5.1]. □

A.2 Linear algebra and geometry of the unit sphere

Lemma A.2.1. *Let* $1 \leq m \leq n$. *If* $\mathbf{A}$ *has full rank, with rows* $R_1, ..., R_m$ *and* $H = \operatorname{span}(R_j, j \neq i)$, *then*

$$\sum_{i=1}^{m} s_i(\mathbf{A})^{-2} = \sum_{i=1}^{m} \operatorname{dist}(R_i, H_i)^{-2}.$$

Proof. See [41, Lemma A.4]. □

Definition A.2.2. *(Compressible and incompressible vectors) Let* $\delta, \tau \in (0, 1)$. *A vector* $x \in \mathbb{R}^n$ *is called sparse if* $|supp(x)| \leq \delta n$. *A vector* $x \in S^{n-1}$ *is called compressible if* x *is within Euclidian distance* τ *from the set of all sparse vectors. A vector* $x \in S^{n-1}$ *is called incompressible if it is not compressible. The sets of sparse, compressible and incompressible vectors will be denoted by Sparse = Sparse* (δ), *Comp = Comp* (δ, τ) *and Incomp = Incomp*(δ, τ) *respectively.*

Lemma A.2.3. *If* $x \in Incomp(\delta, \tau)$ *then at least* $\frac{1}{2}\delta\tau^2 n$ *coordinates* x_k *of* x *satisfy*

$$\frac{\tau}{\sqrt{2n}} \leq |x_k| \leq \frac{1}{\sqrt{\delta n}}.$$

Remark A.2.4. *We can fix some constant* c_0 *such that*

$$\frac{1}{4}\delta\tau^2 \leq c_0 \leq \frac{1}{4}.$$

Then for every vector $x \in Incomp(\delta, \tau)$ $|\operatorname{spread}(x)| = [2c_0 n]$.

Proof. See in [37]. □

Appendix B

Methods

B.1 Moment method

In this section we present results which give the method to investigate under what conditions the convergence of moments of all fixed orders implies the weak convergence of the sequence of the distribution functions. Let $\{F_n\}$ be a sequence of distribution functions.

Theorem B.1.1. *A sequence of distribution functions $\{F_n\}$ converges weakly to a limit if the following conditions are satisfied:*

1. *each F_n has finite moments of all orders.*
2. *For each fixed integer $k \geq 0$ the k-th moment of F_n converges to a finite limit β_k as $n \to \infty$.*
3. *If two right-continuous functions F and G have the same moment sequence $\{\beta_k\}$, then $F = G + const$.*

Proof. See [4]. □

One need to verify condition 3) of the Theorem B.1.1. The following theorem gives condition that implies 3).

Theorem B.1.2 (Carleman)**.** *Let $\{\beta_k = \beta_k(F)\}$ be the sequence of moments of the distribution function F. If the Carleman condition*

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

is satisfied, then F is uniquely determined by the moment sequence $\{\beta_k\}$.

Proof. See [4]. □

B.2 Stieltjes transform method

Definition B.2.1. *The Stieltjes transform of the distribution function G is a function*

$$S_G(\alpha) = \int_{\mathbb{R}} \frac{dG(x)}{x - \alpha}, \quad \alpha \in \mathbb{C}^+.$$

Theorem B.2.2 (Inversion formula)**.** *For any continuity points $a < b$ of G, we have*

$$G([a,b]) = \lim_{\varepsilon \to 0+} \frac{1}{\pi} \int_a^b \operatorname{Im} S_G(x + i\varepsilon) dx,$$

Proof. See [4]. □

For the ESD of the random matrix $n^{-1/2}\mathbf{X}_n$ one has

$$S^{\mathbf{X}_n}(z) = \int_{\mathbb{R}} \frac{1}{x-z} d\mathcal{F}^{\mathbf{X}_n} = \frac{1}{n} \operatorname{Tr} \left(\frac{1}{\sqrt{n}} \mathbf{X}_n - z\mathbf{I} \right)^{-1}.$$

The following theorem gives the method to investigate convergence of the ESD to some limit.

