Extreme Eigenvalues of Sample Covariance and Correlation Matrices

Johannes Heiny

PhD Thesis

This thesis has been submitted to the PhD School of the Faculty of Science, University of Copenhagen

January 10, 2017

Department of Mathematical Sciences
Faculty of Science
University of Copenhagen

Johannes Heiny
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
2100 København Ø
Denmark
johannes.heiny@gmail.com

Supervisor:	Prof. Thomas Mikosch University of Copenhagen
Co-supervisor:	Prof. Olivier Wintenberger University Pierre and Marie Curie
Assessment committee:	Prof. Claudia Klüppelberg Technical University Munich
	Prof. Richard Davis Columbia Univerity Ney York
	Associate Prof. Jeffrey Collamore (chairman) University of Copenhagen

Abstract

This thesis is concerned with asymptotic properties of the eigenvalues of high-dimensional sample covariance and correlation matrices under an infinite fourth moment of the entries.

In the first part, we study the joint distributional convergence of the largest eigenvalues of the sample covariance matrix of a p-dimensional heavy-tailed time series when p converges to infinity together with the sample size n. We generalize the growth rates of p existing in the literature. Assuming a regular variation condition with tail index $\alpha < 4$, we employ a large deviations approach to show that the extreme eigenvalues are essentially determined by the extreme order statistics from an array of iid random variables. The asymptotic behavior of the extreme eigenvalues is then derived routinely from classical extreme value theory. The resulting approximations are strikingly simple considering the high dimension of the problem at hand.

We develop a theory for the point process of the normalized eigenvalues of the sample covariance matrix in the case where rows and columns of the data are linearly dependent. Based on the weak convergence of this point process we derive the limit laws of various functionals of the eigenvalues.

In the second part, we show that the largest and smallest eigenvalues of a high-dimensional sample correlation matrix possess almost sure non-random limits if the truncated variance of the entry distribution is "almost slowly varying", a condition we describe via moment properties of self-normalized sums. We compare the behavior of the eigenvalues of the sample covariance and sample correlation matrices and argue that the latter seems more robust, in particular in the case of infinite fourth moment.

Resumé

Denne afhandling beskæftiger sig med de asymptotiske egenskaber af egenværdierne for højdimensionale empiriske korrelations- og kovariansmatricer, under antagelse af at matrixindgangene har uendeligt fjerde moment.

I første del undersøger vi konvergens i fordeling af de største egenværdier for den observerede kovarians af en p-dimensional tunghalet tidsrække, når p sammen med stikprøvestørrelsen går mod uendelig. Vi generaliserer de eksisterende vækstrater af p fra litteraturen. Under antagelse af regulær variation med haleindeks $\alpha < 4$, bruger vi en large deviations-tilgang for at vise at de ekstremale egenværdier er essentielt bestemt ud fra de største værdier i et array af iid. stokastiske variable. Herefter udleder vi rutinemæssigt de ekstremale egenværdiers asymptotiske egenskaber ved brug af klassisk ekstremværditeori.

Vi fremfører en teori for punktprocesser af de normalisererde egenværdier af den empiriske kovariansmatrix i tilfældet, hvor dens rækker og søjler er afhængige. Med udgangspunkt i, at denne punktproces konverger svagt, udleder vi grænsemomenter af forskellige funktionaler af egenværdierne.

I den anden del af afhandlingen viser vi at de største og mindste egenværdier fra en højdimensional korrelationsmatrix har næsten sikre grænser, hvis den trunkerede varians af indgangsfordelingen er „næsten langsom varierende", en betingelse som kan indentificeres ud fra momentegenskaber af selv-normaliserende summer. Vi sammenligner, hvordan egenværdierne for henholdsvis den empiriske korrelations- og den empiriske kovariansmatrix opfører sig, og argumenterer for at sidstnævnte tilfælde er mere robust, hvilket særligt gælder i tilfældet med uendeligt fjerde moment.

Acknowledgments

First and foremost, I would like to thank my supervisor Thomas Mikosch for many fascinating discussions, his guidance and experience. I could not have wished for a better person to work with and learn from. Furthermore I am grateful for the support I received from my co-supervisor Olivier Wintenberger.

I want to thank Richard Davis and the Statistics Department of Columbia University in the City of New York, where I spent a research visit in 2016, for their great hospitality. The group meetings on Wednesdays led to a vivid exchange of ideas. I would also like to thank Claudia Klüppelberg. Her advice is highly appreciated.

Next, I want to thank Gennady Samorodnitsky and Mark Podolskij for inspiring discussions. I am also grateful to Jeffrey Collamore, Jesper Lund Pedersen, Mogens Steffensen and Jostein Paulsen.

It is my special pleasure to thank our lunch group at the University of Copenhagen which was an integral part of daily university life. I also want to thank my fellow PhD students at the University of Copenhagen and Columbia University for many fun hours. Special thanks go to all my office mates and in particular Xie Xiaolei.

I acknowledge the support by Danish Research Council Grant DFF-4002-00435 “Large random matrices with heavy tails and dependence”.

Last but not least, I want to thank my family for the unconditional love and the invaluable support I have received over the years.

Johannes Heiny *Copenhagen, January* 2017.

Summary

This PhD thesis provides asymptotic theory for the eigenvalues of the sample covariance matrix of a high-dimensional time series with infinite fourth moment. Its main part consists of the first three of the following research papers. They were written from March 2014 until December 2016.

[P1] Heiny, J., and Mikosch, T. Eigenvalues and eigenvectors of heavy-tailed sample covariance matrices with general growth rates: the iid case. *Stochastic Process. Appl.* (2016), 29. [pdf]

[P2] Davis, R. A., Heiny, J., Mikosch, T., and Xie, X. Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series. *Extremes 19*, 3 (2016), 517–547. [pdf]

[P3] Heiny, J., and Mikosch, T. Almost sure convergence of the largest and smallest eigenvalues of high-dimensional sample correlation matrices under infinite fourth moment. *Submitted for publication.*

[P4] Heiny, J., and Mikosch, T. Limit theory for the singular values of the sample autocovariance matrix function of multivariate time series. *In preparation.*

[P5] Heiny, J., and Mikosch, T. Asymptotic theory for high-dimensional stochastic volatility matrices. *In preparation.*

The aforementioned infinite fourth moment is ensured by a *regular variation condition.* We say that a random variable X and its distribution are *regularly varying with index* $\alpha > 0$ if

$$\mathbb{P}(X > x) \sim p_+ \frac{L(x)}{x^\alpha} \quad \text{and} \quad \mathbb{P}(X < -x) \sim p_- \frac{L(x)}{x^\alpha}\,, \qquad x \to \infty\,, \qquad (RV_\alpha)$$

where $p_\pm$ are non-negative constants such that $p_+ + p_- = 1$ and L is a slowly varying function. In particular, if $\alpha < 4$ we have $\mathbb{E}[X^4] = \infty$. The regular variation condition is needed for proving asymptotic theory for the eigenvalues of the sample covariance matrix. Moreover, we will often use the concept of *heavy tails.* A distribution is called *heavy-tailed* if certain moments are infinite. By construction, any regularly varying distribution is heavy-tailed.

Now we explain the structure of this thesis. In Chapter 1 we provide an introduction to Random Matrix Theory and present the classical results in the light-tailed case. We give examples of high-dimensional statistical inference problems and indicate how the asymptotic theory applies. The empirical distribution of the eigenvalues of the widely used sample covariance matrix is studied. Furthermore the a.s. limits of its largest and smallest eigenvalues are identified under finite fourth moment. Then we explain the typical behavior of its eigenvectors. *Section 1.5 constitutes the main part of Chapter 1.* There we present the contribution of this thesis and the novelties of [P1, P2, P3].

Chapters 2-4 consist of these 3 papers, respectively. Each chapter is self-contained with its own introduction and references. Chapter 1 can be used as a joint introduction to Chapters 2-4.

In Chapter 2 we study the joint distributional convergence of the largest eigenvalues of the sample covariance matrix of a p-dimensional time series with iid entries when p converges to infinity together with the sample size n. We generalize the growth rates of p in the literature. Assuming the regular variation condition with $\alpha < 4$, we employ a large deviations approach to show that only the diagonal of the sample covariance matrix is relevant for the asymptotic behavior of the largest eigenvalues and the corresponding eigenvectors. The resulting approximations are strikingly simple considering the high dimension of the problem at hand.

In Chapter 3 we generalize the results from the iid case that were presented in the previous chapter. We develop a theory for the point process of the normalized eigenvalues of the sample covariance matrix in the case when rows and columns of the data are linearly dependent. We provide limit results for the weak convergence of these point processes to Poisson or cluster Poisson processes. Based on this convergence we can also derive the limit laws of various functionals of the ordered eigenvalues such as the joint convergence of a finite number of the largest order statistics, the joint limit law of the largest eigenvalue and the trace, limit laws for successive ratios of ordered eigenvalues, etc. We also develop some limit theory for the singular values of the sample autocovariance matrices and their sums of squares. The theory is illustrated for simulated data and for the components of the S&P 500 stock index. Further generalizations of the results of this chapter are made in [P4] and [P5], but are not part of this thesis.

In Chapter 4, we show that the largest and smallest eigenvalues of a sample correlation matrix stemming from n independent observations of a p-dimensional time series with iid components converge almost surely to $(1+\sqrt{\gamma})^2$ and $(1-\sqrt{\gamma})^2$, respectively, as $n \to \infty$, if $p/n \to \gamma \in (0,1]$ and the truncated variance of the entry distribution is "almost slowly varying", a condition we describe via moment properties of self-normalized sums. We compare the behavior of the eigenvalues of the sample covariance and sample correlation matrices and argue that the latter seems more robust, in particular in the case of infinite fourth moment. We briefly address some practical issues for the estimation of extreme eigenvalues in a simulation study.

Chapter 4 is the most technical one of this thesis. In our proofs we use the method of moments combined with a Path-Shortening Algorithm, which efficiently uses the structure of sample correlation matrices, to calculate precise bounds for matrix norms. We believe that this new approach could be of further use in Random Matrix Theory.

Johannes Heiny *Copenhagen, January* 2017.

Contents

Chapter 1

Introduction

In this thesis we study the largest and smallest eigenvalues of high-dimensional sample covariance and correlation matrices of heavy-tailed time series.

1.1 Random Matrix Theory

The field of Random Matrix Theory (RMT) is concerned with the spectral properties of high-dimensional random matrices. Its development was motivated by applications. In quantum mechanics, for example, the energy levels of particles in a large system can be characterized by the eigenvalues of a random infinite-dimensional Hermitean operator $\mathbf{W}$ on a Hilbert space. It is common to work with discretizations in a finite-dimensional space. In this case $\mathbf{W}$ turns into a high-dimensional Hermitean random matrix. Such matrices are called Wigner matrices named after Eugene Paul Wigner. He proved in the 1950s that if $\mathbf{W}$ has independent standard normal entries on and above the diagonal, then the expected empirical distribution of the eigenvalues of $\mathbf{W}$ tends to the so-called semi-circle law defined in (1.18), as the dimension of $\mathbf{W}$ goes to infinity; see for example [72, 73]. Since then a great variety of asymptotic results has been proved for various classes of random matrices under different assumptions on the distribution of the entries and their dependence structure. The discovery of many results was triggered by the enormous improvement of computation power which led to numerous conjectures. In the second half of the 20th century, the research on asymptotic spectral properties of large-dimensional random matrices attracted considerable interest among physicists, computer scientists and mathematicians. A breakthrough in the theory of spectral distributions of sample covariance matrices was achieved by Marčenko and Pastur in 1967.

For many years the main focus of research in RMT has been on limiting spectral distributions. More recently, the focus turned to linear spectral statistics, eigenvectors, limiting distributions of extreme eigenvalues and their spacings. This thesis addresses these four topics in the special setting of random matrices with heavy-tailed entries.

RMT is a versatile and useful tool in many fields of modern sciences that are faced with high-dimensional data sets. It employs techniques from probability theory, multivariate statistics, number theory and combinatorics. Moreover, RMT finds applications in quantum physics, signal processing, wireless communications and finance; see Bai et al. [5] for more detailed examples.

1.1.1 Limiting spectral distributions

For any random $p \times p$ matrix $\mathbf{A}$ with real eigenvalues $\lambda_1(\mathbf{A}), \ldots, \lambda_p(\mathbf{A})$ the *empirical spectral distribution* is defined by

$$F_{\mathbf{A}}(x) = \frac{1}{p} \sum_{i=1}^{p} \mathbf{1}_{\{\lambda_i(\mathbf{A}) \le x\}}, \qquad x \in \mathbb{R}\,.$$

A major problem in random matrix theory is to find the weak limit of $(F_{\mathbf{A}_n})$, the so-called *limiting spectral distribution*, for suitable sequences of Hermitean $p \times p$ matrices $(\mathbf{A}_n)$. By weak convergence of a sequence of probability distributions $(F_{\mathbf{A}_n})$ to a probability distribution F, we mean $\lim_{n\to\infty} F_{\mathbf{A}_n}(x) = F(x)$ a.s. for all continuity points of F. Although the eigenvalues of $\mathbf{A}_n$ are continuous functions of the entries of $\mathbf{A}_n$ there are no closed-form expressions if the dimension is larger than 4. Therefore methods to identify and characterize the limiting spectral distribution are needed. We will briefly discuss the two most common ones: the method of moments and Stieltjes transforms.

By Lemma B.3 in [6], a distribution function F is uniquely characterized by its sequence of moments

$$\beta_k = \int_0^\infty x^k \,\mathrm{d}F(x)\,, \quad k = 1, 2, \dots$$

if Carleman's condition $\sum_{k=1}^\infty \beta_{2k}^{-1/(2k)} = \infty$ is satisfied. In this case, weak convergence of $(F_{\mathbf{A}_n})$ to F is equivalent to the convergence of moments, that is

$$\beta_k(\mathbf{A}_n) = \int_0^\infty x^k \,\mathrm{d}F_{\mathbf{A}_n}(x) = \frac{1}{p}\operatorname{tr}(\mathbf{A}_n^k) \to \beta_k\,, \quad n \to \infty,\, k = 1, 2, \dots \tag{1.1}$$

In many cases the calculation of $\operatorname{tr}(\mathbf{A}_n^k)$ is very demanding. Often its mean and variance are estimated by combinatorial techniques. On the positive side, if F has finite support, Carleman's condition holds automatically.

Another useful tool is the *Stieltjes transform* of the empirical spectral distribution $F_{\mathbf{A}}$:

$$s_{F_{\mathbf{A}}}(z) = \int_{\mathbb{R}} \frac{1}{x-z}\,\mathrm{d}F_{\mathbf{A}}(x) = \frac{1}{p}\operatorname{tr}\left((\mathbf{A} - z\mathbf{I})^{-1}\right), \quad z \in \mathbb{C}^+\,,$$

where $\mathbb{C}^+$ denotes the complex numbers with positive imaginary part. Weak convergence of $(F_{\mathbf{A}_n})$ to a distribution function F is equivalent to $s_{F_{\mathbf{A}_n}}(z) \to s_F(z)$ a.s. for all $z \in \mathbb{C}^+$. Notice that the Stieltjes transform s_F determines a distribution function F at all continuity points a, b of F via

$$F(b) - F(a) = \lim_{v \to 0^+} \frac{1}{\pi} \int_a^b \Im s_F(x + \mathrm{i}v)\,\mathrm{d}x\,.$$

1.2 Sample covariance matrices

For a sample of n column vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ of a p-dimensional time series the (non-normalized) *sample covariance matrix* is usually defined as

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

where $\overline{\mathbf{x}} = n^{-1}\sum_i \mathbf{x}_i$ is the sample mean and

$$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n) = (X_{it})_{i=1,\dots,p;\, t=1,\dots,n}$$

is the *data matrix*. The sample covariance matrix is of crucial importance in multivariate statistics, for instance in principal component analysis, canonical correlation analysis, multivariate regression, factor analysis, hypothesis testing and discriminant analysis. The case of multivariate normal observations has played a particular role in the development of statistical theory. Already in 1928, Wishart [74] studied sample covariance matrices with normal entries. Through the 20th century non-asymptotic procedures for Gaussian observations such as Fisher's test, Student's test, and the analysis of variance were developed. In practice, however, observations are often not normally distributed and asymptotic methods based on limit theorems for certain model parameters are employed instead of exact results which are difficult to obtain.

Most of the classical limit theorems are derived under the assumption that the dimension p is fixed and the sample size n goes to infinity. If the assumptions of the law of large of large numbers are satisfied, then $n^{-1}\mathbf{S}$ converges a.s. to the covariance matrix Σ of $\mathbf{x}_1$.

If p is moderately large, it is known that $n^{-1}\mathbf{S}$ ceases to be a good estimate for Σ. This means that classical methods based on fixed dimension and large sample limits may lead to wrong conclusions when applied to high-dimensional data. One would need appropriate adjustments. RMT provides limit theory in the case of large p. New statistical methods can be built on these results.

In RMT one assumes that $p = p_n$ grows with n. The most common condition in the literature is

$$\lim_{n\to\infty} \frac{p_n}{n} \to \gamma \in (0, \infty)\,. \tag{1.2}$$

The asymptotic spectral behaviors of two large matrices are the same if their difference is of finite rank. Therefore we will refer to $\mathbf{XX}'$ as the *sample covariance matrix* from now on. Indeed, by the rank inequality (see Bai and Silverstein [6, Theorem A.44]) we have for the supremum norm

$$\|F_{n^{-1}\mathbf{S}} - F_{n^{-1}\mathbf{XX}'}\| \le p^{-1}\mathrm{rank}((\overline{\mathbf{x}}, \ldots, \overline{\mathbf{x}}))\,,$$

which means that the respective limiting spectral distributions coincide under (1.2). For the same reason we can assume without loss of generality that the entries (X_{it}) are centered whenever the expectations exist.

The limiting spectral distribution of normalized sample covariance matrices was found by Marčenko and Pastur.

Theorem 1.1 (Debashis and Aue [57])**.** *Suppose that* $\mathbf{X}$ *has iid entries with common mean and variance* 1*. If* (1.2) *holds, then, with probability one,* $(F_{n^{-1}\mathbf{XX}'})$ *converges weakly to a non-random distribution, the so-called Marčenko–Pastur law* F_γ*. If* $\gamma \in (0, 1]$*,* F_γ *has density,*

$$f_\gamma(x) = \begin{cases} \frac{1}{2\pi x\gamma}\sqrt{(b-x)(x-a)}\,, & \text{if } a \le x \le b, \\ 0\,, & \text{otherwise,} \end{cases} \tag{1.3}$$

where $a = (1-\sqrt{\gamma})^2$ *and* $b = (1+\sqrt{\gamma})^2$*. If* $\gamma > 1$*, the Marčenko–Pastur law is a mixture of a point mass at* 0 *and the density function* $f_{1/\gamma}$ *with weights* $1 - 1/\gamma$ *and* $1/\gamma$*, respectively.*

The Marčenko–Pastur law describes the global behavior of the eigenvalues of $\mathbf{XX}'$. Theorem 1.1 quantifies the spread of the eigenvalues around their mean 1. Note that the range of the deviation increases when γ increases from 0 to ∞.

If $p/n \to 0$, the limiting spectral distribution in Theorem 1.1 is the Dirac measure at 1. After an appropriate transformation of the sample covariance matrix one can obtain the semi-circle law defined in (1.18) as a non-degenerate limiting spectral distribution in this case.

The crucial assumptions in Theorem 1.1 are the finiteness of the variance and that p and n tend to infinity at the same rate. By Theorem 2.8 in Bai [8], the conclusion still holds if the entries are independent, have common mean and satisfy the Lindeberg-type condition

$$\lim_{n\to\infty} \frac{1}{\delta^2 np} \sum_{i,t} \mathbb{E}[X_{it}^2 \mathbf{1}_{\{|X_{it}|>\delta\sqrt{n}\}}] = 0\,, \quad \delta > 0\,.$$

Many important test statistics in multivariate analysis are functions of the eigenvalues $\lambda_{(1)} \ge \cdots \ge \lambda_{(p)}$ of the sample covariance matrix and can be expressed by means of the empirical spectral distribution of $n^{-1}\mathbf{X}\mathbf{X}'$.

Example 1.2. Let (X_{it}) be iid standard normal and consider

$$T_n := \log(\det(n^{-1}\mathbf{X}\mathbf{X}')) = \log \prod_{i=1}^{p} \frac{\lambda_{(i)}}{n} = p \int_0^\infty \log x \, \, \mathrm{d}F_{n^{-1}\mathbf{X}\mathbf{X}'}(x)\,.$$

On the one hand, if p is fixed we know from Example 1.1 in Yao et a. [78] that

$$\sqrt{\frac{n}{p}}\, T_n \xrightarrow{\mathrm{d}} Y \sim N(0,2)\,, \quad n \to \infty\,. \tag{1.4}$$

On the other hand, if $p/n \to \gamma \in (0,1)$, one can use Theorem 1.1 to obtain asymptotic values for T_n. On gets a.s.

$$\frac{T_n}{p} \to \int_0^\infty \log x \, \, \mathrm{d}F_\gamma(x) = \frac{\gamma - 1}{\gamma} \log(1-\gamma) - 1 < 0\,,$$

which implies $\sqrt{n/p}\, T_n \to -\infty$ a.s. In view of (1.4) the asymptotic behavior of the test statistic T_n crucially depends on the dimension p.

A characterization of the limiting spectral distribution of sample covariance matrices with general population covariance has been derived for many settings. As an example we state Theorem 1.1 in Bai and Zhou [7].

Theorem 1.3. *Assume* (1.2) *and the following conditions.*

- *For all k, $\mathbb{E}[X_{jk}X_{lk}] = T_{lj}$, and for any non-random $p \times p$ matrix $\mathbf{B}$ with bounded norm, $\mathbb{E}[(\mathbf{x}_k'\mathbf{B}\mathbf{x}_k - \mathrm{tr}(\mathbf{B}\mathbf{T}))^2] = o(n^{-2})$, where $\mathbf{T} = \mathbf{T}_n = (T_{jl})$ and $\mathbf{x}_k$ are the columns of $\mathbf{X}$.*
- *The norm of $\mathbf{T}_n$ is uniformly bounded and $F_{\mathbf{T}_n}$ tends to a non-random probability distribution H.*

Then, with probability 1, *$F_{n^{-1}\mathbf{X}\mathbf{X}'}$ tends to a probability distribution, whose Stieltjes transform $m(z)$ satisfies*

$$m(z) = \int_{\mathbb{R}} \frac{1}{t(1-\gamma-\gamma z m(z)) - z} \, \mathrm{d}H(t)\,. \tag{1.5}$$

If $\underline{m}(z) = -(1-\gamma)/z + \gamma m(z)$, then (1.5) *becomes*

$$z = -\frac{1}{\underline{m}(z)} + \gamma \int_{\mathbb{R}} \frac{1}{1+\underline{m}(z)\, t} \, \mathrm{d}H(t)\,. \tag{1.6}$$

For historical reasons, (1.5) is often called *Marčenko–Pastur equation*. In practice, the inversion of such integral equations can be very difficult. Therefore, a characterization of the limiting spectral distribution via (1.5) is of limited use. For numerical procedures, the *Silverstein equation* (1.6) is sometimes preferred. Roughly speaking, the only known explicit examples of non-degenerate limiting spectral distributions are the Marčenko–Pastur law, the circular law, the semi-circle law and the multivariate F-matrix. In contrast, the number of existence results is huge. We refer to the discussion in Yao et al. [78] for further details.

In Theorem 1.1 we have seen that the sequence $(F_{n^{-1}\mathbf{XX}'})$ converges to the Marčenko–Pastur law if the iid entries possess a finite second moment. Now we will discuss the situation when the entries are still iid, but have an infinite variance. Here we assume the entries to be regularly varying with index $\alpha \in (0,2)$; see (RV_α) on page v for the definition of regular variation. Assuming (1.2) with $\gamma \in (0,1]$ in this infinite variance case, Belinschi et al. [15, Theorem 1.10] showed that the sequence $(F_{a_{n+p}^{-2}\mathbf{XX}'})$ converges with probability one to a non-random probability measure with density ρ_α^γ satisfying

$$\rho_\alpha^\gamma(x)x^{1+\alpha/2} \to \frac{\alpha\gamma}{2(1+\gamma)}, \quad x \to \infty\,; \tag{1.7}$$

see also Ben Arous and Guionnet [17, Theorem 1.6]. Here the normalization (a_k) is defined such that

$$\mathbb{P}(|X| > a_k) \sim k^{-1}\,, \quad k \to \infty\,. \tag{1.8}$$

An application of the Potter bounds (see Bingham et al. [20, p. 25]) shows that $a_{n+p}^2/n \to \infty$. To the best of our knowledge, explicit expressions or computational methods for the limiting spectral distribution in the infinite variance case are not available at this moment.

1.3 Limits of extreme eigenvalues under finite fourth moment

After the limiting spectral distribution for sample covariance matrices had been found, the focus shifted to the asymptotic behavior of the largest and smallest eigenvalues $\lambda_{(1)}$ and $\lambda_{(p)}$, respectively, of $\mathbf{XX}'$. In this section, we assume that the entries of the data matrix $\mathbf{X}$ are iid with generic element X. Furthermore, suppose $p \le n$; otherwise $\lambda_{(p)} = 0$ since $\mathbf{XX}'$ has at most $\min(n,p)$ non-zero eigenvalues. We will discuss the setting $\mathbb{E}[X^4] < \infty$, while the case $\mathbb{E}[X^4] = \infty$ is treated in Section 1.5.

Under condition (1.2) with $\gamma \le 1$, one can infer from Theorem 1.1 that

$$\liminf_{n\to\infty} \frac{\lambda_{(1)}}{n} \ge (1+\sqrt{\gamma})^2 \quad \text{and} \quad \limsup_{n\to\infty} \frac{\lambda_{(p)}}{n} \le (1-\sqrt{\gamma})^2 \quad \text{a.s.} \tag{1.9}$$

We follow Bai and Silverstein [6] and derive necessary conditions for the a.s. convergence of $n^{-1}\lambda_{(1)}$. Since the largest diagonal entry of a matrix is bounded by its largest eigenvalue, we have

$$\frac{\lambda_{(1)}}{n} \ge \max_{i=1,\dots,p} \frac{1}{n}\sum_{t=1}^n X_{ii}^2\,. \tag{1.10}$$

If $\mathbb{E}[X^4] = \infty$, then by Lemma B.25 in [6]

$$\limsup_{n\to\infty} \max_{i=1,\dots,p} \frac{1}{n}\sum_{t=1}^n X_{ii}^2 = \infty \quad \text{a.s.} \tag{1.11}$$

If $\mathbb{E}[X^4] < \infty$ and $\mathbb{E}[X] = c \neq 0$, then

$$\frac{1}{\sqrt{n}}\|\mathbf{X}\|_2 \geq \frac{1}{\sqrt{n}}\|\mathbb{E}[\mathbf{X}]\|_2 - \frac{1}{\sqrt{n}}\|\mathbf{X} - \mathbb{E}[\mathbf{X}]\|_2 \geq \frac{|c|p}{\sqrt{n}} - \frac{1}{\sqrt{n}}\|\mathbf{X} - \mathbb{E}[\mathbf{X}]\|_2 \to \infty \quad \text{a.s.},$$

where for any matrix $\mathbf{A}$, $\|\mathbf{A}\|_2$ denotes its spectral norm, i.e., its largest singular value. This and (1.11) show that $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^4] < \infty$ are necessary conditions for the a.s. convergence of $n^{-1}\lambda_{(1)}$.

In what follows, we assume $\mathbb{E}[X] = 0$ *and* $\mathbb{E}[X^2] = 1$*, whenever the respective moments exist.*

1.3.1 Sample covariance matrices

Extending work by Geman [38], Bai et al. [79] showed under (1.2) that

$$\frac{\lambda_{(1)}}{n} \stackrel{\text{a.s.}}{\to} \left(1 + \sqrt{\gamma}\right)^2, \qquad n \to \infty, \tag{1.12}$$

which is the optimal result in view of (1.11). Later Bai and Yin [10] proved the following result under the additional assumption $\gamma \in (0, 1)$:

$$\limsup_{n\to\infty} \|n^{-1}\mathbf{X}\mathbf{X}' - (1+\gamma)\mathbf{I}\|_2 \leq 2\sqrt{\gamma} \quad \text{a.s.} \tag{1.13}$$

Because of

$$\|\mathbf{R} - (1+\gamma)\mathbf{I}\|_2 = \max\{\lambda_{(1)}/n - (1+\gamma), -\lambda_{(p)}/n + (1+\gamma)\},$$

equations (1.13) and (1.9) imply

$$\lim_{n\to\infty} \frac{\lambda_{(1)}}{n} = (1+\sqrt{\gamma})^2 \quad \text{and} \quad \lim_{n\to\infty} \frac{\lambda_{(p)}}{n} = (1-\sqrt{\gamma})^2 \quad \text{a.s.} \tag{1.14}$$

The approach based on (1.13) provides a lower bound on the smallest eigenvalue, which is difficult to obtain in general. Unfortunately, one cannot gain any information about the minimal conditions for the existence of a limit of $\lambda_{(p)}/n$ since the method treats $\lambda_{(1)}$ and $\lambda_{(p)}$ simultaneously and therefore it can (at best) only be applied in the most general setting for $\lambda_{(1)}$, losing sharpness for $\lambda_{(p)}$. It was finally discovered by Tikhomirov [70] that the a.s. limit of $n^{-1}\lambda_{(p)}$ is given by (1.14) if $\mathbb{E}[X^2] = 1$, whereas higher moments can be infinite.

Under suitable moment assumptions, $\lambda_{(1)}$ and $\lambda_{(p)}$ possess *Tracy–Widom* fluctuations around their almost sure limits. For instance, Johnstone [48] complemented (1.14) by the corresponding central limit theorem in the special case of iid standard normal entries:

$$n^{2/3} \frac{(\sqrt{\gamma})^{1/3}}{\left(1+\sqrt{\gamma}\right)^{4/3}} \Big(\frac{\lambda_{(1)}}{n} - \big(1 + \sqrt{\tfrac{p}{n}}\big)^2\Big) \stackrel{\text{d}}{\to} \xi,$$

where the limiting random variable has a *Tracy–Widom distribution* of order 1. Its distribution function F_1 is given by

$$F_1(s) = \exp\Big\{ -\frac{1}{2}\int_s^\infty [q(x) + (x-s)q^2(x)]\,\mathrm{d}x\Big\},$$

where $q(x)$ is the unique solution to the Painlevé II differential equation

$$q''(x) = xq(x) + 2q^3(x),$$

where $q(x) \sim \mathrm{Ai}(x)$ as $x \to \infty$ and $\mathrm{Ai}(\cdot)$ is the Airy kernel; see Tracy and Widom [71] for details.

1.3.2 Sample correlation matrices

In comparison with the eigenvalues of $\mathbf{XX}'$, much less is known about the ordered eigenvalues

$$\mu_{(1)} \geq \cdots \geq \mu_{(p)}$$

of the *sample correlation matrix* $\mathbf{R} = \mathbf{YY}'$ with entries

$$R_{ij} = \sum_{t=1}^{n} \frac{X_{it}X_{jt}}{\sqrt{D_i}\sqrt{D_j}} = \sum_{t=1}^{n} Y_{it}Y_{jt}\,, \quad i,j = 1,\dots,p\,. \tag{1.15}$$

In this thesis we will often make use of the notation $\mathbf{Y} = (Y_{it}) = (X_{it}/\sqrt{D_i})$ and

$$D_i = D_i^{(n)} = \sum_{t=1}^{n} X_{it}^2\,, \qquad i = 1,\dots,p;\ n \geq 1\,. \tag{1.16}$$

With $\mathbf{F} = \operatorname{diag}(1/D_1,\dots,1/D_p)$, we have $\mathbf{R} = \mathbf{F}^{1/2}\mathbf{XX}'\mathbf{F}^{1/2}$ which has the same eigenvalues as $\mathbf{XX}'\mathbf{F}$. Weyl's inequality (see [19]) yields

$$\begin{aligned} \max_{i=1,\dots,p} |\mu_{(i)} - n^{-1}\lambda_{(i)}| &\leq \|\mathbf{XX}'\mathbf{F} - n^{-1}\mathbf{XX}'\|_2 \\ &\leq n^{-1}\|\mathbf{XX}'\|_2\|n\mathbf{F} - \mathbf{I}\|_2 \\ &= n^{-1}\lambda_{(1)} \max_{i=1,\dots,p} \Big|\frac{n}{D_i} - 1\Big|\,, \end{aligned} \tag{1.17}$$

which converges a.s. to 0 if $\mathbb{E}[X^4] < \infty$; see Chapter 4 for details. This approach was used by Jiang [46], and Xiao and Zhou [77] to derive

$$\mu_{(1)} \to (1+\sqrt{\gamma})^2 \quad \text{and} \quad \mu_{(p)} \to (1-\sqrt{\gamma})^2 \quad \text{a.s.}$$

1.4 Eigenvectors

Eigenvectors of large random matrices and graphs play an essential role in statistical analysis, physics and computer science. Many properties of a matrix or a graph can be deduced from its eigenvectors. Popular algorithms for data analysis such as spectral clustering, principal component analysis, PageRank and community detection are based on the eigenvector-eigenvalue decomposition of a matrix.

If the data matrix $\mathbf{X}$ has iid standard normal entries, then $\mathbf{XX}'$ is a *Wishart matrix*, whose eigenvectors are well studied; see for example Bai and Silverstein [6, Ch. 10]. Due to the invariance of $\mathbf{XX}'$ under orthogonal transformations, its matrix of properly normalized eigenvectors is Haar distributed, i.e., the distribution is uniform on the space of orthogonal $p \times p$ matrices. This result has been extended to different classes of matrices $\mathbf{X}$ by direct comparison with Wishart matrices. Such statements are called universality results and they often require that the new entry distribution is in some sense similar to the standard normal distribution. This is often achieved by moment conditions. Silverstin [64] showed that the matrix of eigenvectors is asymptotically Haar distributed as $p/n \to \gamma \in (0,\infty)$ if the first four moments of the iid entries coincide with those of the standard normal distribution. On the one hand, this means that eigenvectors of $\mathbf{XX}'$ are completely unstructured for light-tailed entry distributions. On the other hand, the extreme eigenvalues converge to constants a.s.

Although eigenvectors play a minor role in this thesis, we summarize some results from the literature on the light-tailed case. Consequently, our approximations of eigenvectors in the heavy-tailed case are put into context. The majority of studies on eigenvectors of large random matrices is conducted on Wigner matrices $\mathbf{W} = (W_{ij})$. They are symmetric, real-valued, $n \times n$ matrices with entries $W_{ij}, 1 \le i \le j \le n$, that are iid, mean zero and unit variance random variables.

Roughly speaking, the spectral properties of the square of a Wigner matrix and a sample covariance matrix with $p = n$ are quite similar if they share the same sufficiently light-tailed entry distribution. For the purpose of exposition, we look at $\mathbf{W}^2 = \mathbf{W}\mathbf{W}'$ instead of $\mathbf{X}\mathbf{X}'$. Indeed, apart from the additional symmetry restriction for $\mathbf{W}$ they have the same structure.

By Theorem 2.1 in [5], the limiting spectral distribution of $(\mathbf{W}/\sqrt{n})$ is the *semi-circle law* G with density

$$g(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\,\mathbf{1}_{\{|x|\le 2\}}\,. \tag{1.18}$$

The semi-circle law and the Marčenko–Pastur law F_1 are linked in the following way: if $Y \sim G$ then $Y^2 \sim F_1$.

Since $\mathbf{W}^2$ and $\mathbf{W}$ have the same eigenvectors it will be sufficient to study the latter. In the remainder of this section we list some properties of $\mathbf{v}_1, \ldots, \mathbf{v}_n$, where $\mathbf{v}_i$ is an eigenvector associated to the ith largest eigenvalue of $\mathbf{W}$. In addition, we assume $\mathbf{v}_i$ are unit vectors, i.e. $\|\mathbf{v}_i\|_{\ell_2} = 1$, and that the first non-zero component of $\mathbf{v}_i$ is positive.

Often the entries are assumed to be *sub-exponential.* We call a random variable W *sub-exponential* with exponent $\alpha > 0$ if there exists a constant $\beta > 0$ such that

$$\mathbb{P}(|W| > x) \le \beta \exp(-x^\alpha/\beta)\,, \quad x > 0\,.$$

Theorem 1.4 in O'Rourke et al. [56] focuses on Wigner matrices with sub-exponential entries with exponent α. There exists a constant $C_\alpha > 0$ such that the probability that the spectrum of $\mathbf{W}$ is simple and that every coordinate of every $\mathbf{v}_i$ is non-zero is at least $1 - C_\alpha n^{-\alpha}$. By our sign convention, the eigenvectors are unique with high probability. In particular, if W_{11} is standard normal, $\mathbf{v}_i$ is uniformly distributed on

$$\mathcal{S}_+^{n-1} := \{\mathbf{x} = (x_1, \ldots, x_n)' : \|\mathbf{x}\|_{\ell_2} = 1 \text{ and } x_1 > 0\}\,.$$

An eigenvector of a Wigner matrix with light-tailed entry distribution (for instance sub-exponential) behaves like a random vector uniformly distributed on $\mathcal{S}_+^{n-1}$. More precise quantitative statements are difficult to obtain. For details we refer to [56].

If $\mathbf{v} = (v_1, \ldots, v_n)'$ is a random vector uniformly distributed on $\mathcal{S}_+^{n-1}$, probabilistic bounds on its coordinates are available. By Theorem 2.1 in [56], we have

$$v_{\max} := \max_{i=1,\ldots,n} |v_i| \le C\sqrt{\frac{\log n}{n}} \quad \text{and} \quad v_{\min} := \min_{i=1,\ldots,n} |v_i| \ge \frac{c}{n^{3/2}} \tag{1.19}$$

with probability $1 - o(1)$ for any $C > 1$ and $c \in [0, 1)$.

In RMT it is common to study the so-called *bulk* and *edge* spectra separately. For $\varepsilon \in (0, 1)$ one distinguishes between the *bulk* eigenvectors $\mathbf{v}_i, i \in \{1 \le t \le n : \varepsilon n \le t \le (1-\varepsilon)n\} := B_\varepsilon$ and the *edge* eigenvectors $\mathbf{v}_i, i \in \{1, \ldots, n\}\backslash B_\varepsilon$. The associated eigenvalues are usually referred to as bulk and edge spectrum, respectively. Roughly speaking, the limiting spectral distribution of a sequence of random matrices depends on the bulk spectrum, while the edge spectrum influences the behavior of functionals of the eigenvalues.

The behavior of the largest coordinates of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ was studied in the case of certain sub-exponential entries with exponent $\alpha = 2$ whose first four moments match those of the standard normal distribution. By Theorems 4.1 and 4.3 in [56], one has for any bulk eigenvector $\mathbf{v}_i$ and appropriate constants $c_1, c_2 > 0$,

$$c_1\sqrt{\frac{\log n}{n}} \le \mathbf{v}_{i,\max} \le c_2\sqrt{\frac{\log n}{n}}$$

with probability $1-o(1)$. This result is astonishingly precise. Note that up to logarithmic corrections $\mathbf{v}_{i,\max}$ is of the smallest possible magnitude $n^{-1/2}$. This property is called *complete delocalization*; see [62]. For edge eigenvectors, however, it is proved that for a constant $c_3 > 0$,

$$\mathbf{v}_{i,\max} \le c_3\frac{\log n}{\sqrt{n}}$$

with probability $1-o(1)$. The optimal bound in the edge case remains an open problem. Additionally, by Corollary 5.4 in [56] the ℓ_p-norms, $1 \le p \le 2$, of the $(\mathbf{v}_i)$ are of the same order of magnitude:

$$c_0 n^{1/p-1/2} \le \min_{i=1,\dots,n} \|\mathbf{v}_i\|_{\ell_p} \le \max_{i=1,\dots,n} \|\mathbf{v}_i\|_{\ell_p} \le C_0 n^{1/p-1/2} \tag{1.20}$$

with probability $1-o(1)$ for positive constants c_0, C_0.

If the entry distribution of $\mathbf{W}$ has heavy tails, the behavior of its eigenvectors is completely different. In [18], Benaych-Georges and Péché assumed W_{11} to be regularly varying with index $\alpha \in (0,4)$; see (RV_α) on page v, and provided an approximation of $\mathbf{v}_k$ for any fixed k. From their asymptotic result one can deduce that $\mathbf{v}_{k,\max}$ converges to $1/\sqrt{2}$ a.s. Moreover, asymptotically there are only two coordinates of $\mathbf{v}_k$ with non-zero mass. Both are of magnitude $1/\sqrt{2}$. This is the opposite of complete delocalization: *complete localization*. The number of non-zero coordinates is bounded.

Note that in the presence of heavy tails the eigenvectors of $\mathbf{W}$ and $\mathbf{XX}'$ are very different; see (1.37) in Section 1.5.

1.5 Contribution of this thesis

In this section we summarize our contribution to the spectral theory of high-dimensional sample covariance and correlation matrices. We focus on the case where the entries of the $(p \times n)$-dimensional data matrix $\mathbf{X}$ have an infinite fourth moment.

1.5.1 Sample covariance matrices: the iid case

While the finite second moment is the central assumption to obtain the Marčenko–Pastur law as the limiting spectral distribution, the finite fourth moment plays a crucial role when studying the eigenvalues

$$\lambda_{(1)} \ge \cdots \ge \lambda_{(p)} \tag{1.21}$$

of the sample covariance matrix $\mathbf{XX}'$. Unless stated otherwise, the entries (X_{it}) are iid regularly varying random variables with index $\alpha \in (0,4)$ (see (RV_α)) and X is a generic random variable with the same distribution. This implies $\mathbb{E}[X^4] = \infty$. Here and in what follows, we will refer to this setting as the heavy-tailed case, in contrast to the light-tailed case in which $\mathbb{E}[X^4]$ is finite. Moreover, we assume $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = 1$, whenever $\mathbb{E}[X^2] < \infty$.

We normalize the eigenvalues $(\lambda_{(i)})$ by (a_{np}^2) where the sequence (a_k) is chosen such that

$$\mathbb{P}(|X| > a_k) \sim k^{-1}\,, \quad k \to \infty.$$

Standard theory for regularly varying functions (e.g. Bingham et al. [20], Feller [37]) yields that $a_n = n^{1/\alpha}\ell(n)$ where ℓ is a slowly varying function. Assuming the usual growth condition (1.2) for p, the Potter bounds (see [20, p. 25]) yield for $\alpha \in (0,4)$ that

$$\frac{a_{np}^2}{n} \sim \frac{n^{4/\alpha}\gamma^{2/\alpha}\,\ell^2(n^2\gamma)}{n} \to \infty, \qquad n \to \infty\,, \tag{1.22}$$

i.e., the normalization a_{np}^2 is stronger than n.

By (1.10), we have $\lambda_{(1)} \ge X^2_{(1),np}$, where $X^2_{(1),np} \ge \cdots \ge X^2_{(np),np}$ denote the order statistics of $(X_{it}^2)_{i=1,\ldots,p;t=1,\ldots,n}$. Classical extreme value theory yields that $a_{np}^{-2}X^2_{(1),np}$ converges weakly to a *Fréchet distribution* with parameter $\alpha/2$:

$$\Phi_{\alpha/2}(x) = \mathrm{e}^{-x^{-\alpha/2}}\,, \qquad x > 0\,. \tag{1.23}$$

The theory for the largest eigenvalues of sample covariance matrices with heavy tails is less developed than in the light-tailed case. Pioneering work for $\lambda_{(1)}$ under the growth condition (1.2) and $\alpha \in (0,2)$ is due to Soshnikov [65, 66]. For $k \ge 1$ fixed, he showed that

$$\frac{\lambda_{(m)}}{X^2_{(m),np}} \stackrel{\mathbb{P}}{\to} 1\,, \quad n \to \infty\,, 1 \le m \le k\,, \tag{1.24}$$

which reveals that the limiting distribution of $a_{np}^{-2}X^2_{(1),np}$ is the Fréchet distribution (1.23). Furthermore he proved the point process convergence

$$N_n = \sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}\lambda_i} \stackrel{\mathrm{d}}{\to} N_\Gamma = \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}}\,, \qquad n \to \infty\,. \tag{1.25}$$

Here ε_y is the Dirac measure at y,

$$\Gamma_i = E_1 + \cdots + E_i\,, \qquad i \ge 1\,, \tag{1.26}$$

and (E_i) is a sequence of iid standard exponential random variables. In other words, N_Γ is a Poisson point process on $(0,\infty)$ with mean measure $\mu(x,\infty) = x^{-\alpha/2}$, $x > 0$. Convergence in distribution of point processes is understood in the sense of weak convergence in the space of point measures equipped with the vague topology; see Resnick [60, 61].

Later Auffinger et al. [4] established (1.25) also for $\alpha \in [2,4)$. In their proofs they used truncation techniques and a combinatorial approach.

General growth rates for p_n

In many applications it is not realistic to assume that the dimension p of the data and the sample size n grow at the same rate. In the light-tailed case little is known when p and n grow at different rates, i.e., $\lim p/n \in \{0,\infty\}$. Notable exceptions are El Karoui [30] who proved that Johnstone's result in [48] (assuming iid standard normal entries) remains valid when $p/n \to 0$ or $n/p \to \infty$, and Péché [58] who showed universality results for the largest eigenvalues of some sample covariance matrices with non-Gaussian entries.

The aforementioned results of Soshnikov [65, 66] and Auffinger et al. [4] already indicate that the value γ in the usual growth rate (1.2) does not appear in the distributional limits. In the heavy-tailed case, more general growth of p than prescribed by (1.2) has been used in Davis et al. [24, 25]. In Chapters 2 and 3 we will consider power-law growth rates on the dimension (p_n). To be precise, we assume an integer sequence

$$p = p_n = n^\beta l(n), \quad n \geq 1\,, \tag{1.27}$$

where l is a slowly varying function and $\beta \geq 0$. If $\beta = 0$, we also assume $l(n) \to \infty$. Our condition (1.27) is more general than the growth conditions in the literature; see [4, 24, 25].

Note that the matrices $\mathbf{XX}'$ and $\mathbf{X}'\mathbf{X}$ have the same non-zero eigenvalues. Therefore it is sufficient to consider $\beta \in [0, 1]$. For details we refer to Chapters 2 and 3.

Our contribution

In the heavy-tailed case and under the growth condition (1.27) with $\beta \in [0, 1]$ we prove with considerable technical effort that

$$a_{np}^{-2} \|\mathbf{XX}' - \operatorname{diag}(\mathbf{XX}')\|_2 \xrightarrow{\mathbb{P}} 0\,, \qquad n \to \infty\,, \tag{1.28}$$

where $\operatorname{diag}(\mathbf{XX}')$ denotes the diagonal of $\mathbf{XX}'$; see Theorem 2.5.

The employed techniques originate from extreme value analysis and large deviation theory. The (i, j) entry of $\mathbf{XX}'$ is

$$(\mathbf{XX}')_{ij} = \sum_{t=1}^{n} X_{it} X_{jt}\,, \quad i, j = 1 \ldots, p\,.$$

From Embrechts and Veraverbeke [36] we know that X^2 and $X_{11}X_{12}$ are regularly varying with indices $\alpha/2$ and α, respectively. By large deviation theory (see (3.12)), the diagonal and off-diagonal entries of $\mathbf{XX}'$ inherit the tails of X_{it}^2 and $X_{it}X_{jt}$, $i \neq j$, respectively, above some high threshold. Therefore the random variables in the diagonal of $\mathbf{XX}'$ have the heaviest tail. They dominate the spectral behavior of $\mathbf{XX}'$ and thus (1.28) is not unexpected.

Equation (1.28) has some immediate consequences for the approximation of the eigenvalues of $\mathbf{XX}'$ by those of $\operatorname{diag}(\mathbf{XX}')$. Indeed, let $\mathbf{C}$ be a symmetric $p \times p$ matrix with eigenvalues

$$\lambda_{(1)}(\mathbf{C}) \geq \cdots \geq \lambda_{(p)}(\mathbf{C})\,. \tag{1.29}$$

Then for any symmetric $p \times p$ matrices $\mathbf{A}, \mathbf{B}$, by *Weyl's inequality* (see Bhatia [19]),

$$\max_{i=1,\ldots,p} \big|\lambda_{(i)}(\mathbf{A} + \mathbf{B}) - \lambda_{(i)}(\mathbf{A})\big| \leq \|\mathbf{B}\|_2\,.$$

If we now choose $\mathbf{A} + \mathbf{B} = \mathbf{XX}'$ and $\mathbf{A} = \operatorname{diag}(\mathbf{XX}')$ we obtain

$$a_{np}^{-2} \max_{i=1,\ldots,p} \big|\lambda_{(i)} - \lambda_{(i)}(\operatorname{diag}(\mathbf{XX}'))\big| \xrightarrow{\mathbb{P}} 0\,, \quad n \to \infty\,. \tag{1.30}$$

Thus the problem of deriving limit theory for the order statistics of $\mathbf{XX}'$ has been reduced to limit theory for the order statistics of the iid row-sums

$$D_i = (\mathbf{XX}')_{ii} = \sum_{t=1}^{n} X_{it}^2\,, \qquad i = 1, \ldots, p\,,$$

which are the eigenvalues of diag($\mathbf{XX}'$). This theory is completely described by the point processes constructed from the points D_i/a_{np}^2 $i = 1, \dots, p$. Necessary and sufficient conditions for the weak convergence of these point processes are provided by Lemma 3.22. In combination with the Nagaev-type large deviation results of Theorem 3.21 they yield the following result under (1.27):

$$\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}(D_i - c_n)} \xrightarrow{\mathrm{d}} N_\Gamma \,, \qquad n \to \infty \,, \tag{1.31}$$

where N_Γ was defined in (1.25) and $c_n = 0$ if $\mathbb{E}[D] = \infty$ and $c_n = \mathbb{E}[D] = n\mathbb{E}[Z^2]$ otherwise. Note that the centering c_n in the finite variance case can be avoided if $n/a_{np}^2 \to 0$. The latter condition is satisfied if

$$\beta > \alpha/2 - 1 \,. \tag{1.32}$$

Combining (1.28), (1.30) and (1.31), we conclude that (1.25) holds under the general growth rate (1.27) with $\beta \in [0,1]$, where for $\alpha \in [2,4)$ one needs to require (1.32).

The limiting point process (1.25) yields a plethora of ancillary results. For example, one can easily derive the limiting distribution of $a_{np}^{-2}\lambda_{(k)}$ for fixed $k \geq 1$:

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(a_{np}^{-2}\lambda_{(k)} \leq x) &= \lim_{n\to\infty} \mathbb{P}(N_n(x,\infty) < k) = \mathbb{P}(N(x,\infty) < k) \\ &= \mathbb{P}(\Gamma_k^{-2/\alpha} \leq x) = \sum_{s=0}^{k-1} \frac{\big(x^{-\alpha/2}\big)^s}{s!} \, \mathrm{e}^{-x^{-\alpha/2}}, \quad x > 0 \,. \end{aligned}$$

Another immediate consequence of (1.25) is

$$a_{np}^{-2}\big(\lambda_{(1)}, \dots, \lambda_{(k)}\big) \xrightarrow{\mathrm{d}} \big(\Gamma_1^{-2/\alpha}, \dots, \Gamma_k^{-2/\alpha}\big) \tag{1.33}$$

for any fixed $k \geq 1$ and $\alpha \in (0,2]$. In Chapter 3 we show for $\alpha \in (2,4)$

$$a_{np}^{-2}\big(\lambda_{(1)} - n\,\mathbb{E}[Z^2], \dots, \lambda_{(k)} - n\,\mathbb{E}[Z^2]\big) \xrightarrow{\mathrm{d}} \big(\Gamma_1^{-2/\alpha}, \dots, \Gamma_k^{-2/\alpha}\big) \,. \tag{1.34}$$

Equations (1.33) and (1.34) yield that for $\alpha \in (0,4)$ and any fixed $k \geq 1$,

$$a_{np}^{-2}\big(\lambda_{(1)} - \lambda_{(2)}, \dots, \lambda_{(k)} - \lambda_{(k+1)}\big) \xrightarrow{\mathrm{d}} \big(\Gamma_1^{-2/\alpha} - \Gamma_2^{-2/\alpha}, \dots, \Gamma_k^{-2/\alpha} - \Gamma_{k+1}^{-2/\alpha}\big). \tag{1.35}$$

Related results were also derived for linear spectral statistics such as the trace $a_{np}^{-2}(\lambda_1 + \cdots + \lambda_p)$. We refer to Chapter 3 and Davis et al. [24] for details on the proofs and more examples.