Theorem B.2.3. *Let $\mathcal{F}^{\mathbf{X}_n}$ be the ESD of the random matrix $n^{-1/2}\mathbf{X}_n$ and set $F^{\mathbf{X}_n} = \mathbb{E}\mathcal{F}^{\mathbf{X}_n}$. Then*

1. *$\mathcal{F}^{\mathbf{X}_n}(x)$ converges almost surely to $F(x)$ in the vague topology if and only if $S^{\mathbf{X}_n}(z)$ converges almost surely to $S(z)$ for every z in the upper half-plane;*

2. *$\mathcal{F}^{\mathbf{X}_n}(x)$ converges in probability to $F(x)$ in the vague topology if and only if $S^{\mathbf{X}_n}(z)$ converges in probability to $S(z)$ for every z in the upper half-plane;*

3. *$F^{\mathbf{X}_n}(x)$ converges almost surely to $F(x)$ in the vague topology if and only if $\mathbb{E}S^{\mathbf{X}_n}(z)$ converges almost surely to $S(z)$ for every z in the upper half-plane.*

Proof. See [42]. □

B.3 Logarithmic potential

Definition B.3.1. *The logarithmic potential U_m of measure $m(\cdot)$ is a function $U_m : \mathbb{C} \to (-\infty, +\infty]$ defined for all $z \in \mathbb{C}$ by*

$$U_m(z) = -\int_{\mathbb{C}} \log|z - w| m(dw).$$

Definition B.3.2. *The function $f : \mathbb{T} \to \mathbb{R}$, where $\mathbb{T} = \mathbb{C}$ or $\mathbb{T} = \mathbb{R}$, is uniformly integrable in probability with respect to the sequence of random measures $\{m_n\}_{n\geq 1}$ on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ if for all $\varepsilon > 0$:*

$$\lim_{t\to\infty} \overline{\lim_{n\to\infty}} \, \mathbb{P} \left(\int_{|f|>t} |f(x)| m_n(dx) > \varepsilon \right) = 0.$$

Let $s_1(n^{-1/2}\mathbf{X}_n - z\mathbf{I}) \geq s_2(n^{-1/2}\mathbf{X}_n - z\mathbf{I}) \geq ... \geq s_n(n^{-1/2}\mathbf{X}_n - z\mathbf{I})$ be the singular values of $n^{-1/2}\mathbf{X}_n - z\mathbf{I}$ and define the empirical spectral measure of singular values by

$$\nu_n(z, B) = \frac{1}{n}\#\{i \geq 1 : s_i(n^{-1/2}\mathbf{X}_n - z\mathbf{I}) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}).$$

We can rewrite the logarithmic potential of μ_n via the logarithmic moments of measure ν_n by

$$U_{\mu_n}(z) = -\int_{\mathbb{C}} \log|z - w|\mu_n(dw) = -\frac{1}{n}\log\left|\det\left(\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I}\right)\right|$$

$$= -\frac{1}{2n}\log\det\left(\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I}\right)^*\left(\frac{1}{\sqrt{n}}\mathbf{X}_n - z\mathbf{I}\right) = -\int_0^\infty \log x\nu_n(dx).$$

This allows us to consider the Hermitian matrix $(n^{-1/2}\mathbf{X}_n - z\mathbf{I})^*(n^{-1/2}\mathbf{X}_n - z\mathbf{I})$ instead of the asymmetric matrix $n^{-1/2}\mathbf{X}$.

Lemma B.3.3. *Let $(\mathbf{X}_n)_{n\geq 1}$ be a sequence of $n \times n$ random matrices. Suppose that for a.a. $z \in \mathbb{C}$ there exists a probability measure ν_z on $[0, \infty)$ such that*
a) $\nu_n \xrightarrow{weak} \nu_z$ as $n \to \infty$ in probability
b) log *is uniformly integrable in probability with respect to $\{\nu_n\}_{n\geq 1}$.*
Then there exists a probability measure μ such that
a) $\mu_n \xrightarrow{weak} \mu$ as $n \to \infty$ in probability
b) for a.a. $z \in \mathbb{C}$

$$U_\mu(z) = -\int_0^\infty \log x\nu_z(dx).$$

Proof. See [6, Lemma 4.3] for the proof. □

Appendix C

Stochastic processes

In this chapter we introduce all necessary definitions and theorems from the theory of stochastic processes. See [18] for the discussion of stochastic processes.