In the case of fixed p, Janssen et al. [45] related the limiting distribution of the eigenvalues $(\lambda_{(i)})$ to stable distributions. They also used (1.28) and (1.30). In this case it is clear that for example

$$\frac{\lambda_{(2)}}{X_{(2),np}^2} \xrightarrow{\mathbb{P}} 1 \,, \quad n \to \infty \,,$$

cannot hold. If $X_{(1),np}^2$ and $X_{(2),np}^2$ lie in the same row of $\mathbf{X}$, then they appear on the same spot on the diagonal of $\mathbf{XX}'$. Then $X_{(2),np}^2$ cannot be used for the approxiamtion of $\lambda_{(2)}$ in view of (1.30). Indeed, the probability that this happens is approximately $1/p$ which does not tend to 0 if p is fixed. This is in contrast to (1.24).

Recall that if p is fixed, one has to use the approximation of the eigenvalues provided by (1.30), while if $p/n \to \gamma$ one can use either (1.30) or (1.24). There is a phase change of the behavior of $(\lambda_{(i)})$ when going from finite p to p proportional to n. In our condition (1.27) the growth of p is essentially described by the parameter β.

Under (1.27) our Theorem 2.1 asserts that

$$a_{np}^{-2} \max_{i=1,\dots,p} \big|\lambda_{(i)} - X^2_{(i),np}\big| \xrightarrow{\mathbb{P}} 0\,, \tag{1.36}$$

provided (1.32) holds. Due to its simplicity, (1.36) is an elegant result. It reveals, for example, that the largest eigenvalue of a high-dimensional heavy-tailed matrix behaves like the maximum of its iid entries.

In view of Lemma 2.22, condition (1.32) describes the precise β-region, up to the slowly varying function in the tail of X, where $\max_i a_{np}^{-2}|D_{L_i} - X^2_{(i),np}|$ is sufficiently small. Here L_i encodes the location of the ith largest diagonal element of $\mathbf{XX}'$; see (2.12) for the formal definition of L_i. Therefore the critical value of β at which the aforementioned phase change occurs is $\max(0, \alpha/2 - 1)$.

The study of eigenvectors of heavy-tailed sample covariance matrices is a fresh topic which has not been explored in the literature listed here. Our approximation of $\mathbf{XX}'$ in (1.28) and the limiting distributions of the spacings (1.35) can be applied to approximate the unit eigenvectors $(\mathbf{v}_j)$ of $\mathbf{XX}'$, where $\mathbf{v}_j$ is associated to $\lambda_{(j)}$. As for the eigenvectors of Wigner matrices we use the convention that their first non-zero coordinate is positive. From (1.28) we know that $\mathbf{XX}'$ is approximated in spectral norm by $\operatorname{diag}(\mathbf{XX}')$. The unit eigenvectors of $\operatorname{diag}(\mathbf{XX}')$ are the canonical basis vectors $\mathbf{e}_j \in \mathbb{R}^p$, $j = 1, \dots, p$.

By Theorem 2.11, $(\mathbf{e}_j)$ approximate the eigenvectors $(\mathbf{v}_j)$. For $\beta \in [0,1]$ and any fixed $k \ge 1$, we have

$$\|\mathbf{v}_k - \mathbf{e}_{L_k}\|_{\ell_2} \xrightarrow{\mathbb{P}} 0\,, \quad n \to \infty\,. \tag{1.37}$$

1.5.2 Sample covariance matrices: the non-iid case

Davis et al. [25] extended the results of Soshnikov [65, 66] and Auffinger et al. [4] to the case where the rows of $\mathbf{X}$ are iid linear processes with iid regularly varying noise. After a multiplication of the mean measure μ by a constant the Poisson point process convergence result of (1.25) remains valid. Pfaffel and Schlemm [59] described the Stieltjes transform of the limiting spectral distribution in this model. Different limit processes can only be expected if there is dependence in both directions: in Chapter 3 we use a model for (X_{it}) which allows for linear dependence across the rows and through time (see also [24]):

$$X_{it} = \sum_{l\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} h_{kl} Z_{i-k,t-l}\,, \qquad i,t \in \mathbb{Z}\,, \tag{1.38}$$

where $(Z_{it})_{i,t\in\mathbb{Z}}$ is a field of iid regularly varying random variables with index $\alpha \in (0,4)$ and $(h_{kl})_{k,l\in\mathbb{Z}}$ is an array of real numbers. Moreover, we require the summability condition

$$\sum_{l\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} |h_{kl}|^\delta < \infty \tag{1.39}$$

for some $\delta \in (0, \min(\alpha/2, 1))$ which ensures the a.s. absolute convergence of the series in (1.38). Under the condition (1.39), the marginal and finite-dimensional distributions of the field (X_{it}) are regularly varying with index α; see Embrechts et al. [35], Appendix A3.3.

From the field (X_{it}) we construct the $p \times n$ matrices

$$\mathbf{X}_n(s) = (X_{i,t+s})_{i=1,\ldots,p;t=1,\ldots,n}\,, \qquad s = 0,1,2,\ldots\,,$$

As before, we will write $\mathbf{X} = \mathbf{X}_n(0)$. Now we can introduce the (non-normalized) *sample autocovariance matrices*

$$\mathbf{X}_n(0)\mathbf{X}_n(s)'\,, \qquad s = 0,1,2,\ldots\,.$$

We will refer to s as the *lag*. For $s = 0$, we obtain the *sample covariance matrix.* In [P4] (see page v) and Chapter 3, we study the asymptotic behavior of the eigen- and singular values of the sample covariance and autocovariance matrices under the growth condition (1.27).

Theorem 3.7 provides a general approximation result for the ordered singular values of $\mathbf{X}_n(0)\mathbf{X}_n(s)'$. Their behavior is determined by the sums $(\sum_{t=1}^n Z_{it}^2)$ and the singular values of the matrix $\mathbf{M}$ given by

$$(\mathbf{M}(s))_{ij} = \sum_{l\in\mathbb{Z}} h_{i,l}h_{j,l+s}, \qquad i,j \in \mathbb{Z}\,.$$

In Section 3.3.5 the limiting point process of the singular values of $\mathbf{X}_n(0)\mathbf{X}_n(s)'$ is derived. Finally, we mention that our paper [P4] deals with the eigenvectors of $\mathbf{X}_n(0)\mathbf{X}_n(s)'$; see also (1.37). One obtains more interesting structures than in the latter result. In fact, by choosing (h_{kl}) accordingly one can obtain arbitrary eigenvectors. This constitutes a valuable property in principal component analysis.

1.5.3 Sample correlation matrices

In Chapter 4 we study the spectrum of the sample correlation matrix $\mathbf{R}$ defined in (1.15). We assume that the underlying data matrix $\mathbf{X}$ has iid centered entries and the usual growth condition (1.2) holds. Recall the notation $Y_{it} = X_{it}/\sqrt{D_i}$ from p. 7. We will sometimes write $(Y_1,\ldots,Y_n) = (Y_{11},\ldots,Y_{1n})$ and $Y = Y_1$.

Consider the following domain of attraction type-condition for the Gaussian law:

$$\mathbb{E}\big[Y_1Y_2\big] = o(n^{-2}) \quad \text{and} \quad \mathbb{E}\big[Y^4\big] = o(n^{-1})\,, \qquad n\to\infty\,. \tag{1.40}$$

By Giné et al. [39], condition (1.40) holds if the distribution of X is in the domain of attraction of the normal law. We use Theorem 1.3 to show that under (1.40) the sequence $(F_{\mathbf{R}})$ converges weakly to the Marčenko–Pastur law F_γ defined in (1.3); see Theorem 4.3. We prove that the condition (1.40) is necessary. When (1.40) is not valid, the limiting spectral distribution of $(F_{\mathbf{R}})$ (if it exists) must have mean 1, by virtue of the method of moments (see (1.1)). This follows from the fact that the diagonal elements of $\mathbf{R}$ are 1. This together with our approximation of $\mathbb{E}[\beta_k(\mathbf{R})]$ provides some information about this distribution; compare also with (1.7) for the sample covariance case.

Our analysis of the almost sure convergence of the extreme eigenvalues $\mu_{(1)}$ and $\mu_{(p)}$ of $\mathbf{R}$ is carried out for symmetric X. Then condition (1.40) turns into

$$n\,\mathbb{E}\big[Y^4\big] \to 0\,, \qquad n\to\infty\,. \tag{1.41}$$

Theorem 4.5 asserts

$$\mu_{(1)} \to (1+\sqrt{\gamma})^2 \quad \text{and} \quad \mu_{(p)} \to (1-\sqrt{\gamma})^2 \quad \text{a.s.} \tag{1.42}$$

under some condition (C_q) on p. 79 which is slightly more restrictive than (1.41). Condition (C_q) essentially means that the convergence rate of $n\,\mathbb{E}\big[Y^4\big]$ is at least logarithmic. A detailed discussion is given in Section 4.2.

Our proof requires an adequate bound on $\mathbb{E}[\mu_{(1)}^{k_n}]$, where $k_n \to \infty$. To this end, we use the inequality

$$\mathbb{E}[\mu_{(1)}^{k_n}] \le \mathbb{E}[\operatorname{tr}\mathbf{R}^{k_n}] = \sum_{i_1,\dots,i_{k_n}=1}^{p} \sum_{t_1,\dots,t_{k_n}=1}^{n} \mathbb{E}[Y_{i_1 t_{k_n}} Y_{i_1 t_1} \cdots Y_{i_{k_n} t_{k_n-1}} Y_{i_{k_n} t_{k_n}}]$$

and determine those summands on the right-hand side which are largest when weighted by their multiplicities. Employing our *Path-Shortening Algorithm*, which is a novel technique that efficiently uses the inherent structure of sample correlation matrices, their contribution is calculated explicitly. The other summands can –with considerable technical effort– be controlled by (C_q). Note that because of the identity $\mathbb{E}[\operatorname{tr}\mathbf{R}^{k_n}] = p\,\mathbb{E}\big[\int x^{k_n} F_{\mathbf{R}}(dx)\big]$ the behavior of the moments of the empirical spectral distribution is closely linked to the above upper bound.

Equation (1.42) indicates that the a.s. convergence of the extreme eigenvalues of $\mathbf{R}$ does not depend on the finiteness of the fourth or even second moments. This is in stark contrast to the a.s. behavior of $n^{-1}\lambda_{(1)}$, the largest eigenvalue of the sample covariance matrix $n^{-1}\mathbf{X}\mathbf{X}'$. Note that there is a phase transition of the a.s. asymptotic behavior of the extreme eigenvalues at the border between finite and infinite fourth moment of X, while such a transition occurs for the empirical spectral distribution at the border between finite and infinite variance.

The eigenvalues of sample correlation matrices exhibit a "more robust" behavior than their sample covariance analogs. This is perhaps not surprising in view of the *self-normalizing property* of sample correlations. Self-normalization also has the advantage that one does not have to worry about the correct normalization. This is a crucial problem in the study of sample covariance matrices in the case of an infinite fourth moment where one needs a normalization stronger than the classical one; see (1.22). A simulation study in Section 4.3 shows that the asymptotic results for $\mu_{(1)}, \mu_{(p)}, \lambda_{(1)}$ and $\lambda_{(p)}$ work nicely. They can be used to design new statistical tests; see for example (4.13).

1.6 Outlook

The main objective of our work was to find explicit limiting distributions of the eigenvalues of large random matrices and functionals thereof. Thus our theory can be applied in a straightforward way. Sections 1.1-1.5 listed fields where our results can be used. Another example is the analysis of high-frequency data which receives significant interest; see for example Podolskij and Heinrich [40], and Xia and Zheng [75, 76].

For practical purposes it is important to work with arbitrary population covariance matrices. Numerous generalizations and estimation techniques have been developed. For many models the limiting spectral distribution can only be characterized in terms of an integral equation (=Marčenko–Pastur equation) for its Stieltjes transform. Explicit solutions are more involved; see the discussion after Theorem 1.3. In Dobriban [28], an algorithm for calculating the spectral distribution based on certain approximate integral equations for its Stieltjes transform was presented. Contributions like this one breathe life into abstract theoretical results. Research in this direction will attract major interest from the industry.

From a more theoretical point of view, it is interesting to study models with heavy tails in which the asymptotic behavior of the sample covariance matrix $\mathbf{X}\mathbf{X}'$ is not dominated by the squares of the entries of $\mathbf{X}$. Moreover, different tail indices of the rows of $\mathbf{X}$ could make the model more appealing to practitioners. In Chapter 3, we will see that a simple transformation, such as the rank transform, does not entirely overcome this issue.

As regards the sample correlation matrix our methods can be applied to more general models such as spiked covariance/correlation structures. Our technical results in Section 4.4 are of independent interest. They provide a *Path-Shortening Algorithm* for the calculation of bounds for the very high moments of $\mu_{(1)}$. This technique is novel and will be of further use for proving results in Random Matrix Theory. We also conjecture that (1.42) may be proved under (1.40) only.

Finally, in our working papers [P4] and [P5] (see page v), we analyze sample autocovariance matrices and provide limit theory for high-dimensional stochastic volatility matrices. Again, we utilize the large deviations approach propagated in Chapters 2 and 3, now for dependent heavy-tailed time series. The corresponding large deviations theory is available in Mikosch and Wintenberger [53].

Chapter 2

Eigenvalues and eigenvectors of heavy-tailed sample covariance matrices with general growth rates: the iid case

JOHANNES HEINY & THOMAS MIKOSCH
Stochastic Process. Appl. (2016), 29.

Abstract

In this paper we study the joint distributional convergence of the largest eigenvalues of the sample covariance matrix of a p-dimensional time series with iid entries when p converges to infinity together with the sample size n. We consider only heavy-tailed time series in the sense that the entries satisfy some regular variation condition which ensures that their fourth moment is infinite. In this case, Soshnikov [65, 66] and Auffinger et al. [4] proved the weak convergence of the point processes of the normalized eigenvalues of the sample covariance matrix towards an inhomogeneous Poisson process which implies in turn that the largest eigenvalue converges in distribution to a Fréchet distributed random variable. They proved these results under the assumption that p and n are proportional to each other. In this paper we show that the aforementioned results remain valid if p grows at any polynomial rate. The proofs are different from those in [4, 65, 66]; we employ large deviation techniques to achieve them. The proofs reveal that only the diagonal of the sample covariance matrix is relevant for the asymptotic behavior of the largest eigenvalues and the corresponding eigenvectors which are close to the canonical basis vectors. We also discuss extensions of the results to sample autocovariance matrices.

Keywords: Regular variation, sample covariance matrix, independent entries, largest eigenvalues, eigenvectors, point process convergence, compound Poisson limit, Fréchet distribution.

2.1 Introduction

In recent years we have seen a vast increase in the number and sizes of data sets. Science (meteorology, telecommunications, genomics, ...), society (social networks, finance, military and civil intelligence, ...) and industry need to extract valuable information from high-dimensional data sets which are often too large or complex to be processed by traditional means. In order to explore the structure of data one often studies the dependence via (sample) covariances and correlations. Often dimension reduction techniques facilitate further analyzes of large data matrices. For example, *principal component analysis* (PCA) transforms the data linearly such that only a few of the resulting vectors contain most of the variation in the data. These *principal component vectors* are the eigenvectors associated with the largest eigenvalues of the sample covariance matrix.

The aim of this paper is to investigate the asymptotic properties of the largest *eigenvalues* and their corresponding *eigenvectors* for sample covariance matrices of high-dimensional heavy-tailed time series with iid entries. Special emphasis is given to the case when the dimension p and the sample size n tend to infinity simultaneously, not necessarily at the same rate.

Throughout we consider the $p \times n$ data matrix

$$\mathbf{Z} = \mathbf{Z}_n = \big(Z_{it}\big)_{i=1,\ldots,p;t=1,\ldots,n}$$

A column of $\mathbf{Z}$ represents an observation of a p-dimensional time series. We assume that the entries Z_{it} are real-valued, independent and identically distributed (iid), unless stated otherwise. We write Z for a generic element and assume $\mathbb{E}[Z] = 0$ and $\mathbb{E}[Z^2] = 1$ if the first and second moments of Z are finite, respectively. We are interested in limit theory for the eigenvalues $\lambda_1, \ldots, \lambda_p$ of the *sample covariance matrix* $\mathbf{Z}\mathbf{Z}'$ and their ordered values

$$\lambda_{(1)} \geq \cdots \geq \lambda_{(p)}\,. \tag{2.1}$$

In this notation we suppress the dependence of (λ_i) on n. We will only discuss the case when $p \to \infty$; for the finite p case we refer to [3, 54].

2.1.1 The light-tailed case

In random matrix theory a lot of attention has been given to the *empirical spectral distribution function* of the sequence $(n^{-1}\mathbf{Z}\mathbf{Z}')$:

$$F_{n^{-1}\mathbf{Z}\mathbf{Z}'}(x) = \frac{1}{p}\,\#\{1 \leq j \leq p : n^{-1}\lambda_j \leq x\}, \qquad x \geq 0\,, \quad n \geq 1.$$

In the literature convergence results for $(F_{n^{-1}\mathbf{Z}\mathbf{Z}'})$ are established under the assumption that p and n grow at the same rate:

$$\frac{p}{n} \to \gamma \qquad \text{for some } \gamma \in (0,\infty). \tag{2.2}$$

Suppose that the iid entries Z_{it} have mean 0 and variance 1. If (2.2) holds then, with probability one, $(F_{n^{-1}\mathbf{Z}\mathbf{Z}'})$ converges weakly to the Marčenko–Pastur law F_γ. If $\gamma \in (0,1]$, F_γ has density,

$$f_\gamma(x) = \begin{cases} \frac{1}{2\pi x\gamma}\sqrt{(b-x)(x-a)}\,, & \text{if } a \leq x \leq b, \\ 0\,, & \text{otherwise,} \end{cases} \tag{2.3}$$

where $a = (1-\sqrt{\gamma})^2$ and $b = (1+\sqrt{\gamma})^2$. If $\gamma > 1$, the Marčenko–Pastur law is a mixture of a point mass at 0 and the density function $f_{1/\gamma}$ with weights $1-1/\gamma$ and $1/\gamma$, respectively. This mass is intuitively explained by the fact that, with probability 1, $\min(p,n)$ eigenvalues λ_i are non-zero. When $n = (1/\gamma)\, p$ and $\gamma > 1$ the fraction of non-zero eigenvalues is $1/\gamma$ while the fraction of zero eigenvalues is $1-1/\gamma$.

The moment condition $\mathbb{E}[Z^2] < \infty$ is crucial for deriving the Marčenko–Pastur limit law. When studying the largest eigenvalues of the sample covariance matrix $\mathbf{ZZ}'$ the moment condition $\mathbb{E}[Z^4] < \infty$ plays a similarly important role; we assume it in the remainder of this subsection. If (2.2) holds and the iid entries Z_{it} have zero mean and unit variance, Geman [38] showed that

$$\frac{\lambda_{(1)}}{n} \stackrel{\text{a.s.}}{\to} \left(1+\sqrt{\gamma}\right)^2, \qquad n \to \infty\,. \tag{2.4}$$

This means that $\lambda_{(1)}/n$ converges to the right endpoint of the Marčenko–Pastur law in (2.3). Johnstone [48] complemented this strong law of large numbers by the corresponding central limit theorem in the special case of iid standard normal entries:

$$\frac{\lambda_{(1)} - \mu_{n,p}}{\sigma_{n,p}} \stackrel{\text{d}}{\to} \xi, \tag{2.5}$$

where the limiting random variable has a *Tracy–Widom distribution* of order 1 and the centering and scaling constants are

$$\mu_{n,p} = (\sqrt{n-1}+\sqrt{p})^2, \quad \sigma_{n,p} = (\sqrt{n-1}+\sqrt{p})\Big(\frac{1}{\sqrt{n-1}}+\frac{1}{\sqrt{p}}\Big)^{1/3};$$

see Tracy and Widom [71] for details. Ma [50] showed Berry–Esseen-type bounds for (2.5).

Asymptotic theory for the largest eigenvalues of sample covariance matrices with non-Gaussian entries is more complicated; pioneering work is due to Johansson [47]. Johnstone's result was extended to matrices $\mathbf{Z}$ with iid non-Gaussian entries by Tao and Vu [68, Theorem 1.16], assuming that the first four moments of Z match those of the normal distribution. Tao and Vu's result is a consequence of the so-called *Four Moment Theorem* which describes the insensitivity of the eigenvalues with respect to changes in the distribution of the entries. To some extent (modulo the strong moment matching conditions) it shows the universality of Johnstone's limit result (2.5).

In the light-tailed case little is known when p and n grow at different rates, i.e., $\lim p/n \in \{0,\infty\}$. Notable exceptions are El Karoui [30] who proved that Johnstone's result (assuming iid standard normal entries) remains valid when $p/n \to 0$ or $n/p \to \infty$, and Péché [58] who showed universality results for the largest eigenvalues of some sample covariance matrices with non-Gaussian entries.

2.1.2 The heavy-tailed case

Distributions of which certain moments cease to exist are often called heavy-tailed. So far we reviewed theoretical results where the data matrix $\mathbf{Z}$ was "light-tailed" in the following sense: for the distributional convergence of the empirical spectral distribution and the largest eigenvalue of the sample covariance matrix towards the Marčenko–Pastur and Tracy-Widom distributions, respectively, we required finite second/fourth moments of the entries.

The behavior of the largest eigenvalue $\lambda_{(1)}$ changes dramatically when $\mathbb{E}[Z^4] = \infty$. Bai and Silverstein [9] proved for an $n \times n$ matrix $\mathbf{Z}$ with iid centered entries that

$$\limsup_{n\to\infty} \frac{\lambda_{(1)}}{n} = \infty \qquad \text{a.s.} \tag{2.6}$$

This is in stark contrast to Geman's result (2.4).

Following classical limit theory for partial sum processes and maxima, we require more than an infinite fourth moment. We assume a *regular variation condition* on the tail of Z:

$$\mathbb{P}(Z > x) \sim p_+ \frac{L(x)}{x^\alpha} \qquad \text{and} \qquad \mathbb{P}(Z < -x) \sim p_- \frac{L(x)}{x^\alpha}\,, \qquad x \to \infty\,, \tag{2.7}$$

for some $\alpha \in (0,4)$, where $p_\pm$ are non-negative constants such that $p_+ + p_- = 1$ and L is a slowly varying function. We will also refer to Z as a regularly varying random variable, $\mathbf{Z}$ as a regularly varying matrix, etc. *Here and in what follows, we normalize the eigenvalues* (λ_i) *by* (a_{np}^2) *where the sequence* (a_k) *is chosen such that*

$$\mathbb{P}(|Z| > a_k) \sim k^{-1}\,, \quad k \to \infty.$$

Standard theory for regularly varying functions (e.g. Bingham et al. [20], Feller [37]) yields that $a_n = n^{1/\alpha}\ell(n)$ where ℓ is a slowly varying function. Assuming (2.2) for p, the Potter bounds (see [20, p. 25]) yield for $\alpha \in (0,4)$ that

$$\frac{a_{np}^2}{n} \sim \frac{n^{4/\alpha}\gamma^{2/\alpha}\,\ell^2(n^2\gamma)}{n} \to \infty, \qquad n \to \infty\,, \tag{2.8}$$

i.e., the normalization a_{np}^2 is stronger than n.

The eigenvalues (λ_i) of a heavy-tailed matrix $\mathbf{ZZ}'$ were studied first by Soshnikov [65, 66]. He showed under (2.2) and (2.7) for $\alpha \in (0,2)$ that

$$\frac{\lambda_{(1)}}{a_{np}^2} \xrightarrow{\mathrm{d}} \zeta, \quad n \to \infty, \tag{2.9}$$

where ζ follows a *Fréchet distribution* with parameter $\alpha/2$:

$$\Phi_{\alpha/2}(x) = \mathrm{e}^{-x^{-\alpha/2}}\,, \qquad x > 0\,.$$

Later Auffinger et al. [4] established (2.9) also for $\alpha \in [2,4)$ under the additional assumption that the entries are centered. Both Soshnikov [65, 66] and Auffinger et al. [4] proved convergence of the point processes of normalized eigenvalues, from which one can easily infer the joint limiting distribution of the k largest eigenvalues. Davis et al. [24, 25] extended these results allowing for more general growth of p than dictated by (2.2) and a linear dependence structure between the rows and columns of $\mathbf{Z}$; see also Chakrabarty et al. [21] and the overview paper Davis et al. [23]. *The study of eigenvectors of heavy-tailed sample covariance matrices is a fresh topic, which has not been explored in the literature listed here.*

For the sake of completeness we mention that, under (2.2) with $\gamma \in (0,1]$, (2.7) with $\alpha \in (0,2)$ and $\mathbb{E}[Z] = 0$ if the latter expectation is defined, the empirical spectral distribution $F_{a_{n+p}^{-2}\mathbf{ZZ}'}$ converges weakly with probability one to a deterministic probability measure whose density ρ_α^γ satisfies

$$\rho_\alpha^\gamma(x) x^{1+\alpha/2} \to \frac{\alpha\gamma}{2(1+\gamma)}, \qquad x \to \infty\,,$$

see Belinschi et al. [15, Theorem 1.10] and Ben Arous and Guionnet [17, Theorem 1.6].

2.1.3 Structure of the paper

The primary objective of this paper is to study the joint distribution of the largest eigenvalues of the sample covariance matrix $\mathbf{ZZ}'$ in the case of iid regularly varying entries with infinite fourth moment. We make a connection between extreme value theory, point process convergence and the behavior of the largest eigenvalues. We study these eigenvalues under polynomial growth rates of the dimension p relative to the sample size n. It turns out that they are essentially determined by the extreme diagonal elements of $\mathbf{ZZ}'$ or, alternatively, by the extreme order statistics of the squared entries of $\mathbf{Z}$.

In Section 2.2 we consider power-law growth rates of (p_n), thereby generalizing proportional growth as prescribed by (2.2). Our main results are presented in Section 2.3. Theorem 2.1 provides approximations of the ordered eigenvalues of the sample covariance matrix either by the ordered diagonal elements of $\mathbf{ZZ}'$ or $\mathbf{Z}'\mathbf{Z}$, or by the order statistics of the squared entries of $\mathbf{Z}$. These approximations provide a clear picture where the largest eigenvalues of the sample covariance matrix originate from. Our results generalize those in Soshnikov [65, 66] and Auffinger et al. [4] who assume proportionality of p and n. The employed techniques originate from extreme value analysis and large deviation theory; the proofs differ from those in the aforementioned literature. The same techniques can be applied when the entries of $\mathbf{Z}$ are heavy-tailed and allow for dependence through the rows and across the columns; see Davis et al. [24, 25] for some recent attempts when the entries satisfy some linear dependence conditions. In the iid case, these results are covered by the present paper and we also show that they remain valid under much more general growth conditions than in [24, 25]. In particular, we make clear that centering of the sample covariance matrix (as assumed in [24, 25] when Z has a finite second moment) is not needed. Thus, our techniques are applicable under rather general dependence structures. We refer to the recent work by Janssen et al. [45] on eigenvalues of stochastic volatility matrix models, where non-linear dependence was allowed.

The convergence of the point processes of the properly normalized eigenvalues in Section 2.3.2 yields a multitude of useful findings connected to the joint distribution of the eigenvalues. As an application, the structure of the eigenvectors of $\mathbf{ZZ}'$ is explored in Section 2.3.3. Technical proofs are collected in Section 2.4. Section 2.5 is devoted to an extension of the results to the singular values of the *sample autocovariance matrices* which are a generalization of the traditional autocovariance function for time series to high-dimensional matrices. In applications, the analysis of sample autocovariance matrices for different lags might help to detect dependencies in the data; see Lam and Yao [49] for related work. We conclude with Appendix 2.6 which contains useful facts about regular variation and point processes.

2.2 Preliminaries

In this section we will discuss growth rates for $p = p_n \to \infty$ and introduce some notation.

2.2.1 Growth rates for p

In many applications it is not realistic to assume that the dimension p of the data and the sample size n grow at the same rate, i.e., condition (2.2) is unlikely to be satisfied. The aforementioned results of Soshnikov [65, 66] and Auffinger et al. [4] already show that the value γ in the growth rate (2.2) does not appear in the distributional limits. This observation is in contrast to the light-tailed case; see (2.3) and (2.4). Davis et al. [24, 25] allowed for more general rates for $p_n \to \infty$ than linear growth in n. *However, they could* *not completely solve the technical difficulties arising with general growth rates of p.* In what follows, we specify the growth rate of (p_n):

$$p = p_n = n^\beta \ell(n), \qquad n \geq 1, \tag{$C_p(\beta)$}$$

where ℓ is a slowly varying function and $\beta \geq 0$. If $\beta = 0$, we also assume $\ell(n) \to \infty$. Condition $C_p(\beta)$ is more general than the growth conditions in the literature; see [4, 24, 25].

2.2.2 Notation

Recall that $\mathbf{Z} = \mathbf{Z}_n = (Z_{it})_{i=1,\ldots,p;t=1,\ldots,n}$ is a $p \times n$ matrix with iid entries satisfying the regular variation condition (2.7) for some $\alpha \in (0,4)$. The sample covariance matrix $\mathbf{Z}\mathbf{Z}'$ has eigenvalues $\lambda_1, \ldots, \lambda_p$ whose order statistics were defined in (2.1).

Important roles are played by the quantities $(Z_{it}^2)_{i=1,\ldots,p;t=1,\ldots,n}$ and their order statistics

$$Z^2_{(1),np} \geq Z^2_{(2),np} \geq \cdots \geq Z^2_{(np),np}, \qquad n,p \geq 1\,. \tag{2.10}$$

As important are the row-sums

$$D_i^{\rightarrow} = D_i^{(n),\rightarrow} = \sum_{t=1}^n Z_{it}^2\,, \qquad i = 1,\ldots,p\,; \quad n = 1,2,\ldots\,, \tag{2.11}$$

with generic element $D^{\rightarrow}$ and their ordered values

$$D^{\rightarrow}_{(1)} = D^{\rightarrow}_{L_1} \geq \cdots \geq D^{\rightarrow}_{(p)} = D^{\rightarrow}_{L_p}\,, \tag{2.12}$$

where we assume without loss of generality that $(L_1, \ldots, L_p)$ is a permutation of $(1, \ldots, p)$ for fixed n.

Finally, we introduce the column-sums

$$D_t^{\downarrow} = D_t^{(n),\downarrow} = \sum_{i=1}^p Z_{it}^2\,, \qquad t = 1,\ldots,n\,; \quad p = 1,2,\ldots\,, \tag{2.13}$$

with generic element $D^{\downarrow}$ and we also adapt the notation from (2.12) to these quantities.

Norms

For any p-dimensional vector $\mathbf{v}$, $\|\mathbf{v}\|_{\ell_2}$ denotes its Euclidean norm. For any $p \times p$ matrix $\mathbf{C}$, we write $\lambda_i(\mathbf{C})$ for its p singular values and we denote their order statistics by

$$\lambda_{(1)}(\mathbf{C}) \geq \cdots \geq \lambda_{(p)}(\mathbf{C})\,.$$

For any $p \times n$ matrix $\mathbf{A} = (a_{ij})$, we will use the *spectral norm* $\|\mathbf{A}\|_2 = \sqrt{\lambda_{(1)}(\mathbf{A}\mathbf{A}')}$, the *Frobenius norm* $\|\mathbf{A}\|_F = \Big(\sum_{i=1}^p \sum_{j=1}^n |a_{ij}|^2\Big)^{1/2}$ and the *max-row sum norm* $\|\mathbf{A}\|_\infty = \max_{i=1,\ldots,p} \sum_{j=1}^n |a_{ij}|\,.$

2.3 Main results

2.3.1 Basic approximations

We commence with some basic approximation results for the eigenvalues and eigenvectors of $\mathbf{Z}\mathbf{Z}'$. The approximating quantities have a simple structure and their asymptotic behavior is inherited by the eigenvalues and has influence on the eigenvectors.

Theorem 2.1. *Consider a $p \times n$-dimensional matrix $\mathbf{Z}$ with iid entries. We assume the following conditions:*

- *The regular variation condition* (2.7) *for some $\alpha \in (0,4)$.*
- $\mathbb{E}[Z] = 0$ *for $\alpha \geq 2$.*
- *The integer sequence (p_n) has growth rate $C_p(\beta)$ for some $\beta \geq 0$.*

Then the following statements hold:

1. *If $\beta \in [0,1]$, then*

$$a_{np}^{-2} \max_{i=1,\dots,p} \big|\lambda_{(i)} - D_{(i)}^{\rightarrow}\big| \xrightarrow{\mathbb{P}} 0\,. \tag{2.14}$$

2. *If $\beta > 1$, then*

$$a_{np}^{-2} \max_{i=1,\dots,n} \big|\lambda_{(i)} - D_{(i)}^{\downarrow}\big| \xrightarrow{\mathbb{P}} 0\,. \tag{2.15}$$

3. *If $\min(\beta, \beta^{-1}) \in ((\alpha/2-1)_+, 1]$, then*

$$a_{np}^{-2} \max_{i=1,\dots,p} \big|\lambda_{(i)} - Z_{(i),np}^2\big| \xrightarrow{\mathbb{P}} 0\,. \tag{2.16}$$

Remark 2.2. In (2.15) we have chosen to take maxima over the index set $\{1,\dots,n\}$. We notice that $\lambda_{(i)} = 0$ for $i = p \wedge n + 1, \dots, p \vee n$. This is due to the fact that the $p \times p$ matrix $\mathbf{Z}\mathbf{Z}'$ and the $n \times n$ matrix $\mathbf{Z}'\mathbf{Z}$ have the same positive eigenvalues. Moreover, for n sufficiently large, $p \wedge n = p$ for $\beta \in (0,1)$ and $p \wedge n = n$ for $\beta > 1$, i.e., only in the case $\beta = 1$ both cases $n \leq p$ or $p \leq n$ are possible.

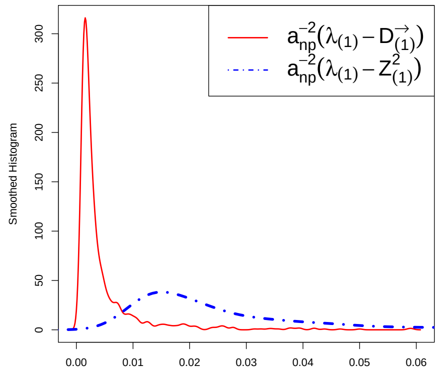
Remark 2.3. The condition $\min(\beta, \beta^{-1}) \in ((\alpha/2-1)_+, 1]$ in part (3) is only a restriction when $\alpha \in (2,4)$. We notice that this condition implies $(n \vee p)/a_{np}^2 \to 0$. In turn, this means that centering of the quantities $a_{np}^{-2} D_i^{\rightarrow}$ and $a_{np}^{-2} D_i^{\downarrow}$ in the limit theorems can be avoided. This argument is relevant in various parts of the proofs.

Remark 2.4. In Figure 2.1 we illustrate the different approximations of the eigenvalues $(\lambda_{(i)})$ by $(D_{(i)}^{\rightarrow})$ as suggested by (2.14) and $(Z_{(i),np}^2)$ as suggested by (2.16). For Z we choose the density

$$f_Z(x) = \begin{cases} \frac{\alpha}{(4|x|)^{\alpha+1}}\,, & \text{if } |x| > 1/4 \\ 1\,, & \text{otherwise.} \end{cases} \tag{2.17}$$

In the left graph, we focus on the largest eigenvalue $\lambda_{(1)}$. We show smoothed histograms of the approximation errors $a_{np}^{-2}(\lambda_{(1)} - D_{(1)}^{\rightarrow})$, $a_{np}^{-2}(\lambda_{(1)} - Z_{(1),np}^2)$. By Cauchy's interlacing theorem (see [69, Lemma 22]), the considered differences are non-negative.

In the right graph, we take the maxima as in (2.14) and (2.16) and show smoothed histograms of the approximation errors $a_{np}^{-2} \max_{i \leq p} |\lambda_{(i)} - D_{(i)}^{\rightarrow}|$, $a_{np}^{-2} \max_{i \leq p} |\lambda_{(i)} - Z_{(i),np}^2|$. We take absolute values to deal with negative differences. Figure 2.1 indicates that $(D_{(i)}^{\rightarrow})$ yield a much better approximation to $(\lambda_{(i)})$ than $(Z_{(i),np}^2)$. Notice the different scaling on the x- and y-axes.



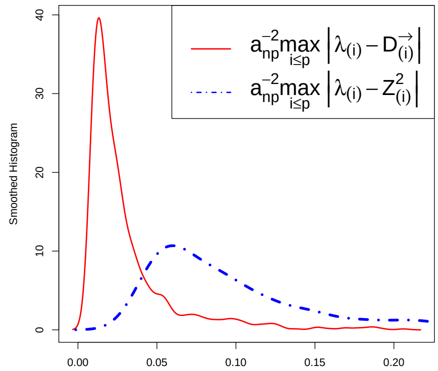


Figure 2.1: Smoothed histograms of the approximation errors for the normalized eigenvalues $(a_{np}^{-2}\lambda_{(i)})$ for entries Z_{it} with density (2.17), $\alpha = 1.6$, $\beta = 1$, $n = 1,000$ and $p = 200$.

The proof of Theorem 2.1 will be given in Section 2.4. A main step in the proof is provided by the following result whose proof will also be given in Section 2.4; a version of this theorem was proved in Davis et al. [24] under more restrictive conditions on the growth rate of (p_n).

Theorem 2.5. *Assume the conditions of Theorem 2.1 on* $\mathbf{Z}$ *and* (p_n)*.*

1. *If* $\beta \in [0,1]$ *we have*

$$a_{np}^{-2}\|\mathbf{Z}\mathbf{Z}' - \operatorname{diag}(\mathbf{Z}\mathbf{Z}')\|_2 \xrightarrow{\mathbb{P}} 0\,, \qquad n \to \infty\,.$$

2. *If* $\beta \geq 1$ *we have*

$$a_{np}^{-2}\|\mathbf{Z}'\mathbf{Z} - \operatorname{diag}(\mathbf{Z}'\mathbf{Z})\|_2 \xrightarrow{\mathbb{P}} 0\,, \qquad n \to \infty\,.$$

The second part of this theorem follows from the first one by an interchange of n and p. Indeed, if $\beta \geq 1$, we can write $n = p^{1/\beta}\ell(p)$ for some slowly varying function ℓ and then part (2) follows from part (1).

Remark 2.6. *Theorem 2.5 shows that the largest eigenvalues of* $\mathbf{Z}\mathbf{Z}'$ *are determined by the largest diagonal entries. In the case of heavy-tailed Wigner matrices, however, the diagonal elements do not play any particular role.*

From this theorem one immediately obtains a result about the approximation of the eigenvalues of $\mathbf{Z}\mathbf{Z}'$ and $\mathbf{Z}'\mathbf{Z}$ by those of $\operatorname{diag}(\mathbf{Z}\mathbf{Z}')$ and $\operatorname{diag}(\mathbf{Z}'\mathbf{Z})$, respectively. Indeed, for any symmetric $p \times p$ matrices $\mathbf{A}, \mathbf{B}$, by *Weyl's inequality* (see Bhatia [19]),

$$\max_{i=1,\ldots,p} \left|\lambda_{(i)}(\mathbf{A}+\mathbf{B}) - \lambda_{(i)}(\mathbf{A})\right| \leq \|\mathbf{B}\|_2\,. \tag{2.18}$$

If we now choose $\mathbf{A}+\mathbf{B} = \mathbf{Z}\mathbf{Z}'$ and $\mathbf{A} = \operatorname{diag}(\mathbf{Z}\mathbf{Z}')$ (or $\mathbf{A}+\mathbf{B} = \mathbf{Z}'\mathbf{Z}$ and $\mathbf{A} = \operatorname{diag}(\mathbf{Z}'\mathbf{Z})$) we obtain the following result.

Corollary 2.7. *Assume the conditions of Theorem 2.1 on* $\mathbf{Z}$ *and* (p_n).

1. *If* $\beta \in [0,1]$ *we have*

$$a_{np}^{-2} \max_{i=1,\ldots,p} \big|\lambda_{(i)} - \lambda_{(i)}(\mathrm{diag}(\mathbf{Z}\mathbf{Z}'))\big| \stackrel{\mathbb{P}}{\to} 0\,, \qquad n \to \infty\,.$$

2. *If* $\beta > 1$ *we have*

$$a_{np}^{-2} \max_{i=1,\ldots,n} \big|\lambda_{(i)} - \lambda_{(i)}(\mathrm{diag}(\mathbf{Z}'\mathbf{Z}))\big| \stackrel{\mathbb{P}}{\to} 0\,, \qquad n \to \infty\,.$$

Now (2.14) and (2.15) are immediate consequences of this corollary. Indeed, we have $\lambda_{(i)}(\mathrm{diag}(\mathbf{Z}\mathbf{Z}')) = D_{(i)}^{\to}$ and $\lambda_{(i)}(\mathbf{Z}'\mathbf{Z}) = D_{(i)}^{\downarrow}$, $i = 1,\ldots,p \wedge n$.

2.3.2 Point process convergence

In this section we want to illustrate how the approximations from Theorem 2.1 can be used to derive asymptotic theory for the largest eigenvalues of $\mathbf{Z}\mathbf{Z}'$ via the weak convergence of suitable point processes. The limiting point process involves the points of the Poisson process

$$N_\Gamma = \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}}\,, \qquad n \to \infty\,, \tag{2.19}$$

where ε_y is the Dirac measure at y,

$$\Gamma_i = E_1 + \cdots + E_i\,, \qquad i \geq 1\,,$$

and (E_i) is a sequence of iid standard exponential random variables. In other words, N_Γ is a Poisson point process on $(0,\infty)$ with mean measure $\mu(x,\infty) = x^{-\alpha/2}$, $x > 0$.

Lemma 2.8. *Assume the conditions of Theorem 2.5 hold.*

1. *If* $\beta \geq 0$*, then*

$$\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}(D_i^{\to} - c_n)} \stackrel{\mathrm{d}}{\to} N_\Gamma\,, \qquad n \to \infty\,, \tag{2.20}$$

where $c_n = 0$ *if* $\mathbb{E}[D^{\to}] = \infty$ *and* $c_n = \mathbb{E}[D^{\to}] = n\,\mathbb{E}[Z^2]$ *otherwise.*

2. *If* $\beta \geq 0$*, then*

$$\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2} Z_{(i),np}^2} \stackrel{\mathrm{d}}{\to} N_\Gamma\,, \qquad n \to \infty\,, \tag{2.21}$$

The weak convergence of the point processes holds in the space of point measures with state space $(0,\infty)$ *equipped with the vague topology; see Resnick [60].*

Remark 2.9. Similar results were used in the proofs of Davis et al. [23, 24]. We also mention that the centering c_n in the finite variance case can be avoided if $n/a_{np}^2 \to 0$. The latter condition is satisfied if $\beta > \alpha/2 - 1$.

Proof. Part (1) follows from Lemma 2.19. As regards part (2), we observe that

$$\sum_{i=1}^{p}\sum_{t=1}^{n}\varepsilon_{a_{np}^{-2}Z_{it}^2}\xrightarrow{\mathrm{d}} N_\Gamma\,; \tag{2.22}$$

see e.g. Resnick [61], Proposition 3.21. On the other hand, $a_{np}^{-2}Z_{(p),np}^2 \xrightarrow{\mathbb{P}} 0$ which together with (2.22) yields part (2). □

Theorem 2.1 and arguments similar to the proofs in Davis et al. [23, 24] enable one to derive the weak convergence of the point processes of the normalized eigenvalues.

Theorem 2.10. *Assume the conditions of Theorem 2.1. If* $\min(\beta,\beta^{-1})\in((\alpha/2-1)_+,1]$ *then*

$$\sum_{i=1}^{p}\varepsilon_{a_{np}^{-2}\lambda_i}\xrightarrow{\mathrm{d}} N_\Gamma\,, \tag{2.23}$$

in the space of point measures with state space $(0,\infty)$ *equipped with the vague topology.*

Proof. The limit relation (2.23) follows from (2.21) in combination with (2.16). Alternatively, one can exploit (2.20) both for $(D_i^{\rightarrow})$ and $(D_t^{\downarrow})$ (notice that the point process convergence for the latter sequence follows by interchanging the roles of n and p), the fact that $(n\vee p)/a_{np}^2\to 0$ if $\min(\beta,\beta^{-1})\in((\alpha/2-1)_+,1]$ (hence centering of the points $(D_i^{\rightarrow})$ and $(D_t^{\downarrow})$ in (2.20) can be avoided for $\mathbb{E}[Z^2]<\infty$) and finally using the approximations (2.14) or (2.15). □

The weak convergence of the point processes of the normalized eigenvalues of $\mathbf{ZZ}'$ in Theorem 2.10 allows one to use the conventional tools in this field; see Resnick [60, 61]. An immediate consequence is

$$a_{np}^{-2}\big(\lambda_{(1)},\ldots,\lambda_{(k)}\big)\xrightarrow{\mathrm{d}}\big(\Gamma_1^{-2/\alpha},\ldots,\Gamma_k^{-2/\alpha}\big) \tag{2.24}$$

for any fixed $k\ge 1$. Using the methods of Davis et al. [23] shows for $\alpha\in(2,4)$

$$a_{np}^{-2}\big(\lambda_{(1)}-(p\vee n)\mathbb{E}[Z^2],\ldots,\lambda_{(k)}-(p\vee n)\mathbb{E}[Z^2]\big)\xrightarrow{\mathrm{d}}\big(\Gamma_1^{-2/\alpha},\ldots,\Gamma_k^{-2/\alpha}\big)\,. \tag{2.25}$$

Equations (2.24) and (2.25) yield that for $\alpha\in(0,4)$ and any fixed $k\ge 1$,

$$a_{np}^{-2}\big(\lambda_{(1)}-\lambda_{(2)},\ldots,\lambda_{(k)}-\lambda_{(k+1)}\big)\xrightarrow{\mathrm{d}}\big(\Gamma_1^{-2/\alpha}-\Gamma_2^{-2/\alpha},\ldots,\Gamma_k^{-2/\alpha}-\Gamma_{k+1}^{-2/\alpha}\big)\,. \tag{2.26}$$

Related results can also be derived for an increasing number of order statistics, e.g. the joint convergence of the largest eigenvalue $a_{np}^{-2}\lambda_{(1)}$ and the trace $a_{np}^{-2}(\lambda_1+\cdots+\lambda_p)$. In particular, one obtains for $\alpha\in(0,2)$ under the conditions of Theorem 2.10 that

$$\frac{\lambda_{(1)}}{\lambda_1+\cdots+\lambda_p}\xrightarrow{\mathrm{d}}\frac{\Gamma_1^{-2/\alpha}}{\Gamma_1^{-2/\alpha}+\Gamma_2^{-2/\alpha}+\cdots}\,.$$

We refer to Davis et al. [24] for details on the proofs and more examples.

In the next subsection we will show how the above results on the joint convergence of eigenvalues can be applied to approximate the eigenvectors of $\mathbf{ZZ}'$.

2.3.3 Eigenvectors

In this section we assume the conditions of Theorem 2.5 and $\beta \in [0,1]$. From Theorem 2.5(1) we know that $\mathbf{Z}\mathbf{Z}'$ is approximated in spectral norm by $\mathrm{diag}(\mathbf{Z}\mathbf{Z}')$. The unit eigenvectors of a $p \times p$ diagonal matrix are the canonical basis vectors $\mathbf{e}_j \in \mathbb{R}^p$, $j = 1, \dots, p$. This raises the question as to whether $(\mathbf{e}_j)$ are good approximations of the eigenvectors $(\mathbf{v}_j)$ of $\mathbf{Z}\mathbf{Z}'$. By $\mathbf{v}_j$ we denote the unit eigenvector associated with the jth largest eigenvalue $\lambda_{(j)}$. The unit eigenvector associated with the jth largest eigenvalue of $\mathrm{diag}(\mathbf{Z}\mathbf{Z}')$ is $\mathbf{e}_{L_j}$, where L_j is defined in (2.12). Our guess that $\mathbf{v}_j$ is approximated by $\mathbf{e}_{L_j}$ is confirmed by the following result.

Theorem 2.11. *Assume the conditions of Theorem 2.1 and let $\beta \in [0,1]$. Then for any fixed $k \ge 1$,*

$$\|\mathbf{v}_k - \mathbf{e}_{L_k}\|_{\ell_2} \stackrel{\mathbb{P}}{\to} 0\,, \quad n \to \infty\,.$$

Indeed, $\mathbf{v}_j$ and $\mathbf{e}_{L_j}$ share another property: they are *localized* which means that they are concentrated only in a few components. Vectors which are not localized are called *delocalized*. Figure 2.2 shows the outcome of a simulation example in which we visualize the components of the unit eigenvector associated with the largest eigenvalue of $\mathbf{Z}\mathbf{Z}'$ for a simulated data matrix $\mathbf{Z}$ with iid Pareto(0.8) entries. In the right graph we see that only one of the $p = 200$ components is significant. Hence we can find a canonical basis vector $\mathbf{e}_k$ such that $\|\mathbf{e}_k - \mathbf{v}_1\|_{\ell_2}$ is small. Therefore the eigenvector is localized. This is in stark contrast to the case of iid standard normal entries; see the left graph. Then many components are of similar magnitude, hence the eigenvector is delocalized. Typically, the eigenvectors tend to be localized when the entry distribution has an infinite fourth moment, while they tend to be delocalized otherwise; see Benaych-Georges and Péché [18] for the case of Wigner matrices.

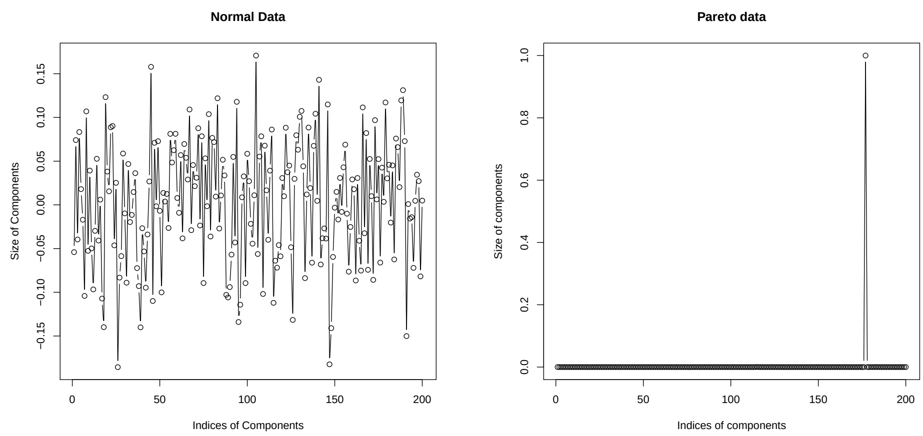


Figure 2.2: The components of the eigenvector $\mathbf{v}_1$. Right: The case of iid Pareto(0.8) entries. Left: The case of iid standard normal entries. We choose $p = 200$ and $n = 1,000$.