C.1 Some facts from stochastic processes

Definition C.1.1. *Function of two variables $X(t, \omega) = \xi(t)$, defined for all $t \in T, \omega \in \Omega$, taking values in a metric space X, $\mathcal{F}$-measurable for all $t \in T$, is called stochastic process. The set T is a domain of stochastic process and the space X is a codomain of stochastic process.*

Definition C.1.2. *Stochastic processes $X_1(t, \omega)$ and $X_2(t, \Omega)$ determined on the common probability space are stochastically equivalent if for all $t \in T$*

$$\mathbb{P}(X_1(t, \omega) \neq X_2(t, \omega)) = 0.$$

Let us consider a sequence of random variables X_i, then it is well known that $\sup_i X_i$ is a random variable. It follows immediately from

$$\{\omega \in \Omega : \sup_i X_i > x\} = \bigcup_{i=1}^{\infty} \{\omega \in \Omega : X_i > x\} \in \mathcal{F}.$$

Now let us consider stochastic process X_t. In this case it may occur that $\sup_t X_t$ is not a random variable. To overcome this difficulty we should introduce the definition of a separable process.

Definition C.1.3. *Stochastic process is called separable it there exist the countable and dense set of points $\{t_j\}_{j \geq 1} \subset T$ and the set $N \subset \Omega$, $\mathbb{P}(N) = 0$, such that for all open $G \subset T$ and all closed set $F \in X$ the sets*

$$\{\omega : X_{t_j}(\omega) \in F, t_j \in G\}, \quad \{\omega : X_t(\omega) \in F, t \in G\}$$

differ only on subsets of N.

Separability is not very strict condition. Under rather general assumptions on T and X there exists a separable process which is equivalent to a given one.

Theorem C.1.4. *Let X be a separable locally compact space and T – arbitrary separable space. For all $X_t(\omega)$ defined on T, taking values in X, there exists a stochastically equivalent copy $\tilde{X}_t(\omega)$ taking values in $\tilde{X}$ which is a compact extension of X.*

Proof. See [18]. □

In this thesis we consider the case $T = \mathbb{R}_+$ and $X = \mathbb{R}$. Then the statement of Theorem C.1.4 is automatically satisfied.

Definition C.1.5. *Stochastic process X_t is called process with independent increments if for all $t_0 < t_1 < ... < t_k$ from T the random variables $X(t_0), X(t_1) - X(t_0), ..., X(t_k) - X(t_{k-1})$ are independent.*

Definition C.1.6. *Stochastic process X_t is called homogenous if*

$$\mathrm{Law}(X_{t+s} - X_s) = \mathrm{Law}(X_t - X_0), s, t \in T.$$

Definition C.1.7. *Stochastic process $X_t, t \in T$ defined on the stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in T}, \mathbb{P})$ is called martingale, if X_t is $\mathcal{F}_t$-measurable, $\mathbb{E}X_t < \infty$, $t \in T$, and*

$$\mathbb{E}(X_t|\mathcal{F}_s) = X_s, \quad s \leq t, s, t \in T.$$

Bibliography

[1] R. Adamczak. On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries. *Electron. J. Probab.*, 16:no. 37, 1068–1095, 2011. 50

[2] G. Akemann, J. Baik, and P. Di Francesco. *The Oxford Handbook of Random Matrix Theory.* Oxford Unversity Press, London, 2011. 6, 16

[3] L. Arnold. On Wigner's semicircle law for the eigenvalues of random matrices. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 19:191–198, 1971. 4

[4] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices.* Springer, New York, second edition, 2010. 5, 9, 37, 56, 75, 76

[5] V. Bentkus. A new approach to approximations in probability theory and operator theory. *Liet. Mat. Rink.*, 43(4):444–470, 2003. 52

[6] C. Bordenave and D. Chafaï. Around the circular law. *arXiv:1109.3343.* 6, 9, 77

[7] H. D. Brunk. The strong law of large numbers. *Duke Math. J.*, 15:181–195, 1948. 11

[8] S. D. Chatterji. An L^p-convergence theorem. *Ann. Math. Statist.*, 40:1068–1070, 1969. 68

[9] Y. S. Chow. On a strong law of large numbers for martingales. *Ann. Math. Statist.*, 38:610, 1967. 11

[10] Yuan Shih Chow and Henry Teicher. *Probability theory.* Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1988. Independence, interchangeability, martingales. 11, 64, 65

[11] John B. Conway. *Functions of one complex variable*, volume 11. Springer-Verlag, New York, second edition, 1978. 35

[12] K. Costello. Bilinear and quadratic variants on the Littlewood-Offord problem. *Submitted.* 71

[13] D. L. Doob. *Stochastic processes.* Izdat. Inostr. Lit., Moscow, 1956. 66, 68

[14] A. Edelman. The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law. *J. Multivariate Anal.*, 60(2):203–232, 1997. 5