Proof of Theorem 2.11. Fix $k \ge 1$. Since $p \to \infty$ we can assume $k \le p$ for sufficiently large n. We observe that for $j=1,\dots,p$,

$$\mathbf{Z}\mathbf{Z}'\,\mathbf{e}_j - D_j^{\rightarrow}\,\mathbf{e}_j = \Big(\sum_{t=1}^n Z_{1t}Z_{jt},\dots,\sum_{t=1}^n Z_{j-1,t}Z_{jt},0,\sum_{t=1}^n Z_{j+1,t}Z_{jt},\dots,\sum_{t=1}^n Z_{pt}Z_{jt}\Big)',$$

are the columns of $\mathbf{Z}\mathbf{Z}' - \operatorname{diag}(\mathbf{Z}\mathbf{Z}')$. By Theorem 2.5(1),

$$a_{np}^{-2}\max_{j=1,\dots,p}\|\mathbf{Z}\mathbf{Z}'\mathbf{e}_j - D_j^{\rightarrow}\mathbf{e}_j\|_{\ell_2} \le a_{np}^{-2}\|\mathbf{Z}\mathbf{Z}' - \operatorname{diag}(\mathbf{Z}\mathbf{Z}')\|_2 \stackrel{\mathbb{P}}{\to} 0\,,\; n\to\infty. \tag{2.27}$$

If we set $\mathbf{H}^{(n)} = a_{np}^{-2}\mathbf{Z}\mathbf{Z}'$, $\mathbf{v}^{(n)} = \mathbf{e}_{L_k}\in\mathbb{R}^p$ and $\lambda^{(n)} = a_{np}^{-2}D_{L_k}^{\rightarrow}$, we see that

$$a_{np}^{-2}\mathbf{Z}\mathbf{Z}'\mathbf{e}_{L_k} = a_{np}^{-2}D_{L_k}^{\rightarrow}\mathbf{e}_{L_k} + \varepsilon^{(n)}\mathbf{w}^{(n)},$$

where $\mathbf{w}^{(n)} = \|\mathbf{Z}\mathbf{Z}'\mathbf{e}_{L_k} - D_{L_k}^{\rightarrow}\mathbf{e}_{L_k}\|_{\ell_2}^{-1}(\mathbf{Z}\mathbf{Z}'\mathbf{e}_{L_k} - D_{L_k}^{\rightarrow}\mathbf{e}_{L_k})$ is a unit vector and $\varepsilon^{(n)} = a_{np}^{-2}\|\mathbf{Z}\mathbf{Z}'\mathbf{e}_{L_k} - D_{L_k}^{\rightarrow}\mathbf{e}_{L_k}\|_{\ell_2} \stackrel{\mathbb{P}}{\to} 0$ by (2.27).

Before we can apply Proposition 2.23 we need to show that with probability converging to 1, there are no other eigenvalues in a suitably small interval around $\lambda_{(k)}$. Let $s>1$. We define the set

$$\Omega_n = \Omega_n(k,s) = \{a_{np}^{-2}|\lambda_{(k)} - \lambda_{(i)}| > s\,\varepsilon^{(n)} \,:\, i\neq k = 1,\dots,p\}\,.$$

From (2.27) we get $s\,\varepsilon^{(n)}\to 0$. Then using this and (2.26), we obtain

$$\lim_{n\to\infty}\mathbb{P}(\Omega_n^c) = \lim_{n\to\infty}\mathbb{P}(a_{np}^{-2}\min\{\lambda_{(k-1)} - \lambda_{(k)},\lambda_{(k)} - \lambda_{(k+1)}\}\le s\,\varepsilon^{(n)}) = 0$$

By Proposition 2.23 the unit eigenvector $\mathbf{v}_k$ associated with $\lambda_{(k)}$ and the projected vector $\mathbf{P}_{\mathbf{e}_{L_k}}(\mathbf{v}_k) = (\mathbf{v}_k)_{L_k}\mathbf{e}_{L_k}$ satisfy for fixed $\delta>0$:

$$\begin{aligned}
&\limsup_{n\to\infty}\mathbb{P}(\|\mathbf{v}_k - (\mathbf{v}_k)_{L_k}\mathbf{e}_{L_k}\|_{\ell_2} > \delta)\\
&\quad\le \limsup_{n\to\infty}\mathbb{P}(\{\|\mathbf{v}_k - (\mathbf{v}_k)_{L_k}\mathbf{e}_{L_k}\|_{\ell_2} > \delta\}\cap\Omega_n) + \limsup_{n\to\infty}\mathbb{P}(\Omega_n^c)\\
&\quad\le \limsup_{n\to\infty}\mathbb{P}(\{2\varepsilon^{(n)}/(s\,\varepsilon^{(n)} - \varepsilon^{(n)}) > \delta\}\cap\Omega_n)\\
&\quad\le \limsup_{n\to\infty}\mathbb{P}(\{2/(s-1) > \delta\}) = \mathbf{1}_{\{2/(s-1)>\delta\}}.
\end{aligned}$$

The right-hand side is zero for sufficiently large s. Since both $\mathbf{v}_k$ and $\mathbf{e}_{L_k}$ are unit vectors this means that

$$\|\mathbf{v}_k - \mathbf{e}_{L_k}\|_{\ell_2} \stackrel{\mathbb{P}}{\to} 0\,,\quad n\to\infty\,.$$

This proves our result on eigenvectors. □

2.4 Proof of Theorem 2.1

In what follows, c stands for any constant whose value is not of interest. We write (Z_t) for an iid sequence with the same distribution as Z.

The plan of the proof is as follows:

1. We prove Theorem 2.5 which implies (2.14) and (2.15); see Corollary 2.7. In view of the arguments after Theorem 2.1 it suffices to consider only the case $\beta\in[0,1]$.
2. We prove (2.16).

2.4.1 Proof of Theorem 2.5

We proceed in several steps.

The case $\alpha \in (0, 8/3)$. If $\alpha \in [1,2)$ and $\mathbb{E}[|Z|] < \infty$, we have

$$a_{np}^{-1}\|\mathbf{Z} - (\mathbf{Z} - \mathbb{E}[\mathbf{Z}])\|_2 = |\mathbb{E}[Z]| \frac{\sqrt{np}}{a_{np}} \to 0\,, \quad n \to \infty.$$

Therefore, without loss of generality $\mathbb{E}[Z]$ can be assumed 0 in this case.

From now on we assume $\mathbb{E}[Z] = 0$ whenever $\mathbb{E}[|Z|]$ exists. Since the Frobenius norm $\|\cdot\|_F$ is an upper bound of the spectral norm we have

$$\begin{aligned}
\|\mathbf{ZZ}' - \operatorname{diag}(\mathbf{ZZ}')\|_2^2 &\le \|\mathbf{ZZ}' - \operatorname{diag}(\mathbf{ZZ}')\|_F^2 \\
&= \sum_{i,j=1;i\neq j}^{p} \sum_{t=1}^{n} Z_{it}^2 Z_{jt}^2 + \sum_{i,j=1;i\neq j}^{p} \sum_{t_1,t_2=1;t_1\neq t_2}^{n} Z_{i,t_1} Z_{j,t_1} Z_{i,t_2} Z_{j,t_2} \\
&= \sum_{i,j=1;i\neq j}^{p} \sum_{t=1}^{n} Z_{it}^2 Z_{jt}^2 \big[\mathbf{1}_{\{Z_{it}^2 Z_{jt}^2 > a_{np}^4\}} + \mathbf{1}_{\{Z_{it}^2 Z_{jt}^2 \le a_{np}^4\}}\big] + I_2^{(n)} \\
&= I_{11}^{(n)} + I_{12}^{(n)} + I_2^{(n)}\,.
\end{aligned}$$

Thus it suffices to show that each of the expressions on the right-hand side when normalized with a_{np}^4 converges to zero in probability. We have for any $\epsilon > 0$,

$$\mathbb{P}\big(I_{11}^{(n)} > \epsilon\, a_{np}^4\big) \le p^2\, n\, \mathbb{P}(Z_1^2 Z_2^2 > a_{np}^4) \to 0\,.$$

Here we also used the fact that $Z_1 Z_2$ is regularly varying with index α; see Embrechts and Goldie [34]. An application of Markov's inequality and Lyapunov's moment inequality with $\gamma \in (\alpha/2, 4/3)$ if $\alpha \in [2, 8/3)$ and $\gamma = 1$ otherwise shows that

$$\mathbb{P}\big(I_{12}^{(n)} > \epsilon\, a_{np}^4\big) \le c\, \frac{p^2 n}{a_{np}^4} \Big(\mathbb{E}[|Z_1 Z_2|^{2\gamma} \mathbf{1}_{\{|Z_1 Z_2| \le a_{np}^2\}}]\Big)^{\frac{1}{\gamma}} \le c\, p^{2-\frac{2}{\gamma}}\, n^{1-\frac{2}{\gamma}+\delta} \to 0,$$

where we used Karamata's theorem (see Bingham et al. [20]), and the constant $\delta > 0$ can be chosen arbitrarily small due to the Potter bounds.

In the case $\alpha \in (0,2)$ the probability $P_2^{(n)} = \mathbb{P}(I_2^{(n)} > \epsilon\, a_{np}^4)$ can be handled analogously. Next, we turn to $P_2^{(n)}$ in the case $\alpha \in (2, 8/3)$. In particular, $\mathbb{E}[Z^2] < \infty$. With Čebychev's inequality, also using the fact that $\mathbb{E}[Z] = 0$, we find that

$$P_2^{(n)} \le c\frac{1}{a_{np}^8} \mathbb{E}\Big[\Big(\sum_{i,j=1;i\neq j}^{p} \sum_{t_1,t_2=1;t_1\neq t_2}^{n} Z_{i,t_1} Z_{j,t_1} Z_{i,t_2} Z_{j,t_2}\Big)^2\Big] \le c\, \frac{(p\, n)^2}{a_{np}^8} \to 0. \tag{2.28}$$

The case $\alpha = 2$ is most difficult because the second moment of Z can be infinite. Without loss of generality we assume that Z is continuous. Otherwise, we add independent centered normal random variables to each of the entries Z_{it}; due the normalization a_{np}^2 the asymptotic properties of the eigenvalues remain the same, i.e., the added normal components are asymptotically negligible. In view of Hult and Samorodnitsky [44, Lemma 4.2] there exist constants $C, K > 0$ and a function $h : [K, \infty) \to (0, \infty)$ such that

$$\mathbb{E}[Z \mathbf{1}_{\{-h(x) \le Z \le x\}}] = 0 \quad \text{and} \quad C^{-1} \le \frac{h(x)}{x} \le C \tag{2.29}$$

for all $x \ge K$.[1] We have

$$I_2^{(n)} = \sum_{i,j=1;i\neq j}^{p} \sum_{t_1,t_2=1;t_1\neq t_2}^{n} Z_{i,t_1} Z_{j,t_1} Z_{i,t_2} Z_{j,t_2} \left[\mathbf{1}_{A^c_{i,j,t_1,t_2}} + \mathbf{1}_{A_{i,j,t_1,t_2}}\right] = I_{21}^{(n)} + I_{22}^{(n)} ,$$

where $A_{i,j,t_1,t_2} = \{-h(a_{np}^4) \le Z_{i,t_1}, Z_{j,t_1}, Z_{i,t_2}, Z_{j,t_2} \le a_{np}^4\}$. We see that

$$\begin{aligned} \mathbb{P}(I_{21}^{(n)} > \epsilon\, a_{np}^4) &\le (p\,n)^2\, \mathbb{P}(A^c_{i,j,t_1,t_2}) \le c\,(p\,n)^2\, \mathbb{P}(|Z| > \min(h(a_{np}^4), a_{np}^4)) \\ &\le c\,(pn)^2\, \mathbb{P}(|Z| > \min(C, C^{-1})\, a_{np}^4) \le c\,(np)^{-2+\delta} \to 0, \end{aligned}$$

where we used the second formula in (2.29). The small constant $\delta > 0$ comes from a Potter bound argument. Finally, using the first condition in (2.29), we may conclude similarly to (2.28) that

$$P_{22}^{(n)} = \mathbb{P}(I_{22}^{(n)} > \epsilon\, a_{np}^4) \le c\, \frac{(pn)^2}{a_{np}^8} \left(\mathbb{E}[Z^2 \mathbf{1}_{\{-h(a_{np}^4)\le Z\le a_{np}^4\}}]\right)^4.$$

Since

$$\mathbb{E}[Z^2 \mathbf{1}_{\{|Z|\le \max(C,C^{-1})x\}}] \ge \mathbb{E}[Z^2 \mathbf{1}_{\{-h(x)\le Z\le x\}}] ,$$

and the left-hand side is slowly varying (see [37]), we have $P_{22}^{(n)} \to 0$. The proof is complete for $\alpha \in (0, 8/3)$. □

The case $\alpha \in [8/3, 4)$

Before we can proceed with the case $\alpha \in [8/3, 4)$ we provide an auxiliary result. Consider the following decomposition

$$[\mathbf{Z}\mathbf{Z}' - \operatorname{diag}(\mathbf{Z}\mathbf{Z}')]^2 = \mathbf{D} + \mathbf{F} + \mathbf{R} ,$$

where

$$\mathbf{D} = (D_{ij})_{i,j=1,\dots,p} = \operatorname{diag}([\mathbf{Z}\mathbf{Z}' - \operatorname{diag}(\mathbf{Z}\mathbf{Z}')]^2) ,$$

The $p \times p$ matrix $\mathbf{F}$ has a zero-diagonal and

$$F_{ij} = \sum_{u=1;u\neq i,j}^{p} \sum_{t=1}^{n} Z_{it}\, Z_{jt}\, Z_{ut}^2, \quad 1 \le i \ne j \le p ,$$

The $p \times p$ matrix $\mathbf{R}$ has a zero-diagonal and

$$R_{ij} = \sum_{u=1;u\neq i,j}^{p} \sum_{t_1=1}^{n} \sum_{t_2=1;t_2\neq t_1}^{n} Z_{i,t_1}\, Z_{j,t_2}\, Z_{u,t_1}\, Z_{u,t_2}, \quad 1 \le i \ne j \le p .$$

Lemma 2.12. *Assume the conditions of Theorem 2.5 and $\alpha \in (2, 4)$. Then $a_{np}^{-4}\big(\|\mathbf{D}\|_2 + \|\mathbf{F}\|_2 + \|\mathbf{R}\|_2\big) \xrightarrow{\mathbb{P}} 0$.*

[1] Here we assume that $p_+\, p_- > 0$. If either $p_+ = 0$ or $p_- = 0$ one can proceed in a similar way by modifying h slightly; we omit details.

In view of this lemma we have

$$a_{np}^{-4}\|\mathbf{Z}\mathbf{Z}' - \operatorname{diag}(\mathbf{Z}\mathbf{Z}')\|_2^2 = a_{np}^{-4}\|[\mathbf{Z}\mathbf{Z}' - \operatorname{diag}(\mathbf{Z}\mathbf{Z}')]^2\|_2 = a_{np}^{-4}\|\mathbf{D}+\mathbf{F}+\mathbf{R}\|_2^2 \stackrel{\mathbb{P}}{\to} 0\,.$$

This finishes the proof of Theorem 2.5. It is left to prove Lemma 2.12.

Proof of the $\mathbf{D}$*-part.* We have for $i=1,\dots,p$,

$$\begin{aligned} D_{ii} &= \sum_{u=1}^{p}\sum_{t=1}^{n} Z_{it}^2 Z_{ut}^2 \mathbf{1}_{\{i\neq u\}} + \sum_{u=1}^{p}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n} Z_{i,t_1}Z_{u,t_1}Z_{u,t_2}Z_{i,t_2}\mathbf{1}_{\{i\neq u\}}\mathbf{1}_{\{t_1\neq t_2\}} \\ &= M_{ii} + N_{ii}\,. \end{aligned}$$

We write $\mathbf{M}$ and $\mathbf{N}$ for diagonal matrices constructed from (M_{ii}) and (N_{ii}) such that $\mathbf{D}=\mathbf{M}+\mathbf{N}$. First bounding $\|\mathbf{N}\|_2$ by the Frobenius norm and then applying Markov's inequality, one can prove that $a_{np}^{-4}\|\mathbf{N}\|_2 \stackrel{\mathbb{P}}{\to} 0$. We have

$$\frac{\mathbb{E}[M_{ii}]}{a_{np}^4} \le c\,\frac{np}{a_{np}^4} \to 0, \qquad n\to\infty.$$

Therefore centering of M_{ii} will not influence the limit of the spectral norm $a_{np}^{-4}\|\mathbf{M}\|_2$. Writing $A_{i,u} = \{|\sum_{t=1}^{n}(Z_{it}^2Z_{ut}^2 - \mathbb{E}[Z_1^2Z_2^2])\mathbf{1}_{\{i\neq u\}}| > a_{np}^2\}$, we have for $i=1,\dots,p$,

$$\begin{aligned} M_{ii} - \mathbb{E}[M_{ii}] &= \sum_{u=1}^{p}\sum_{t=1}^{n}\left(Z_{it}^2Z_{ut}^2 - \mathbb{E}[Z_1^2Z_2^2]\right)\mathbf{1}_{\{i\neq u\}}\left[\mathbf{1}_{A_{i,u}} + \mathbf{1}_{A_{i,u}^c}\right] \\ &= M_{ii}^{(1)} + M_{ii}^{(2)}. \end{aligned}$$

On the one hand, $\|M^{(2)}\|_2 \le p\,a_{np}^2$. Hence $a_{np}^{-4}\|M^{(2)}\|_2 \stackrel{\mathbb{P}}{\to} 0$. On the other hand, we obtain with Markov's inequality, Proposition 2.18 and the Potter bounds for $\epsilon>0$ and small $\delta>0$,

$$\begin{aligned} \mathbb{P}(\|M^{(1)}\|_2 > \epsilon\, a_{np}^4) &= \mathbb{P}(\max_{i=1,\dots,p}|M_{ii}^{(1)}| > \epsilon\, a_{np}^4) \\ &\le \mathbb{P}\Big(\max_{i=1,\dots,p}\sum_{u=1}^{p}\Big|\sum_{t=1}^{n}(Z_{it}^2Z_{ut}^2 - \mathbb{E}[Z_1^2Z_2^2])\mathbf{1}_{\{i\neq u\}}\,\mathbf{1}_{A_{i,u}}\Big| > \epsilon\, a_{np}^4\Big) \\ &\le c\frac{p^2}{a_{np}^4}\mathbb{E}\Big[\Big|\sum_{t=1}^{n}(Z_{1t}^2Z_{2t}^2 - \mathbb{E}[Z_1^2Z_2^2])\Big|\mathbf{1}_{A_{1,2}}\Big] \\ &\sim c\,\frac{p^2}{a_{np}^4}\,n\,a_{np}^2\,\mathbb{P}(Z_1^2Z_2^2 > a_{np}^2) \le \frac{p\,(np)^\delta}{a_{np}^2} \to 0, \end{aligned}$$

since Z_1Z_2 is regularly varying with index α. This finishes the proof of the $\mathbf{D}$-part. □

Proof of the $\mathbf{F}$*-part.* Let $\delta>0$. We will use the following decomposition for $i\neq j$:

$$\begin{aligned} F_{ij} = &\sum_{u=1;u\neq i,j}^{p}\sum_{t=1}^{n} Z_{it}Z_{jt}(Z_{ut}^2 - \mathbb{E}[Z^2\mathbf{1}_{\{Z^2\le a_{np}^{4-2\delta}\}}])+ \\ &+\mathbb{E}[Z^2\mathbf{1}_{\{Z^2\le a_{np}^{4-2\delta}\}}]\,(p-2)\sum_{t=1}^{n} Z_{it}Z_{jt} = \widetilde{F}_{ij} + T_{ij}\,. \end{aligned}$$

We observe that $\mathbf{T} = \mathbb{E}[Z^2 \mathbf{1}_{\{Z^2 \le a_{np}^{4-2\delta}\}}]\,(p-2)\,(\mathbf{Z}_n\mathbf{Z}_n' - \operatorname{diag}(\mathbf{Z}_n\mathbf{Z}_n'))$. We have for some constant $c>0$,

$$\begin{aligned}\|\mathbf{T}\|_2^2 &= \|\mathbf{T}^2\|_2 \le c\,p^2\,\|(\mathbf{Z}_n\mathbf{Z}_n' - \operatorname{diag}(\mathbf{Z}_n\mathbf{Z}_n'))^2\|_2\\ &\le c\,p^2\,\|\mathbf{D}+\widetilde{\mathbf{F}}+\mathbf{R}\|_2 + c\,p^2\,\|\mathbf{T}\|_2\,.\end{aligned}$$

Therefore

$$\frac{\|\mathbf{T}\|_2}{a_{np}^4} \le c\,\frac{p}{a_{np}^2}\Big(\frac{\|\mathbf{D}+\widetilde{\mathbf{F}}+\mathbf{R}\|_2}{a_{np}^4}\Big)^{1/2} + c\,\frac{p}{a_{np}^2}\Big(\frac{\|\mathbf{T}\|_2}{a_{np}^4}\Big)^{1/2}. \tag{2.30}$$

In the course of the proof of this lemma we show that

$$\frac{\|\mathbf{D}+\widetilde{\mathbf{F}}+\mathbf{R}\|_2}{a_{np}^4} \stackrel{\mathbb{P}}{\to} 0\,.$$

Moreover, there is a small $\varepsilon>0$ such that

$$\delta_n = \frac{p}{a_{np}^2} \le n^{1-4/\alpha+\varepsilon}, \quad 1-4/\alpha+\varepsilon<0\,.$$

Therefore iteration of (2.30) yields for $k\ge 1$

$$\begin{aligned}\frac{\|\mathbf{T}\|_2}{a_{np}^4} &\le o_{\mathbb{P}}(1) + c\,\delta_n\Big(\delta_n\Big(\frac{\|\mathbf{D}+\widetilde{\mathbf{F}}+\mathbf{R}\|_2}{a_{np}^4}\Big)^{1/2}\Big)^{1/2} + c\,\delta_n\Big(\delta_n\Big(\frac{\|\mathbf{T}\|_2}{a_{np}^4}\Big)^{1/2}\Big)^{1/2}\\ &= o_{\mathbb{P}}(1) + c\,\Big(\delta_n^{4+2}\frac{\|\mathbf{T}\|_2}{a_{np}^4}\Big)^{1/4}\\ &\le o_{\mathbb{P}}(1) + c\,\Big(\delta_n^{2^k+\cdots+2}\frac{\|\mathbf{T}\|_2}{a_{np}^4}\Big)^{1/2^k}.\end{aligned} \tag{2.31}$$

Using some elementary moment bounds for $\|\mathbf{T}\|_2$ (e.g. a bound by the Frobenius norm), it is not difficult to show that $n^{-l}\|\mathbf{T}\|_2 \stackrel{\mathbb{P}}{\to} 0$ for some sufficiently large l. Thus we achieve that the right-hand side in (2.31) converges to zero in probability.

It remains to show that $a_{np}^{-4}\|\widetilde{\mathbf{F}}\|_2 \stackrel{\mathbb{P}}{\to} 0$. With the notation $B_{u,t} = \{Z_{ut}^2 \le a_{np}^{4-2\delta}\}$ for some small $\delta>0$, we decompose $Z_{it}Z_{jt}(Z_{ut}^2 - \mathbb{E}[Z^2\mathbf{1}_{\{Z^2\le a_{np}^{4-2\delta}\}}])$ as follows:

$$Z_{it}Z_{jt}(Z_{ut}^2\mathbf{1}_{B_{u,t}} - \mathbb{E}[Z^2\mathbf{1}_{\{Z^2\le a_{np}^{4-2\delta}\}}]) + Z_{it}Z_{jt}Z_{ut}^2\mathbf{1}_{B_{u,t}^c}\,.$$

We decompose the matrix $\widetilde{\mathbf{F}}$ accordingly:

$$\widetilde{\mathbf{F}} = \widetilde{\mathbf{F}}^{(1)} + \widetilde{\mathbf{F}}^{(2)}\,,$$

such that, for example,

$$\widetilde{F}_{ij}^{(1)} = \sum_{u=1;u\ne i,j}^{p}\sum_{t=1}^{n} Z_{it}Z_{jt}(Z_{ut}^2\mathbf{1}_{B_{u,t}} - \mathbb{E}[Z^2\mathbf{1}_{\{Z^2\le a_{np}^{4-2\delta}\}}])\,, \quad i\ne j.$$

$\widetilde{\mathbf{F}}^{(1)}$: Bounding the spectral norm by the Frobenius norm, applying Markov's inequality and using Karamata's theorem together with the Potter bounds one can check that for

$\epsilon > 0$ and small $\delta > 0$,

$$
\begin{aligned}
\mathbb{P}(\|\widetilde{\mathbf{F}}^{(1)}\|_2 > \epsilon\, a_{np}^4) &\le c\, a_{np}^{-8}\, \mathbb{E}\Big[\sum_{i,j=1}^{p} (\widetilde{F}_{ij}^{(1)})^2\Big] \\
&\le c\, \frac{p^3\, n}{a_{np}^8} \mathbb{E}[(Z_1 Z_2)^2] \mathbb{E}[(Z^2 \mathbf{1}_{\{Z^2 \le a_{np}^{4-2\delta}\}} - \mathbb{E}[Z^2 \mathbf{1}_{\{Z^2 \le a_{np}^{4-2\delta}\}}])^2] \\
&\le c\, \frac{p^3\, n}{a_{np}^8} \mathbb{E}[Z^4 \mathbf{1}_{\{Z^2 \le a_{np}^{4-2\delta}\}}] \\
&\le c \frac{p^3 n}{a_{np}^{4\delta}} \mathbb{P}(|Z| > a_{np}^{2-\delta}) \to 0\,, \qquad n \to \infty,
\end{aligned}
$$

$\widetilde{\mathbf{F}}^{(2)}$: We have for small $\delta > 0$,

$$
\begin{aligned}
\mathbb{P}(\|\widetilde{\mathbf{F}}^{(2)}\|_2 > \epsilon\, a_{np}^4) \quad &\le \quad \mathbb{P}\Big(\bigcup_{1 \le u \le p\,, 1 \le t \le n} B_{u,t}^c \Big) \\
&\le \quad p\, n\, \mathbb{P}(|Z| > a_{np}^{2-\delta}) \to 0\,, \qquad n \to \infty\,.
\end{aligned}
$$

The proof of the **F**-part is complete. □

Proof of the **R***-part.* We have

$$
\mathbb{E}[\|\mathbf{R}\|_2^2] \le \mathbb{E}[\|\mathbf{R}\|_F^2] \quad \le \quad \sum_{i,j=1}^{p} \sum_{u=1}^{p} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} (\mathbb{E}[Z^2])^4 \le c\, p^3\, n^2\,.
$$

Therefore and by Markov's inequality for $\epsilon > 0$,

$$
\mathbb{P}(\|\mathbf{R}\|_2 > \epsilon\, a_{np}^4) \le c\, \frac{p^3\, n^2}{a_{np}^8} \to 0, \qquad n \to \infty\,, \tag{2.32}
$$

as long as $\alpha \in (2, 16/5)$. For $\alpha \in [16/5, 4)$ we use a similar idea for the truncated entries. Write $\mathbf{R} = \overline{\mathbf{R}} + \widetilde{\mathbf{R}}$, where for $i \ne j$

$$
\begin{aligned}
\overline{R}_{ij} &= \sum_{u=1; u \ne i,j}^{p} \sum_{t_1=1}^{n} \sum_{t_1=1; t_1 \ne t_2}^{n} Z_{i,t_1} Z_{j,t_2} Z_{u,t_1} Z_{u,t_2}\, \mathbf{1}_{A_{i,j,t_1,t_2}}\,, \\
\widetilde{R}_{ij} &= \sum_{u=1; u \ne i,j}^{p} \sum_{t_1=1}^{n} \sum_{t_1=1; t_1 \ne t_2}^{n} Z_{i,t_1} Z_{j,t_2} Z_{u,t_1} Z_{u,t_2}\, \mathbf{1}_{A_{i,j,t_1,t_2}^c}\,,
\end{aligned}
$$

with $A_{i,j,t_1,t_2}^c = \{-h(a_{np}) \le Z_{i,t_1}, Z_{j,t_2}, Z_{u,t_1}, Z_{u,t_2} \le a_{np}\}$ and h as in (2.29). Analogously to (2.32), using the fact that $Z_{i,t_1} Z_{j,t_2} Z_{u,t_1} Z_{u,t_2}\, \mathbf{1}_{A_{i,j,t_1,t_2}}$ are uncorrelated for the considered index set, one obtains for $\epsilon > 0$,

$$
\begin{aligned}
\mathbb{P}(\|\overline{\mathbf{R}}\|_2 > \epsilon\, a_{np}^4) \quad &\le \quad c\, \frac{p^3\, n^2}{a_{np}^8}\, \mathbb{E}[Z^2 \mathbf{1}_{\{|Z| > \min(C, C^{-1})\, a_{np}\}}] \\
&\le \quad c\, \frac{p^3\, n^2}{a_{np}^6}\, \mathbb{P}(|Z| > \min(C, C^{-1})\, a_{np}) \to 0\,, \quad n \to \infty\,,
\end{aligned}
$$

where we used Karamata's theorem and $\mathbb{P}(A_{i,j,t_1,t_2}) \le c\, \mathbb{P}(|Z| > \min(C, C^{-1})\, a_{np})$.

We introduce the truncated random variables $\widetilde{Z}_{it} = Z_{it}\mathbf{1}_{\{-h(a_{np})\le Z_{it}\le a_{np}\}}$ with the generic element $\widetilde{Z}$. We will repeatedly use the following inequality which is valid for a real symmetric matrix $\mathbf{M}$:

$$\|\mathbf{M}\|_2^2 \le \|\mathbf{M}\|_F^2 = \operatorname{tr}(\mathbf{M}^2).$$

Then we have for $k\ge 1$, $\|\widetilde{\mathbf{R}}^{2^{k-1}}\|_2^2 = \|\widetilde{\mathbf{R}}^{2^k}\|_2$ and

$$\|\widetilde{\mathbf{R}}\|_2^{2^k} \le \operatorname{tr}(\widetilde{\mathbf{R}}^{2^k}) = \sum_{i,j=1}^p (\widetilde{R}^{2^{k-1}})_{ij}^2\,.$$

This together with the Markov inequality of order 2^k yields

$$\mathbb{P}(\|\widetilde{\mathbf{R}}\|_2 > ca_{np}^4) \le ca_{np}^{-4\cdot 2^k}\,\mathbb{E}\Big[\sum_{i,j=1}^p (\widetilde{R}^{2^{k-1}})_{ij}^2\Big]. \tag{2.33}$$

Next we study the structure of $\widetilde{\mathbf{R}}^{2^{k-1}}$. The (i,j)-entry of this matrix is

$$(\widetilde{R}^{2^{k-1}})_{ij} = \sum_{i_1=1}^p \cdots \sum_{i_{2^{k-1}-1}=1}^p \widetilde{R}_{i,i_1}\widetilde{R}_{i_1,i_2}\cdots \widetilde{R}_{i_{2^{k-1}-2},i_{2^{k-1}-1}}\widetilde{R}_{i_{2^{k-1}-1},j}. \tag{2.34}$$

In view of (2.34) and by definition of $\widetilde{\mathbf{R}}$, $(\widetilde{R}^{2^{k-1}})_{ij}$ contains exactly 2^k-1 sums running from 1 to p, and 2^k sums running from 1 to n. Now we consider the expectation on the right-hand side of (2.33). The highest and lowest powers of $\widetilde{Z}_{it}$ in this expectation are 2^k and 1. Let $(I,T) = ((i_1,t_1),\dots,(i_{2^k},t_{2^k}))$. We have

$$\mathbb{E}\Big[\sum_{i,j=1}^p (\widetilde{R}^{2^{k-1}})_{ij}^2\Big] = \sum_{(I,T)\in S} \mathbb{E}[\widetilde{Z}_{i_1,t_1}\widetilde{Z}_{i_2,t_2}\cdots\widetilde{Z}_{i_{2^k},t_{2^k}}],$$

where $S\subset \{1,\dots,p\}^{2^k}\times\{1,\dots,n\}^{2^k}$ is the index set that covers all combinations of indices that arise on the left-hand side. Since $\mathbb{E}[\widetilde{Z}]=0$, each $\widetilde{Z}$ in $\widetilde{Z}_{i_1,t_1}\widetilde{Z}_{i_2,t_2}\cdots\widetilde{Z}_{i_{2^k},t_{2^k}}$ must appear at least twice for the expectation of this product to be non-zero. Let $S_1\subset S$ be the set of all those indices that make a non-zero contribution to the sum. From the specific structure of $\widetilde{\mathbf{R}}$, (2.34) and the considerations above it now follows that the cardinality of S_1 has the following bound

$$|S_1| \le c(k)\,p^2\,p^{2^k-1}\,n^{2^k} = c\,p^{2^k+1}\,n^{2^k}\,.$$

For $l=2,3$ we can use $\mathbb{E}[|\widetilde{Z}^l|]\le c$. If $l\ge 4$, we infer with Karamata's theorem

$$\mathbb{E}[|\widetilde{Z}^l|] \le c\,a_{np}^l\,\mathbb{P}(|Z|>a_{np}). \tag{2.35}$$

The subset of S_1 (say S_l) which generates a $\widetilde{Z}^l$ for $l\ge 4$ is much smaller than S_1. Also its cardinality is divided by at least n if we go from l to $l+1$, i.e. $|S_l|\ge n|S_{l+1}|$. Observe that na_{np}^{-1} converges to infinity. This combined with (2.35) tells us that only the case of every $\widetilde{Z}$ appearing exactly twice is of interest since it has most influence on the expectation in (2.33). We conclude that

$$\frac{1}{a_{np}^{4\cdot 2^k}}\,\mathbb{E}\Big[\sum_{i,j=1}^p (\widetilde{R}^{2^{k-1}})_{ij}^2\Big] \le c\,\frac{|S_1|}{a_{np}^{4\cdot 2^k}} \le c\,p\,\Big(\frac{np}{a_{np}^4}\Big)^{2^k} \le c\,(np)\,\Big(\frac{np}{a_{np}^4}\Big)^{2^k}.$$

The expression on the right-hand side converges to 0 if $1+2^k-2^{k+2}/\alpha<0$ or equivalently

$$k > \log\Big(\frac{\alpha}{4-\alpha}\Big)(\log 2)^{-1}.$$

Since k was arbitrary the proof of the **R**-part is finished. □

2.4.2 Proof of (2.16)

We define the $p\times p$ matrix $\mathbf{Y}_n^{\rightarrow}$ as the diagonal matrix with elements

$$(\mathbf{Y}_n^{\rightarrow})_{ii} = \max_{t=1,\dots,n} Z_{it}^2\,, \qquad i=1,\dots,p\,.$$

Correspondingly, we define the $n\times n$ matrix $\mathbf{Y}_n^{\downarrow}$ as the diagonal matrix with elements

$$(\mathbf{Y}_n^{\downarrow})_{tt} = \max_{i=1,\dots,p} Z_{it}^2\,, \qquad t=1,\dots,n\,.$$

Lemma 2.13. *Assume the conditions of Theorem 2.1.*

1. *If $\beta\in((\alpha/2-1)_+,1]$ we have*

$$a_{np}^{-2}\max_{i=1,\dots,p}\big|\lambda_{(i)}-\lambda_{(i)}(\mathbf{Y}_n^{\rightarrow})\big| \stackrel{\mathbb{P}}{\to} 0\,, \qquad n\to\infty\,.$$

2. *If $\beta^{-1}\in((\alpha/2-1)_+,1)$ we have*

$$a_{np}^{-2}\max_{i=1,\dots,n}\big|\lambda_{(i)}-\lambda_{(i)}(\mathbf{Y}_n^{\downarrow})\big| \stackrel{\mathbb{P}}{\to} 0\,, \qquad n\to\infty\,.$$

Proof. We restrict ourselves to the proof in the case $\beta\in(0,1]$; the case $\beta>1$ can again be handled by switching from $\mathbf{ZZ}'$ to $\mathbf{Z}'\mathbf{Z}$. An application of Weyl's inequality (see (2.18)) and the triangle inequality yield

$$a_{np}^{-2}\max_{i=1,\dots,p}\big|\lambda_{(i)}-\lambda_{(i)}(\mathbf{Y}_n^{\rightarrow})\big| \le a_{np}^{-2}\|\mathbf{ZZ}'-\operatorname{diag}(\mathbf{ZZ}')\|_2 + a_{np}^{-2}\|\operatorname{diag}(\mathbf{ZZ}')-\operatorname{diag}(\mathbf{Y}_n^{\rightarrow})\|_2.$$

The first term on the right-hand side converges to 0 in probability by Theorem 2.5(1). As regards the second term we have

$$a_{np}^{-2}\|\operatorname{diag}(\mathbf{ZZ}')-\operatorname{diag}(\mathbf{Y}_n^{\rightarrow})\|_2 \;=\; a_{np}^{-2}\max_{i=1,\dots,p}\Big|D_i^{\rightarrow}-\max_{t=1,\dots,n} Z_{it}^2\Big|\,.$$

The right-hand side converges to zero in probability in view of Lemma 2.22 applied to (Z_{it}^2). □

Now (2.16) follows from the next result.

Lemma 2.14. *Assume the conditions of Theorem 2.1.*

1. *If $\beta\in((\alpha/2-1)_+,1]$ we have*

$$a_{np}^{-2}\max_{i=1,\dots,p}\big|\lambda_{(i)}(\mathbf{Y}_n^{\rightarrow})-Z_{(i),np}^2\big| \stackrel{\mathbb{P}}{\to} 0\,, \qquad n\to\infty\,.$$

2. If $\beta^{-1} \in ((\alpha/2-1)_+,1)$ we have

$$a_{np}^{-2} \max_{i=1,\ldots,n} \big|\lambda_{(i)}(\mathbf{Y}_n^{\downarrow}) - Z^2_{(i),np}\big| \stackrel{\mathbb{P}}{\to} 0\,, \qquad n\to\infty\,.$$

Proof. We focus on part (1). We write $V_{(1)} \ge \cdots \ge V_{(p)}$ for the order statistics of $(\max_{t=1,\ldots,n} Z_{it}^2)$. By definition of the order statistics we have $Z^2_{(i),np} \ge V_{(i)}$ for $i = 1,\ldots,p$. We choose δ such that $1 > \delta > \frac{2+\beta}{2(1+\beta)}$ and define the event

$$B_{np}^{2\delta} = \{\text{There is a row of } (Z_{it}^2) \text{ with at least two entries larger than } a_{np}^{2\delta}.\}\,.$$

By Lemma 2.21, $\mathbb{P}(B_{np}^{2\delta}) \to 0$.

Next, we choose $0<\varepsilon<1-\delta$. Then Lemma 2.20 guarantees the existence of a sequence $k=k_n\to\infty$ such that the event

$$\Omega_n = \{Z^2_{(k),np} > a_{np}^{2(1-\varepsilon)}\}$$

satisfies $\mathbb{P}(\Omega_n^c)\to 0$. On the event $(B_{np}^{2\delta})^c \cap \Omega_n$ we have

$$V_{(i)} - Z^2_{(i),np} = 0\,, \quad i=1,\ldots,k.$$

This shows for $\gamma>0$,

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\big(a_{np}^{-2} \max_{i=1,\ldots,p} |V_{(i)} - Z^2_{(i),np}| > \gamma\big)\\
&\quad\le \limsup_{n\to\infty} \mathbb{P}\big(\{a_{np}^{-2} \max_{i=1,\ldots,p} |V_{(i)} - Z^2_{(i),np}| > \gamma\} \cap (B_{np}^{2\delta})^c \cap \Omega_n\big)\\
&\qquad + \limsup_{n\to\infty} \mathbb{P}(B_{np}^{2\delta}) + \limsup_{n\to\infty} \mathbb{P}(\Omega_n^c)\\
&\quad= \limsup_{n\to\infty} \mathbb{P}\big(\{a_{np}^{-2} \max_{i=k+1,\ldots,p} |V_{(i)} - Z^2_{(i),np}| > \gamma\} \cap (B_{np}^{2\delta})^c \cap \Omega_n\big)\\
&\quad\le \limsup_{n\to\infty} \mathbb{P}\big(2\, a_{np}^{-2}\, Z^2_{(k+1),np} > \gamma\big) = 0.
\end{aligned}$$

□

2.5 Generalization to autocovariance matrices

An important topic in multivariate time series analysis is the study of the covariance structure. From the field (Z_{it}) we construct the $p\times n$ matrices

$$\mathbf{Z}(s,k) = \mathbf{Z}_n(s,k) = (Z_{i-s,t-k})_{i=1,\ldots,p;t=1,\ldots,n}\,, \quad s,k\in\mathbb{Z}\,.$$

We introduce the (non-normalized) *generalized sample autocovariance matrices*

$$(\mathbf{Z}(0,0)\mathbf{Z}(s,k)')\,, \quad s,k\in\mathbb{Z}\,,$$

with entries

$$(\mathbf{Z}(0,0)\mathbf{Z}(s,k)')_{ij} = \sum_{t=1}^n Z_{i,t}\, Z_{j-s,t-k}\,, \qquad i,j=1,\ldots,p\,.$$

If $\min(|s|,|k|)\neq 0$, the generalized sample autocovariance matrix $\mathbf{Z}(0,0)\mathbf{Z}(s,k)'$ is not symmetric and might thus have complex eigenvalues. In what follows, we will be interested in the *singular values* $\lambda_1(s,k),\ldots,\lambda_p(s,k)$ of $\mathbf{Z}(0,0)\mathbf{Z}(s,k)'$. The singular values of a matrix $\mathbf{A}$ are the square roots of the eigenvalues of $\mathbf{A}\mathbf{A}'$. We reuse the notation $(\lambda_i(s,k))$ for the singular values and again write $\lambda_{(1)}(s,k)\ge\cdots\ge\lambda_{(p)}(s,k)$ for their order statistics.

Theorem 2.15. *Assume $s, k \in \mathbb{Z}$. Consider the $p \times n$-dimensional matrices $\mathbf{Z}(0,0)$ and $\mathbf{Z}(s,k)$ with iid entries. We assume the following conditions:*

- *The regular variation condition (2.7) for some $\alpha \in (0,4)$.*
- *$\mathbb{E}[Z] = 0$ for $\alpha \geq 2$.*
- *The integer sequence (p_n) has growth rate $C_p(\beta)$ for some $\beta \geq 0$.*

(1) If $k \neq 0$, then

$$a_{np}^{-2}\, \lambda_{(1)}(s,k) \xrightarrow{\mathbb{P}} 0\,.$$

Now assume $k = 0$ and recall the notation $D^{\rightarrow}_{(i)}$ and $D^{\downarrow}_{(i)}$ from Section 2.2.2. Then the following statements hold:

(2) If $\beta \in [0,1]$, then

$$a_{np}^{-2} \max_{i=1,\ldots,p-|s|} \left|\lambda_{(i)}(s,0) - D^{\rightarrow}_{(i)}\right| \xrightarrow{\mathbb{P}} 0\,. \tag{2.36}$$

(3) If $\beta > 1$, then

$$a_{np}^{-2} \max_{i=1,\ldots,n-|s|} \left|\lambda_{(i)}(s,0) - D^{\downarrow}_{(i)}\right| \xrightarrow{\mathbb{P}} 0\,. \tag{2.37}$$

(4) If $\min(\beta, \beta^{-1}) \in ((\alpha/2-1)_+, 1]$, then

$$a_{np}^{-2} \max_{i=1,\ldots,p-|s|} \left|\lambda_{(i)}(s,0) - Z^2_{(i),np}\right| \xrightarrow{\mathbb{P}} 0\,. \tag{2.38}$$

Proof. We focus on the case $\beta \in [0,1]$. The proof is analogous to the proof of Theorem 2.1 which was given in Section 2.4. This proof relied on the reduction of $\mathbf{Z}\mathbf{Z}'$ to its diagonal. If $k = 0$, we will reduce $\mathbf{Z}(0,0)\mathbf{Z}(s,k)'$ to a $p \times p$ matrix $\mathbf{M}^{(s,k)}$, which only takes values on its sth sub-diagonal. The entries of the sth sub-diagonal of $\mathbf{M}^{(s,k)}$ are $\mathbf{M}^{(s,k)}_{i,i+s}$, $i = 1 + s_-, \ldots, p - s_+$. Here $s_+, s_- \geq 0$ are the positive and negative parts of s, respectively.

We sketch the steps of this reduction. Let $k \in \mathbb{Z}$. For simplicity of notation assume $s \geq 0$. Define the $p \times p$ matrix $\mathbf{M}^{(s,k)}$,

$$\mathbf{M}^{(s,k)}_{i,i+s} = \mathbf{1}_{\{k=0\}}(\mathbf{Z}(0,0)\mathbf{Z}(s,0)')_{i,i+s} = \mathbf{1}_{\{k=0\}} \sum_{t=1}^{n} Z^2_{it}\,, \quad i = 1, \ldots, p-s\,,$$

and $\mathbf{M}^{(s,k)}_{ij} = 0$ for all other i,j. We have

$$\begin{aligned}
\big((\mathbf{Z}(0,0)\mathbf{Z}(s,k)' - &\mathbf{M}^{(s,k)})(\mathbf{Z}(0,0)\mathbf{Z}(s,k)' - \mathbf{M}^{(s,k)})'\big)_{ij} \\
&= \sum_{u=1}^{p}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n} Z_{i,t_1} Z_{j,t_2} Z_{u-s,t_1-k} Z_{u-s,t_2-k} \mathbf{1}_{\{i\neq u-s, j\neq u-s\}} \\
&\quad \times (\mathbf{1}_{\{i=j\}} + \mathbf{1}_{\{i\neq j, t_1=t_2\}} + \mathbf{1}_{\{i\neq j, t_1\neq t_2\}}) \\
&= \mathbf{D}_{ij} + \mathbf{F}_{ij} + \mathbf{R}_{ij}\,.
\end{aligned}$$

Repeating the steps in the proof of Lemma 2.12, one obtains

$$a_{np}^{-4}\|\mathbf{D}+\mathbf{F}+\mathbf{R}\|_2^2 \xrightarrow{\mathbb{P}} 0\,.$$

Therefore we also have

$$\begin{aligned}&a_{np}^{-4}\|\mathbf{Z}(0,0)\mathbf{Z}(s,k)'-\mathbf{M}^{(s,k)}\|_2^2\\&\quad=a_{np}^{-4}\|(\mathbf{Z}(0,0)\mathbf{Z}(s,k)'-\mathbf{M}^{(s,k)})(\mathbf{Z}(0,0)\mathbf{Z}(s,k)'-\mathbf{M}^{(s,k)})'\|_2 \xrightarrow{\mathbb{P}} 0\,.\end{aligned}$$

This proves part (1). Since, with probability tending to 1, the matrix $\mathbf{M}^{(s,k)}$ has the required singular values, part (2) follows by Weyl's inequality.

Finally, part (4) is a consequence of Lemma 2.14. □

We obtain the following result for the weak convergence of the point processes of the points $\lambda_i(s,0)$, $s=0,\dots,l$; the proof is similar to the one of Theorem 2.15.

Corollary 2.16. *Assume the conditions of Theorem 2.15. Then, with the notation of Theorem 2.10, the following point process convergence holds for $l\ge 0$ and $(\beta,\beta^{-1})\in((\alpha/2-1)_+,1]$,*

$$\sum_{i=1}^{p}\varepsilon_{a_{np}^{-2}\left(\lambda_{(i)}(0,0),\dots,\lambda_{(i)}(l,0)\right)} \xrightarrow{\mathrm{d}} \sum_{i=1}^{\infty}\varepsilon_{\Gamma_i^{-2/\alpha}\left(1,\dots,1\right)}\,.$$

The joint convergence of a finite number of the random variables $\lambda_{(i)}(s,0)$, $i\ge 1$, $s\ge 0$, is an immediate consequence of this result.

2.6 Appendix

Let (Z_i) be iid copies of Z whose distribution satisfies

$$\mathbb{P}(Z>x)\sim p_+\frac{L(x)}{x^\alpha}\quad\text{and}\quad\mathbb{P}(Z\le -x)\sim p_-\frac{L(x)}{x^\alpha}\,,\quad x\to\infty\,,$$

for some tail index $\alpha>0$, where $p_+,p_-\ge 0$ with $p_++p_-=1$ and L is a slowly varying function. We say that Z is regularly varying with index α. The monograph [20] contains many properties and useful tools for regularly varying functions. Theorem 1.5.6 therein, which is known as Potter bounds, asserts that a regularly varying function essentially lies between two power laws. In particular, for any $\delta>0$ and $C>1$ we have for x sufficiently large,

$$C^{-1}x^{-\delta}\le L(x)\le Cx^{\delta}\,.$$

Theorem 1.6.1 in [20], widely known as Karamata's theorem, describes the behavior of truncated moments of the regularly varying random variable Z. For $x\to\infty$,

$$\begin{aligned}\mathbb{E}[|Z|^\beta\mathbf{1}_{\{|Z|\le x\}}]&\sim\frac{\alpha}{\beta-\alpha}x^\beta\mathbb{P}(|Z|>x),\quad \beta>\alpha,\\ \mathbb{E}[|Z|^\beta\mathbf{1}_{\{|Z|>x\}}]&\sim\frac{\alpha}{\alpha-\beta}x^\beta\mathbb{P}(|Z|>x),\quad \beta<\alpha.\end{aligned}$$

If $\mathbb{E}[|Z|]<\infty$ also assume $\mathbb{E}[Z]=0$. The product Z_1Z_2 is regular varying with the same index α and $\mathbb{P}(|Z_1Z_2|>x)=x^{-\alpha}L_1(x)$, where L_1 is slowly varying function different from L; see Embrechts and Goldie [34]. Write

$$S_n=Z_1+\cdots+Z_n\,,\quad n\ge 1,$$

and consider a sequence (a_n) such that $\mathbb{P}(|Z|>a_n)\sim n^{-1}$.

2.6.1 Large deviation results

The following theorem can be found in Nagaev [55] and Cline and Hsing [22] for $\alpha > 2$ and $\alpha \le 2$, respectively; see also Denisov et al. [27].

Theorem 2.17. *Under the assumptions on the iid sequence (Z_t) given above the following relation holds*

$$\sup_{x\ge c_n}\left|\frac{\mathbb{P}(S_n > x)}{n\mathbb{P}(|Z| > x)} - p_+\right| \to 0\,,$$

where (c_n) is any sequence satisfying $c_n/a_n \to \infty$ for $\alpha \le 2$ and $c_n \ge \sqrt{(\alpha-2)n\log n}$ for $\alpha > 2$.