[15] L. Erdős. Universality of wigner random matrices: a survey of recent results. *arXiv:1004.0861.* 4

[16] I. Fazekas and O. Klesov. A general approach to the strong laws of large numbers. *Teor. Veroyatnost. i Primenen.*, 45(3):568–583, 2000. 11, 63, 65

[17] Y. V. Fyodorov, B. A. Khoruzhenko, and H. Sommers. Universality in the random matrix spectra in the regime of weak non-hermiticity. *Ann. Inst. Henri Poincare: Phys. Theor.*, 68(4):449–489, 1998. 6, 16

[18] I. I. Gikhman and A. V. Skorokhod. *Introduction to the theory of random processes.* Translated from the Russian by Scripta Technica, Inc. W. B. Saunders Co., Philadelphia, Pa., 1969. 79, 80

[19] J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Mathematical Phys.*, 6:440–449, 1965. 5

[20] V. L. Girko. The circular law. *Teor. Veroyatnost. i Primenen.*, 29(4):669–679, 1984. 5, 9

[21] V. L. Girko. The elliptic law. *Teor. Veroyatnost. i Primenen.*, 30(4):640–651, 1985. 6

[22] V. L. Girko. The strong elliptic law. Twenty years later. *Random Oper. and Stoch. Equ.*, 14(1):59–102, 2006. 6

[23] F. Götze and A. Tikhomirov. Limit theorems for spectra of positive random matrices under dependence. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 311(Veroyatn. i Stat. 7):92–123, 299, 2004. 50

[24] F. Götze and A. Tikhomirov. The circular law for random matrices. *Ann. Probab.*, 38(4):1444–1491, 2010. 5, 10, 33

[25] F. Götze and A. N. Tikhomirov. Limit theorems for spectra of random matrices with martingale structure. *Teor. Veroyatn. Primen.*, 51(1):171–192, 2006. 4, 5, 50

[26] P. Hall and C. C. Heyde. *Martingale limit theory and its application.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980. Probability and Mathematical Statistics. 4

[27] V. M. Kruglov. A generalization of the Brunk-Prokhorov strong law of large numbers. *Teor. Veroyatnost. i Primenen.*, 47(2):347–349, 2002. 11, 63

[28] V. M. Kruglov and V. Yu. Korolev. *Limit theorems for random sums.* Moskov. Gos. Univ., Moscow, 1990. With a foreword by B. V. Gnedenko. 69

[29] M. Ledoux. Complex hermitian polynomials: from the semi-circular law to the circular law. *Commun. Stoch. Anal.*, 2(1):27–32, 2008. 6, 16

[30] Michel Loève. *Probability theory.* Translated from the English by B. A. Sev-ast'janov; edited by Ju. V. Prohorov. Izdat. Inostr. Lit., Moscow, 1962. 11, 12, 63, 64, 69

[31] Madan Lal Mehta. *Random matrices.* Academic Press Inc., Boston, MA, second edition, 1991. 5

[32] H. Nguyen and S. O'Rourke. The Elliptic Law. *arXiv:1208.5883*. 6

[33] S. O'Rourke. A note on the Marchenko-Pastur law for a class of random matrices with dependent entries. *arXiv:1201.3554*. 50

[34] G. Pan and W. Zhou. Circular law, Extreme Singular values and Potential theory. *arXiv:0705.3773*. 5

[35] L. A. Pastur. Spectra of random selfadjoint operators. *Uspehi Mat. Nauk*, 28(1(169)):3–64, 1973. 4

[36] Yu. V. Prohorov. On the strong law of large numbers. *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, 14:523–536, 1950. 11

[37] M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008. 10, 17, 19, 72, 73

[38] Albert N. Shiryaev. *Probability*, volume 95. Springer-Verlag, New York, second edition, 1996. 10, 12, 43, 64

[39] H. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein. Spectrum of large random asymmetric matrices. *Phys. Rev. Lett.*, 60:1895–1898, May 1988. 6, 8, 16

[40] T. Tao and V. Vu. Random Matrices: The circular Law. *arXiv:0708.2895*. 6, 10

[41] T. Tao and V. Vu. Random matrices: universality of local eigenvalue statistics. *Acta Math.*, 206(1):127–204, 2011. 6, 72, 73

[42] Terence Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012. 25, 76

[43] R. Vershynin. Invertibility of symmetric random matrices. *arXiv:1102.0300*. 10, 17, 19, 71, 72

[44] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958. 3