2.6.2 Karamata theory for sums

Proposition 2.18. *Let (c_n) be the threshold sequence in Theorem 2.17 for a given $\alpha > 0$, and let (d_n) be such that $d_n/c_n \to \infty$ for $\alpha > 2$ and $d_n = c_n$ for $\alpha \le 2$. Assume $0 < \gamma < \alpha$. Then we have for a sequence $x_n \ge d_n$*

$$\mathbb{E}[|x_n^{-1}S_n|^\gamma \mathbf{1}_{\{|S_n|>x_n\}}] \sim \frac{\alpha}{\alpha-\gamma} n\mathbb{P}(|Z| > x_n), \qquad n\to\infty. \tag{2.39}$$

Proof. We use the notation $Y_n := |x_n^{-1}S_n|$. Since $Y_n^\gamma \mathbf{1}_{\{Y_n>1\}}$ is a positive random variable one can write

$$\mathbb{E}[Y_n^\gamma \mathbf{1}_{\{Y_n>1\}}] = \int_0^\infty \mathbb{P}(Y_n^\gamma \mathbf{1}_{\{Y_n>1\}} > y)\,\mathrm{d}y.$$

The probability inside the integral is

$$\begin{aligned}\mathbb{P}(Y_n^\gamma \mathbf{1}_{\{Y_n>1\}} > y) &= \mathbb{P}(Y_n^\gamma \mathbf{1}_{\{Y_n>1\}} > y, Y_n > 1) + \mathbb{P}(Y_n^\gamma \mathbf{1}_{\{Y_n>1\}} > y, Y_n < 1)\\ &= \mathbb{P}(Y_n^\gamma > y, Y_n > 1) = \mathbb{P}(Y_n > \max\{y^{1/\gamma}, 1\})\\ &= \begin{cases}\mathbb{P}(Y_n > 1) & \text{if } y \le 1,\\ \mathbb{P}(Y_n > y^{1/\gamma}) & \text{if } y \ge 1.\end{cases}\end{aligned}$$

Therefore, using the uniform convergence result in Theorem 2.17, we conclude that

$$\begin{aligned}\int_0^\infty \mathbb{P}(Y_n^\gamma \mathbf{1}_{\{Y_n>1\}} > y)\,\mathrm{d}y &= \mathbb{P}(Y_n > 1) + \int_1^\infty \mathbb{P}(Y_n > y^{1/\gamma})\,\mathrm{d}y\\ &\sim n\mathbb{P}(|Z| > x_n) + \int_1^\infty y^{-\frac{\alpha}{\gamma}} n\mathbb{P}(|Z| > x_n)\,\mathrm{d}y\\ &= \frac{\alpha}{\alpha-\gamma} n\mathbb{P}(|Z| > x_n), \quad n\to\infty.\end{aligned}$$

□

2.6.3 A point process convergence result

Assume that the conditions at the beginning of Appendix 2.6 hold. Consider a sequence of iid copies $(S_n^{(t)})_{t=1,2,\ldots}$ of S_n and the sequence of point processes

$$N_n = \sum_{t=1}^p \varepsilon_{a_{np}^{-1}S_n^{(t)}}, \quad n = 1,2,\ldots,$$

for an integer sequence $p = p_n \to \infty$. We assume that the state space of the point processes N_n is $\overline{\mathbb{R}}_0 = [\mathbb{R}\cup\{\pm\infty\}]\backslash\{0\}$.

Lemma 2.19. *Assume* $\alpha \in (0,2)$ *and the conditions of Appendix 2.6 on the iid sequence* (Z_t) *and the normalizing sequence* (a_n)*. Then the limit relation* $N_n \stackrel{\rm d}{\to} N$ *holds in the space of point measures on* $\overline{\mathbb{R}}_0$ *equipped with the vague topology (see [61, 60]) for a Poisson random measure* N *with state space* $\overline{\mathbb{R}}_0$ *and intensity measure* $\mu_\alpha(dx) = \alpha |x|^{-\alpha-1}(p_+ \mathbf{1}_{\{x>0\}} + p_- \mathbf{1}_{\{x<0\}})dx$*.*

Proof. According to Resnick [61], Proposition 3.21, we need to show that $p\,\mathbb{P}(a_{np}^{-1} S_n \in \cdot) \stackrel{\rm v}{\to} \mu_\alpha$, where $\stackrel{\rm v}{\to}$ denotes vague convergence of Radon measures on $\overline{\mathbb{R}}_0$. Observe that we have $a_{np}/a_n \to \infty$ as $n \to \infty$. This fact and $\alpha \in (0,2)$ allow one to apply Theorem 2.17:

$$\frac{\mathbb{P}(S_n > x a_{np})}{n\,\mathbb{P}(|Z| > a_{np})} \to p_+ x^{-\alpha} \quad \text{and} \quad \frac{\mathbb{P}(S_n \le -x a_{np})}{n\,\mathbb{P}(|Z| > a_{np})} \to p_-\, x^{-\alpha}\,, \quad x > 0\,.$$

On the other hand, $n\,\mathbb{P}(|Z| > a_{np}) \sim p^{-1}$ as $n \to \infty$. This proves the lemma. □

2.6.4 Auxiliary results

Assume that the non-negative random variable Z is regularly varying with index $\alpha \in (0,2)$ and (a_n) is such that $n\,\mathbb{P}(Z > a_n) \sim 1$. We also write

$$Z_{(1)} \ge \cdots \ge Z_{(n)}\,,$$

for the order statistics of the iid copies $Z_1, \ldots, Z_n$ of Z.

Lemma 2.20. *For every* $\varepsilon \in (0, 0.5)$ *there exists a sequence* $k = k_n \to \infty$*,* $k < n$ *such that*

$$\lim_{n\to\infty} \mathbb{P}(Z_{(k)} > a_n^{1-\varepsilon}) = 1.$$

Proof of Lemma 2.20. From the theory of order statistics we know that

$$\begin{aligned} \mathbb{P}(Z_{(k)} \le a_n^{1-\varepsilon}) &= \sum_{r=0}^{k-1} \binom{n}{r} \mathbb{P}(Z > a_n^{1-\varepsilon})^r\, \mathbb{P}(Z \le a_n^{1-\varepsilon})^{n-r} \\ &\le \big(\mathbb{P}(Z \le a_n^{1-\varepsilon})\big)^n \sum_{r=0}^{k-1} \frac{1}{r!} \Big(\frac{n\,\mathbb{P}(Z > a_n^{1-\varepsilon})}{\mathbb{P}(Z \le a_n^{1-\varepsilon})}\Big)^r. \end{aligned}$$

We observe that

$$\big(\mathbb{P}(Z \le a_n^{1-\varepsilon})\big)^n \sim \mathrm{e}^{-n\,\big[\mathbb{P}(Z > a_n^{1-\varepsilon}) - 0.5(\mathbb{P}(Z > a_n^{1-\varepsilon}))^2(1+o(1))\big]}$$

Writing $\Gamma(k)$ and $\Gamma(k,y)$ for the gamma and incompete gamma functions, we have

$$\mathrm{e}^{-y} \sum_{r=0}^{k-1} \frac{y^r}{r!} = \frac{\Gamma(k,y)}{\Gamma(k)} = \mathbb{P}(\Gamma_k > y), \qquad y \ge 0\,,$$

where $\Gamma_k = E_1 + \cdots + E_k$, $k \ge 1$, for an iid standard exponential sequence (E_i). Therefore

$$\begin{aligned} &\mathbb{P}(Z_{(k)} \le a_n^{1-\varepsilon}) \\ &\le c\,\mathrm{e}^{-n\,\big[\mathbb{P}(Z > a_n^{1-\varepsilon}) - 0.5(\mathbb{P}(Z > a_n^{1-\varepsilon}))^2(1+o(1))\big] + \big[n\,\mathbb{P}(Z > a_n^{1-\varepsilon})/\mathbb{P}(Z \le a_n^{1-\varepsilon})\big]} \\ &\quad \mathbb{P}\big(\Gamma_k > n\,\mathbb{P}(Z > a_n^{1-\varepsilon})/\mathbb{P}(Z \le a_n^{1-\varepsilon})\big) \\ &= c\,\mathrm{e}^{O\big(n\,(\mathbb{P}(Z > a_n^{1-\varepsilon}))^2\big)}\, \mathbb{P}\big(k^{-1}\Gamma_k > k^{-1} n\,\mathbb{P}(Z > a_n^{1-\varepsilon})/\mathbb{P}(Z \le a_n^{1-\varepsilon})\big)\,. \end{aligned}$$

The right-hand side converges to zero if $2\varepsilon < 1$ and $k \le n^{\varepsilon'}$ for some $\varepsilon' < \varepsilon$. □

Now consider a $p \times n$ random matrix $\mathbf{Z}$ with iid non-negative entries Z_{it} and generic element Z as specified above. The number of rows p satisfies the growth condition $C_p(\beta)$.

We write for $\delta > 0$,

$$B_{np}^\delta = \{\text{There is a row of } \mathbf{Z} \text{ with at least two entries larger than } a_{np}^\delta.\}\,. \quad (2.40)$$

Lemma 2.21. *Assume that $p = p_n$ satisfies the growth condition $C_p(\beta)$ with $\beta \in [0,1]$. Then we have*

$$\lim_{n\to\infty} \mathbb{P}(B_{np}^\delta) = 0 \quad \textit{for all} \quad \delta > \frac{2+\beta}{2(1+\beta)}.$$

Proof of Lemma 2.21. Assume $\delta > \frac{2+\beta}{2(1+\beta)}$ and consider the counting variables

$$N_i = \sum_{t=1}^n \mathbf{1}_{\{Z_{it} > a_{np}^\delta\}}, \qquad i = 1, \ldots, p.$$

Clearly, N_i are iid $\text{Bin}(n,q)$ with $q = q_n = \mathbb{P}(Z > a_{np}^\delta) \to 0$ as $n \to \infty$ and

$$\begin{aligned} \mathbb{P}(B_{np}^\delta) &= \mathbb{P}(\max_{i=1,\ldots,p} N_i \geq 2) \\ &= 1 - \big(\mathbb{P}(N_1 \leq 1)\big)^p \\ &= 1 - \big((1-q)^{n-1}(1+(n-1)q)\big)^p\,. \end{aligned}$$

Thus it remains to show that the right-hand side converges to 0. Taking logarithms, we get

$$p \,\log\big((1-q)^{n-1}\,(1+(n-1)q)\big) = p\,[(n-1)\log(1-q) + \log(1+(n-1)q)].$$

A second order Taylor expansion of the logarithm yields

$$p\,(n-1)\,\log(1-q) + p\,\log(1+(n-1)q) = p\,q + p\,\frac{(nq)^2}{2} + O(p\,(nq^2 + (nq)^3)\,)\,. \quad (2.41)$$

By the Potter bounds we conclude that (2.41) converges to zero if $\delta > \frac{2+\beta}{2(1+\beta)}$. The proof is complete. □

For $\varepsilon \in (0,1)$ define the events

$$A_i^{(n)}(\varepsilon) = \Big\{\sum_{t=1}^n Z_{it} - \max_{t=1,\ldots,n} Z_{it} > a_{np}^{1-\varepsilon}\Big\}, \qquad i = 1, \ldots, p\,.$$

The following result generalizes Lemma 5 in Auffinger et al. [4] (which in turn is a modified version of a result in Soshnikov [65]) to the case of regularly varying growth rates (p_n). The method of proof is different from the aforementioned literature.

Lemma 2.22. *Assume that $p = p_n = n^\beta \ell(n)$ where ℓ is a slowly varying function. Assume $\beta \in (0,\infty)$ for $\alpha \in (0,1]$ and $\beta \in (\alpha-1,\infty)$ for $\alpha \in [1,2)$. There exists a constant $\varepsilon \in (0,1)$ such that*

$$\lim_{n\to\infty} \mathbb{P}\Big(\bigcup_{i=1}^p A_i^{(n)}(\varepsilon)\Big) = 0\,.$$

Proof. Write $M_t = \max_{i=1,\dots,t} Z_i$. We observe that

$$\begin{aligned}\mathbb{P}\Big(\bigcup_{i=1}^{p} A_i^{(n)}(\varepsilon)\Big) &\le p\,\mathbb{P}(S_n - M_n > a_{np}^{1-\varepsilon})\\ &= n\,p\,\mathbb{P}(S_{n-1} > a_{np}^{1-\varepsilon}\,, Z_n > M_{n-1})\\ &= n\,p\int_0^\infty \mathbb{P}(S_{n-1} > a_{np}^{1-\varepsilon}\,, z > M_{n-1})\;\mathrm{d}\mathbb{P}(Z\le z)\,.\end{aligned}$$

We split the integration area into disjoint sets:

$$[0,\infty) = [0, a_n/h_n] \cup (a_n/h_n, a_{np}^{\gamma}] \cup (a_{np}^{\gamma}, \infty) = \bigcup_{i=1}^{3} B_i\,.$$

We choose $h_n \to \infty$ such that $n\,\mathbb{P}(Z > a_n/h_n) \sim 2\log(np)$. Then

$$\log(np) - n\,\mathbb{P}(Z > a_n/h_n) \to -\infty\,, \qquad n\,\big(\mathbb{P}(Z > a_n/h_n)\big)^2 \to 0\,. \tag{2.42}$$

Moreover, choose γ and $\varepsilon > 0$ fixed such that $\varepsilon < 1 - (1\vee\alpha)/(1+\beta)$ and

- $\frac{1}{1+\beta} + \varepsilon < \gamma < 1 - \frac{\varepsilon}{1-\alpha}$ if $\alpha \in (0,1)$ and
- $\frac{1}{1+\beta} + \varepsilon < \gamma < 1 - \frac{2\varepsilon}{2-\alpha}$ if $\alpha \in [1,2)$.

By virtue of (2.42) we have

$$\begin{aligned} n\,p\int_{B_1} \mathbb{P}(S_{n-1} > a_{np}^{1-\varepsilon}\,, z > M_{n-1})\;\mathrm{d}\mathbb{P}(Z\le z) &\le n\,p\,\mathbb{P}(M_{n-1} \le a_n/h_n)\\ &= \mathrm{e}^{\log(np) - n\,\mathbb{P}(Z > a_n/h_n) + o(1)} \to 0\,.\end{aligned}$$

By definition of ε, we have $(a_n + n)/a_{np}^{1-\varepsilon} \to 0$ for $\alpha \in (0,2)$. Therefore an application of Theorem 2.17 yields

$$\begin{aligned} n\,p\int_{B_3} \mathbb{P}(S_{n-1} > a_{np}^{1-\varepsilon}\,, z > M_{n-1})\;\mathrm{d}\mathbb{P}(Z\le z) &\le n\,p\,\mathbb{P}(S_{n-1} > a_{np}^{1-\varepsilon})\,\mathbb{P}(Z > a_{np}^{\gamma})\\ &\sim \big(n\,p\,\mathbb{P}(Z > a_{np}^{1-\varepsilon})\big)\,\big(n\,\mathbb{P}(Z > a_{np}^{\gamma})\big)\,.\end{aligned}$$

The right-hand side converges to zero due to the property $\gamma > 1/(1+\beta) + \varepsilon$.

Now assume $\alpha \in (0,1)$. Then we have by Markov's inequality and Karamata's theorem,

$$\begin{aligned} &n\,p\int_{B_2} \mathbb{P}(S_{n-1} > a_{np}^{1-\varepsilon}\,, z > M_{n-1})\;\mathrm{d}\mathbb{P}(Z\le z)\\ \le\;& \frac{n^2\,p}{a_{np}^{1-\varepsilon}}\int_{B_2} \mathbb{E}[Z\mathbf{1}_{\{Z\le z\}}]\,\mathrm{d}\mathbb{P}(Z\le z)\\ \le\;& \frac{n^2\,p}{a_{np}^{1-\varepsilon}}\,\mathbb{E}[Z\mathbf{1}_{\{Z\le a_{np}^{\gamma}\}}]\,\mathbb{P}(Z > a_n/h_n)\\ \sim\;& c\,\frac{n\,p}{a_{np}^{1-\varepsilon}}\,\big[a_{np}^{\gamma}\,\mathbb{P}(Z > a_{np}^{\gamma})\big]\;\log(np)\,.\end{aligned}$$

An application of the Potter bounds and using the fact that $\gamma < 1 - \varepsilon/(1-\alpha)$ shows that the right-hand side converges to zero for the chosen ε.

Now assume $\alpha \in [1,2)$ and $\beta > \alpha - 1$. Due to the latter condition we have $n/a_{np}^{1-\varepsilon} \to 0$. We obtain by Čebyshev's inequality and Karamata's theorem,

$$\begin{aligned}
& n\,p \int_{B_2} \mathbb{P}(S_{n-1} > a_{np}^{1-\varepsilon}\,, z > M_{n-1})\,\mathrm{d}\mathbb{P}(Z \le z) \\
& \le\; n\,p \int_{B_2} \mathbb{P}\Big(\sum_{t=1}^{n} Z_t\,\mathbf{1}_{\{Z_t \le a_{np}^{\gamma}\}} - n\mathbb{E}\big[Z\mathbf{1}_{\{Z \le a_{np}^{\gamma}\}}\big] > a_{np}^{1-\varepsilon} - n\,\mathbb{E}\big[Z\mathbf{1}_{\{Z \le a_{np}^{\gamma}\}}\big]\Big)\,\mathrm{d}\mathbb{P}(Z \le z) \\
& \le\; n\,p \int_{B_2} \mathbb{P}\Big(\sum_{t=1}^{n} Z_t\,\mathbf{1}_{\{Z_t \le a_{np}^{\gamma}\}} - n\,\mathbb{E}\big[Z\,\mathbf{1}_{\{Z \le a_{np}^{\gamma}\}}\big] > c\,a_{np}^{1-\varepsilon}\Big)\,\mathrm{d}\mathbb{P}(Z \le z) \\
& \le\; n\,p\,\frac{\mathbb{E}[Z^2\,\mathbf{1}_{\{Z \le a_{np}^{\gamma}\}}]}{a_{np}^{2(1-\varepsilon)}}\,\big[n\,\mathbb{P}(Z > a_n/h_n)\big] \\
& \sim\; c\,n\,p\,\frac{a_{np}^{2\gamma}\mathbb{P}(Z > a_{np}^{\gamma})}{a_{np}^{2(1-\varepsilon)}}\,\log(np)\,.
\end{aligned}$$

The right-hand side converges to zero since $\gamma < 1 - 2\varepsilon/(2-\alpha)$. This finishes the proof. □

2.6.5 Perturbation theory for eigenvectors

We state Proposition A.1 in Benaych-Georges and Péché [18].

Proposition 2.23. *Let* $\mathbf{H}$ *be a Hermitean matrix and* $\mathbf{v}$ *a unit vector such that for some* $\lambda \in \mathbb{R}$, $\varepsilon > 0$,

$$\mathbf{H}\,\mathbf{v} = \lambda\,\mathbf{v} + \varepsilon\,\mathbf{w}\,,$$

where $\mathbf{w}$ *is a unit vector such that* $\mathbf{w} \perp \mathbf{v}$.

1. *Then* $\mathbf{H}$ *has an eigenvalue* λ_ε *such that* $|\lambda - \lambda_\varepsilon| \le \varepsilon$.

2. *If* $\mathbf{H}$ *has only one eigenvalue* λ_ε *(counted with multiplicity) such that* $|\lambda - \lambda_\varepsilon| \le \varepsilon$ *and all other eigenvalues are at distance at least* $d > \varepsilon$ *from* λ. *Then for a unit eigenvector* $\mathbf{v}_\varepsilon$ *associated with* λ_ε *we have*

$$\|\mathbf{v}_\varepsilon - \mathbf{P}_{\mathbf{v}}(\mathbf{v}_\varepsilon)\|_{\ell_2} \le \frac{2\,\varepsilon}{d - \varepsilon}\,,$$

where $\mathbf{P}_{\mathbf{v}}$ *denotes the orthogonal projection onto* $\mathrm{Span}(\mathbf{v})$.

Chapter 3

Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series

Richard Davis, Johannes Heiny,
Thomas Mikosch & Xiaolei Xie
Extremes 19, 3 (2016), 517–547.

Abstract

We provide some asymptotic theory for the largest eigenvalues of a sample covariance matrix of a p-dimensional time series where the dimension $p = p_n$ converges to infinity when the sample size n increases. We give a short overview of the literature on the topic both in the light- and heavy-tailed cases when the data have finite (infinite) fourth moment, respectively. Our main focus is on the heavy-tailed case. In this case, one has a theory for the point process of the normalized eigenvalues of the sample covariance matrix in the iid case but also when rows and columns of the data are linearly dependent. We provide limit results for the weak convergence of these point processes to Poisson or cluster Poisson processes. Based on this convergence we can also derive the limit laws of various functionals of the ordered eigenvalues such as the joint convergence of a finite number of the largest order statistics, the joint limit law of the largest eigenvalue and the trace, limit laws for successive ratios of ordered eigenvalues, etc. We also develop some limit theory for the singular values of the sample autocovariance matrices and their sums of squares. The theory is illustrated for simulated data and for the components of the S&P 500 stock index.

Keywords: Regular variation, sample covariance matrix, dependent entries, largest eigenvalues, trace, point process convergence, cluster Poisson limit, infinite variance stable limit, Fréchet distribution.

3.1 Estimation of the largest eigenvalues: an overview in the iid case

3.1.1 The light-tailed case

One of the exciting new areas of statistics is concerned with analyses of large data sets. For such data one often studies the dependence structure via covariances and correlations. In this paper we focus on one aspect: the estimation of the eigenvalues of the covariance matrix of a multivariate time series when the dimension p of the series increases with the sample size n. In particular, we are interested in limit theory for the largest eigenvalues of the sample covariance matrix. This theory is closely related to topics from classical extreme value theory such as maximum domains of attraction with the corresponding normalizing and centering constants for maxima; cf. Embrechts et al. [35], Resnick [60, 61]. Moreover, point process convergence with limiting Poisson and cluster Poisson processes enters in a natural way when one describes the joint convergence of the largest eigenvalues of the sample covariance matrix. Large deviation techniques find applications, linking extreme value theory with random walk theory and point process convergence. The objective of this paper is to illustrate some of the main developments in random matrix theory for the particular case of the sample covariance matrix of multivariate time series with independent or dependent entries. We give special emphasis to the heavy-tailed case when extreme value theory enters in a rather straightforward way.

Classical multivariate time series analysis deals with observations which assume values in a p-dimensional space where p is "relatively small" compared to the sample size n. With the availability of large data sets p can be "large" relative to n. One of the possible consequences is that standard asymptotics (such as the central limit theorem) break down and may even cause misleading results.

The dependence structure in multivariate data is often summarized by the covariance matrix which is typically estimated by its sample analog. For example, principal component analysis (PCA) extracts principal component vectors corresponding to the largest eigenvalues of the sample covariance matrix. The magnitudes of these eigenvalues provide an empirical measure of the importance of these components.

If p, n are fixed, a column of the $p \times n$ data matrix

$$\mathbf{X} = \mathbf{X}_n = \big(X_{it}\big)_{i=1,\ldots,p;t=1,\ldots,n}$$

represents an observation of a p-dimensional time series model with unknown parameters. In this section we assume that the real-valued entries X_{it} are iid, unless mentioned otherwise, and we write X for a generic element. One challenge is to infer information about the parameters from the eigenvalues $\lambda_1, \ldots, \lambda_p$ of the *sample covariance matrix* $\mathbf{X}\mathbf{X}'$. In the notation we suppress the dependence of (λ_i) on n and p. If p and n are finite and the columns of $\mathbf{X}$ are iid and multivariate normal, Muirhead [54] derived a (rather complicated) formula for the joint distribution of the eigenvalues (λ_i).

For p fixed and $n \to \infty$, assuming $\mathbf{X}$ has centered normal entries and a diagonal covariance matrix Σ, Anderson [3] derived the joint asymptotic density of $(\lambda_1, \ldots, \lambda_p)$. We quote from Johnstone [48]: "The classic paper by Anderson [3] gives the limiting joint distribution of the roots, but the marginal distribution of the largest eigenvalue is hard to extract even in the null case" (i.e., when the covariance matrix Σ is proportional to the identity matrix).

It turns out that limit theory for the largest eigenvalues becomes "easier" when the dimension p increases with n. Over the last 15 years there has been increasing interest in the case when $p = p_n \to \infty$ as $n \to \infty$. In most of the literature (exceptions are El Karoui [30], Davis et al. [24, 25] and Heiny and Mikosch [41]) one assumes that p and n grow at the same rate:

$$\frac{p}{n} \to \gamma \qquad \text{for some } \gamma \in (0, \infty). \tag{3.1}$$

In random matrix theory, the convergence of the *empirical spectral distributions* $(F_{n^{-1}\mathbf{XX}'})$ of a sequence $(n^{-1}\mathbf{XX}')$ of non-negative definite matrices is the principle object of study. The empirical spectral distribution $F_{n^{-1}\mathbf{XX}'}$ is constructed from the eigenvalues via

$$F_{n^{-1}\mathbf{XX}'}(x) = \frac{1}{p}\,\#\{1 \le j \le p : n^{-1}\lambda_j \le x\}, \quad x \in \mathbb{R}, \quad n \ge 1.$$

In the literature convergence results for the sequence of empirical spectral distributions are established under the assumption that p and n grow at the same rate. Suppose that the iid entries Z_{it} have mean 0 and variance 1. If (3.1) holds, then, with probability one, $(F_{n^{-1}\mathbf{XX}'})$ converges weakly to the celebrated Marčenko–Pastur law F_γ. If $\gamma \in (0, 1]$, F_γ has density,

$$f_\gamma(x) = \begin{cases} \frac{1}{2\pi x\gamma}\sqrt{(b-x)(x-a)}\,, & \text{if } a \le x \le b, \\ 0\,, & \text{otherwise,} \end{cases} \tag{3.2}$$

where $a = (1-\sqrt{\gamma})^2$ and $b = (1+\sqrt{\gamma})^2$. If $\gamma > 1$, the Marčenko–Pastur law is a mixture of a point mass at 0 and the density function $f_{1/\gamma}$ with weights $1 - 1/\gamma$ and $1/\gamma$, respectively. The point mass at 0 is intuitively explained by the fact that, with probability 1, $\min(p, n)$ eigenvalues λ_i are non-zero. When $n = (1/\gamma)\,p$ and $\gamma > 1$ one sees that the proportion of non-zero eigenvalues of the sample covariance matrix is $1/\gamma$ while the proportion of zero eigenvalues is $1 - 1/\gamma$.

While the finite second moment is the central assumption to obtain the Marčenko–Pastur law as the limiting spectral distribution, the finite fourth moment plays a crucial role when studying the largest eigenvalues

$$\lambda_{(1)} \ge \cdots \ge \lambda_{(p)} \tag{3.3}$$

of $\mathbf{XX}'$, where we suppress the dependence on n in the notation.

Assuming (3.1) and iid entries X_{it} with zero mean, unit variance and finite fourth moment, Geman [38] showed that

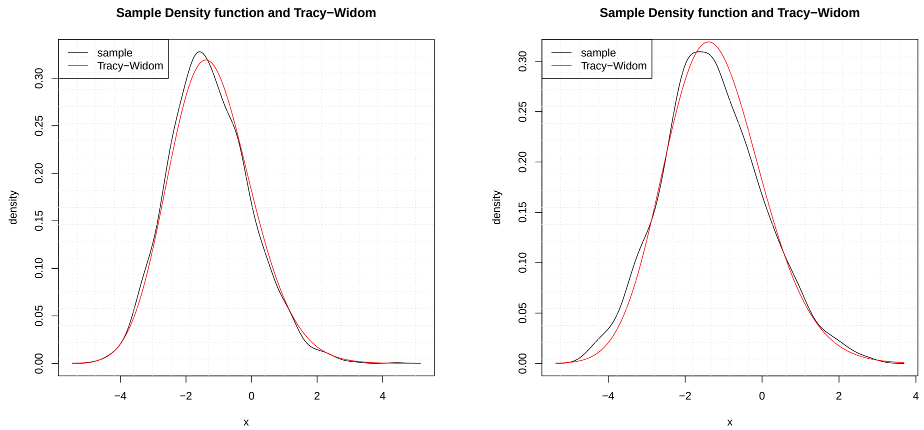
$$\frac{\lambda_{(1)}}{n} \overset{\text{a.s.}}{\to} \big(1+\sqrt{\gamma}\big)^2\,, \qquad n \to \infty\,. \tag{3.4}$$

Johnstone [48] complemented this strong law of large numbers by the corresponding central limit theorem in the special case of iid standard normal entries:

$$n^{2/3}\,\frac{(\sqrt{\gamma})^{1/3}}{\big(1+\sqrt{\gamma}\big)^{4/3}}\Big(\frac{\lambda_{(1)}}{n} - \big(1+\sqrt{\tfrac{p}{n}}\big)^2\Big) \overset{\text{d}}{\to} \text{TW}\,, \tag{3.5}$$

where the limiting random variable has a *Tracy–Widom distribution* of order 1. Notice that the centering $\big(1+\sqrt{\frac{p}{n}}\big)^2$ can in general not be replaced by $(1+\sqrt{\gamma})^2$. This distribution is ubiquitous in random matrix theory. Its distribution function F_1 is given by

$$F_1(s) = \exp\Big\{-\frac{1}{2}\int_s^\infty [q(x) + (x-s)q^2(x)]\,\mathrm{d}x\Big\}\,,$$



(a) Standard normal entries

(b) Entry distribution: $\mathbb{P}(X = \sqrt{3}) = \mathbb{P}(X = -\sqrt{3}) = 1/6$, $\mathbb{P}(X = 0) = 2/3$. Note $\mathbb{E}X = 0$, $\mathbb{E}[X^2] = 1$, $\mathbb{E}[X^3] = 0$ and $\mathbb{E}[X^4] = 3$, i.e., the first 4 moments of X match those of the standard normal distribution.

Figure 3.1: Sample density function of the largest eigenvalue compared with the Tracy–Widom density function. The data matrix $\mathbf{X}$ has dimension 200×1000. An ensemble of 2000 matrices is simulated.

where $q(x)$ is the unique solution to the Painlevé II differential equation

$$q''(x) = xq(x) + 2q^3(x)\,,$$

where $q(x) \sim \mathrm{Ai}(x)$ as $x \to \infty$ and $\mathrm{Ai}(\cdot)$ is the Airy kernel; see Tracy and Widom [71] for details. We notice that the rate $n^{2/3}$ compares favorably to the $\sqrt{n}$-rate in the classical central limit theorem for sums of iid finite variance random variables. The calculation of the spectrum is facilitated by the fact that the distribution of the classical Gaussian matrix ensembles is invariant under orthogonal transformations. The corresponding computation for non-invariant matrices with non-Gaussian entries is more complicated and was a major challenge for several years; a first step was made by Johansson [47]. Johnstone's result was extended to matrices $\mathbf{X}$ with iid non-Gaussian entries by Tao and Vu [68, Theorem 1.16]. Assuming that the first four moments of the entry distribution match those of the standard normal distribution, they showed (3.5) by employing *Lindeberg's replacement method*, i.e., the iid non-Gaussian entries are replaced step-by-step by iid Gaussian ones. This approach is well-known from summation theory for sequences of iid random variables. Tao and Vu's result is a consequence of the so-called *Four Moment Theorem*, which describes the insensitivity of the eigenvalues with respect to changes in the distribution of the entries. To some extent (modulo the strong moment matching conditions) it shows the universality of Johnstone's limit result (3.5). Later we will deal with entries with infinite fourth moment. In this case, the weak limit for the normalized largest eigenvalue $\lambda_{(1)}$ is distinct from the Tracy–Widom distribution: the classical Fréchet extreme value distribution appears. In Figure 3.1 we illustrate how the Tracy–Widom approximation works for Gaussian and non-Gaussian entries of $\mathbf{X}$ and in Figure 3.2 we also illustrate that this approach fails when $\mathbb{E}[X^4] = \infty$.

Figure 3.1 compares the sample density function of the properly normalized largest eigenvalue estimated from 2000 simulated sample covariance matrices $\mathbf{XX}'$ ($n = 1000, p = 200$) with the Tracy–Widom density. If X has infinite fourth moment and further regularity conditions on the tail hold then the Tracy–Widom limiting law needs to be replaced by the Fréchet distribution; see Section 3.1.2 for details. Figure 3.2 illustrates this fact with a simulated ensemble whose entries are distributed according to the heavy-tailed distribution from (3.33) below with $\alpha = 1.6$.

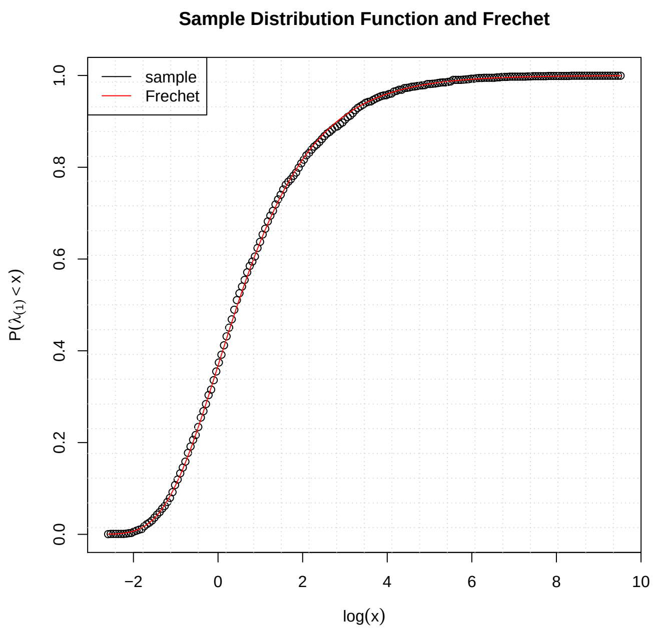


Figure 3.2: Sample distribution function of the largest eigenvalue $\lambda_{(1)}$ compared to the Fréchet distribution (solid line) with $\alpha = 1.6$. The data matrices have dimension 200×1000 and iid entries with infinite fourth moment. The results are based on 2000 replicates.

3.1.2 The heavy-tailed case

So far we focused on "light-tailed" $\mathbf{X}$ in the sense that its entries have finite fourth moment. However, there is statistical evidence that the assumption of finite fourth moment may be violated when dealing with data from insurance, finance or telecommunications. We illustrate this fact in Figure 3.3 where we show the pairs (α_L, α_U) of lower and upper tail indices of $p = 478$ log-return series composing the S&P 500 index estimated from $n = 1,345$ daily observations from 01/04/2010 to 02/28/2015. This means we assume for every row $(X_{it})_{t=1,\dots,n}$ of $\mathbf{X}$ that the tails behave like

$$\mathbb{P}(X_{it} > x) \sim c_U\, x^{-\alpha_U} \qquad \text{and} \qquad \mathbb{P}(X_{it} < -x) \sim c_L\, x^{-\alpha_L}\,, \qquad x \to \infty\,,$$

for non-negative constants c_L, c_U. We apply the Hill estimator (see Embrechts et al. [35], p. 330, de Haan and Ferreira [26], p. 69) to the time series of the gains and losses in a naive way, neglecting the dependence and non-stationarity in the data; we also omit confidence bands. From the figure it is evident that the majority of the return series have tail indices below four, corresponding to an infinite fourth moment. The behavior

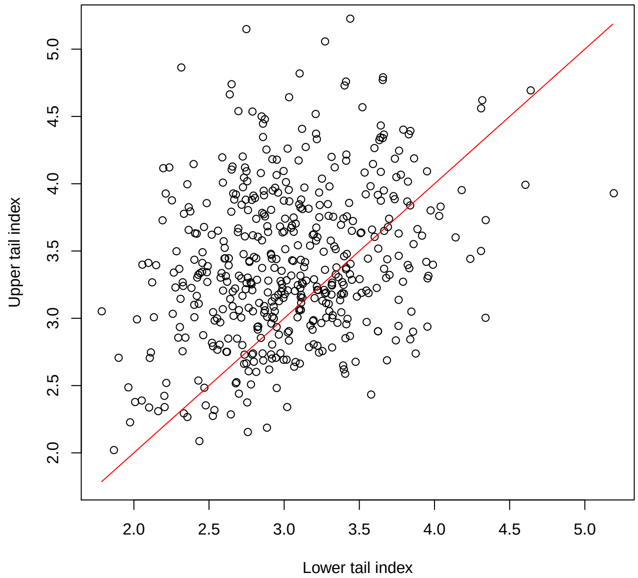


Figure 3.3: Tail indices of log-returns of 478 time series from the S&P 500 index. The values $(\hat{\alpha}_L, \hat{\alpha}_U)$ of the lower and upper tail indices are provided by Hill's estimator. We also draw the line $\hat{\alpha}_U = \hat{\alpha}_L$.

of the largest eigenvalue $\lambda_{(1)}$ changes dramatically when $\mathbf{X}$ has infinite fourth moment. Bai and Silverstein [9] proved for an $n \times n$ matrix $\mathbf{X}$ with iid centered entries that

$$\limsup_{n\to\infty} \frac{\lambda_{(1)}}{n} = \infty \qquad \text{a.s.} \tag{3.6}$$

This is in stark contrast to Geman's result (3.4).

In the heavy-tailed case it is common to assume a *regular variation condition:*

$$\mathbb{P}(X > x) \sim p_+ \frac{L(x)}{x^\alpha} \qquad \text{and} \qquad \mathbb{P}(X < -x) \sim p_- \frac{L(x)}{x^\alpha}\,, \qquad x \to \infty\,, \tag{3.7}$$

where $p_\pm$ are non-negative constants such that $p_+ + p_- = 1$ and L is a slowly varying function. In particular, if $\alpha < 4$ we have $\mathbb{E}[X^4] = \infty$. The regular variation condition on X (we will also refer to X as a regularly varying random variable) is needed for proving asymptotic theory for the eigenvalues of $\mathbf{X}\mathbf{X}'$. This is similar to proving limit theory for sums of iid random variables with infinite variance stable limits; see for example Feller [37].

In (3.2) we have seen that the sequence $(F_{n^{-1}\mathbf{X}\mathbf{X}'})$ of empirical spectral distributions converges to the Marčenko–Pastur law if the centered iid entries possess a finite second moment. Now we will discuss the situation when the entries are still iid and centered, but have an infinite variance. Here we assume the entries to be regularly varying with index $\alpha \in (0, 2)$. Assuming (3.1) with $\gamma \in (0, 1]$ in this infinite variance case, Belinschi et al. [15, Theorem 1.10] showed that the sequence $(F_{a_{n+p}^{-2}\mathbf{XX}'})$ converges with probability one to a non-random probability measure with density ρ_α^γ satisfying

$$\rho_\alpha^\gamma(x)x^{1+\alpha/2} \to \frac{\alpha\gamma}{2(1+\gamma)}, \quad x \to \infty,$$

see also Ben Arous and Guionnet [17, Theorem 1.6]. The normalization (a_k) is chosen such that $\mathbb{P}(|X| > a_k) \sim k^{-1}$ as $k \to \infty$. An application of the Potter bounds (see Bingham et al. [20, p. 25]) shows that $a_{n+p}^2/n \to \infty$.

It is interesting to note that there is a phase change in the extreme eigenvalues in going from finite to infinite fourth moment, while the phase change occurs for the empirical spectral distribution going from finite to infinite variance.

The theory for the largest eigenvalues of sample covariance matrices with heavy tails is less developed than in the light-tailed case. Pioneering work for $\lambda_{(1)}$ in the case of iid regularly varying entries X_{it} with index $\alpha \in (0,2)$ is due to Soshnikov [65, 66]. He showed the point process convergence

$$N_n = \sum_{i=1}^p \varepsilon_{a_{np}^{-2}\lambda_i} \xrightarrow{\mathrm{d}} N = \sum_{i=1}^\infty \varepsilon_{\Gamma_i^{-2/\alpha}}\,, \qquad n \to \infty\,, \tag{3.8}$$

under the growth condition (3.1) on (p_n). Here

$$\Gamma_i = E_1 + \cdots + E_i\,, \qquad i \geq 1\,, \tag{3.9}$$

and (E_i) is an iid standard exponential sequence. In other words, N is a Poisson point process on $(0,\infty)$ with mean measure $\mu(x,\infty) = x^{-\alpha/2}$, $x > 0$. Convergence in distribution of point processes is understood in the sense of weak convergence in the space of point measures equipped with the vague topology; see Resnick [60, 61]. We can easily derive the limiting distribution of $a_{np}^{-2}\lambda_{(k)}$ for fixed $k \geq 1$ from (3.8):

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(a_{np}^{-2}\lambda_{(k)} \leq x) &= \lim_{n\to\infty} \mathbb{P}(N_n(x,\infty) < k) = \mathbb{P}(N(x,\infty) < k) = \mathbb{P}(\Gamma_k^{-2/\alpha} \leq x) \\ &= \sum_{s=0}^{k-1} \frac{\big(\mu(x,\infty)\big)^s}{s!}\, \mathrm{e}^{-\mu(x,\infty)}, \quad x > 0.\end{aligned}$$

In particular,

$$\frac{\lambda_{(1)}}{a_{np}^2} \xrightarrow{\mathrm{d}} \Gamma_1^{-\alpha/2}\,, \qquad n \to \infty\,,$$

where the limit has *Fréchet distribution* with parameter $\alpha/2$ and distribution function

$$\Phi_{\alpha/2}(x) = \mathrm{e}^{\,-x^{-\alpha/2}}\,, \qquad x > 0\,.$$

We mention that the tail balance condition (3.7) may be replaced in this case by the weaker assumption $\mathbb{P}(|X| > x) = L(x)x^{-\alpha}$ for a slowly varying function L. Indeed, it follows from the proof that only the squares X_{it}^2 contribute to the point process limits of the eigenvalues (λ_i). A consequence of the continuous mapping theorem and (3.8) is the joint convergence of the upper order statistics: for any $k \geq 1$,

$$a_{np}^{-2}\big(\lambda_{(1)}, \ldots, \lambda_{(k)}\big) \xrightarrow{\mathrm{d}} \big(\Gamma_1^{-2/\alpha}, \ldots, \Gamma_k^{-2/\alpha}\big)\,, \qquad n \to \infty\,.$$

It follows from standard theory for point processes with iid points (e.g. Resnick [60, 61]) that (3.8) remains valid if we replace N_n by the point process $\sum_{i=1}^p \sum_{t=1}^n \varepsilon_{X_{it}^2/a_{np}^2}$. Then we also have for any $k \geq 1$,

$$a_{np}^{-2}\big(X^2_{(1),np}, \dots, X^2_{(k),np}\big) \xrightarrow{\mathrm{d}} \big(\Gamma_1^{-2/\alpha}, \dots, \Gamma_k^{-2/\alpha}\big)\,, \qquad n \to \infty\,, \tag{3.10}$$

where

$$X^2_{(1),np} \geq \cdots \geq X^2_{(np),np}$$

denote the order statistics of $(X_{it}^2)_{i=1,\dots,p;t=1,\dots,n}$.

Auffinger et al. [4] showed that (3.8) remains valid under the regular variation condition (3.7) for $\alpha \in (2,4)$, the growth condition (3.1) on (p_n) and the additional assumption $\mathbb{E}[X] = 0$. Of course, (3.10) remains valid as well. Davis et al. [25] extended these results to the case when the rows of $\mathbf{X}$ are iid linear processes with iid regularly varying noise. The Poisson point process convergence result of (3.8) remains valid in this case. Different limit processes can only be expected if there is dependence across rows and columns.

In what follows, we refer to the *heavy-tailed case* when we assume the regular variation condition (3.7) for some $\alpha \in (0,4)$.

3.1.3 Overview

The primary objective of this overview is to make a connection between extreme value theory and the behavior of the largest eigenvalues of sample covariance matrices from heavy-tailed multivariate time series. For time series that are linearly dependent through time and across rows, it turns out that the extreme eigenvalues are essentially determined by the extreme order statistics from an array of iid random variables. The asymptotic behavior of the extreme eigenvalues is then derived routinely from classical extreme value theory. As such, explicit joint distributions of the extreme order statistics can be given which yield a plethora of ancillary results. Convergence of the point process of extreme eigenvalues, properly normalized, plays a central role in establishing the main results.

In Section 3.2 we continue the study of the case when the data matrix $\mathbf{X}$ consists of iid heavy-tailed entries. We will consider power-law growth rates on the dimension (p_n) that is more general than prescribed by (3.1). In Section 3.3 we introduce a model for X_{it} which allows for linear dependence across the rows and through time. The point process convergence of normalized eigenvalues is presented in Section 3.3.4. This result lays the foundation for new insight into the spectral behavior of the sample covariance matrix, which is the content of Section 3.4.1.

Sections 3.4.1 and 3.4.3 are devoted to *sample autocovariance matrices.* Motivated by [49], we study the eigenvalues of sums of transformed matrices and illustrate the results in two examples. These results are applied to the time series of S&P 500 in Section 3.4.2. Appendix 3.5 contains useful facts about regular variation and point processes.

3.2 General growth rates for p_n in the iid heavy-tailed case

This section is based on ideas in Heiny and Mikosch [41] where one can also find detailed proofs.

Growth conditions on (p_n)

In many applications it is not realistic to assume that the dimension p of the data and the sample size n grow at the same rate. The aforementioned results of Soshnikov [65, 66] and Auffinger et al. [4] already indicate that the value γ in the growth rate (3.1) does not appear in the distributional limits. This obervation is in contrast to the light-tailed case; see (3.4) and (3.5). Davis et al. [24, 25] and Heiny and Mikosch [41] allowed for more general rates for $p_n \to \infty$ than linear growth in n. Recall that $p = p_n \to \infty$ is the number of rows in the matrix $\mathbf{X}_n$. We need to specify the growth rate of (p_n) to ensure a non-degenerate limit distribution of the normalized singular values of the sample autocovariance matrices. To be precise, we assume

$$p = p_n = n^\beta \ell(n), \quad n \geq 1, \tag{$C_p(\beta)$}$$

where ℓ is a slowly varying function and $\beta \geq 0$. If $\beta = 0$, we also assume $\ell(n) \to \infty$. Condition $C_p(\beta)$ is more general than the growth conditions in the literature; see [4, 24, 25].

Theorem 3.1. *Assume that $\mathbf{X} = \mathbf{X}_n$ has iid entries satisfying the regular variation condition* (3.7) *for some $\alpha \in (0,4)$. If $\mathbb{E}[|X|] < \infty$ we also suppose that $\mathbb{E}[X] = 0$. Let (p_n) be an integer sequence satisfying $C_p(\beta)$ with $\beta \geq 0$. In addition, we require*

$$\min(\beta, \beta^{-1}) \in (\alpha/2 - 1, 1] \qquad \text{for } \alpha \in [2,4), \tag{$\widetilde{C}_\beta(\alpha)$}$$

Then

$$\sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}\lambda_i} \xrightarrow{\text{d}} \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}}\,, \qquad n \to \infty\,, \tag{3.11}$$

where the convergence holds in the space of point measures with state space $(0,\infty)$ equipped with the vague toplogy; see Resnick [60].

A discussion of the case $\beta \in [0,1]$

We mentioned earlier that in the heavy-tailed case, limit theory for the largest eigenvalues of the sample covariance matrix is rather insensitive to the growth rate of (p_n) and that the limits are essentially determined by the diagonal of this matrix. This is confirmed by the following result.

Proposition 3.2. *Assume that $\mathbf{X} = \mathbf{X}_n$ has iid entries satisfying the regular variation condition* (3.7) *for some $\alpha \in (0,4)$. If $\mathbb{E}[|X|] < \infty$ we also suppose that $\mathbb{E}[X] = 0$. Then for any sequence (p_n) satisfying $C_p(\beta)$ with $\beta \in [0,1]$ we have*

$$a_{np}^{-2} \|\mathbf{X}\mathbf{X}' - \operatorname{diag}(\mathbf{X}\mathbf{X}')\|_2 \xrightarrow{\mathbb{P}} 0\,, \qquad n \to \infty\,,$$

where $\|\cdot\|_2$ denotes the spectral norm; see (3.23) *for its definition.*

Proposition 3.2 is not unexpected for two reasons:

- It is well-known from classical theory (see Embrechts and Veraverbeke [36]) that for any iid regularly varying non-negative random variables Y, Y' with index $\alpha' > 0$, YY' is regularly varying with index α' while Y^2 is regularly varying with index $\alpha'/2$. Therefore X^2 and $X_{11}X_{12}$ are regularly varying with indices $\alpha/2$ and α, respectively.

- The aforementioned tail behavior is inherited by the entries of $\mathbf{XX}'$ in the following sense. By virtue of Nagaev-type large deviation results for an iid regularly varying sequence (Y_i) with index $\alpha' \in (0,2)$ where we also assume that $\mathbb{E}[Y_0]=0$ if $\mathbb{E}[|Y_0|]<\infty$ (see Theorem 3.21) we have that $\mathbb{P}(Y_1+\cdots+Y_n>b_n)/(n\,\mathbb{P}(|Y_0|>b_n))$ converges to a non-negative constant provided $b_n/a'_n\to\infty$, where $\mathbb{P}(|Y_0|>a'_n)\sim n^{-1}$ as $n\to\infty$. As a consequence of the tail behaviors of X_{it}^2 and $X_{it}X_{jt}$ for $i\neq j$ and Nagaev's results we have for (b_n) such that $b_n/a_n^2\to\infty$,

$$\frac{\mathbb{P}\big((\mathbf{XX}')_{ij}>b_n\big)}{\mathbb{P}\big((\mathbf{XX}')_{ii}-c_n>b_n\big)}\sim\frac{n\,\mathbb{P}(X_{11}X_{12}>b_n)}{n\,\mathbb{P}(X^2>b_n)}\to 0\,,\qquad n\to\infty\,,\tag{3.12}$$

  where $c_n=0$ or $n\,\mathbb{E}[X^2]$ according as $\alpha\in(0,2)$ or $\alpha\in(2,4)$. This means that the diagonal and off-diagonal entries of $\mathbf{XX}'$ inherit the tails of X_{it}^2 and $X_{it}X_{jt}$, $i\neq j$, respectively, above the high threshold b_n.

Proposition 3.2 has some immediate consequences for the approximation of the eigenvalues of $\mathbf{XX}'$ by those of diag$(\mathbf{XX}')$. Indeed, let C be a symmetric $p\times p$ matrix with eigenvalues $\lambda_1(C),\ldots,\lambda_p(C)$ and ordered eigenvalues

$$\lambda_{(1)}(C)\geq\cdots\geq\lambda_{(p)}(C)\,.\tag{3.13}$$

Then for any symmetric $p\times p$ matrices A,B, by *Weyl's inequality* (see Bhatia [19]),

$$\max_{i=1,\ldots,p}\big|\lambda_{(i)}(A+B)-\lambda_{(i)}(A)\big|\leq\|B\|_2\,.$$

If we now choose $A+B=\mathbf{XX}'$ and $A=\text{diag}(\mathbf{XX}')$ we obtain the following result.

Corollary 3.3. *Under the conditions of Proposition 3.2,*

$$a_{np}^{-2}\max_{i=1,\ldots,p}\big|\lambda_{(i)}-\lambda_{(i)}(\text{diag}(\mathbf{XX}'))\big|\xrightarrow{\mathbb{P}}0\,,\quad n\to\infty\,.$$

Thus the problem of deriving limit theory for the order statistics of $\mathbf{XX}'$ has been reduced to limit theory for the order statistics of the iid row-sums

$$D_i^{\rightarrow}=(\mathbf{XX}')_{ii}=\sum_{t=1}^n X_{it}^2\,,\qquad i=1,\ldots,p\,,$$

which are the eigenvalues of diag$(\mathbf{XX}')$. This theory is completely described by the point processes constructed from the points $D_i^{\rightarrow}/a_{np}^2$ $i=1,\ldots,p$. Necessary and sufficient conditions for the weak convergence of these point processes are provided by Lemma 3.22 which in combination with the Nagaev-type large deviation results of Theorem 3.21 yield the following result; see also Davis et al. [24].

Lemma 3.4. *Assume the conditions of Proposition 3.2 hold. Then*

$$\sum_{i=1}^p\varepsilon_{a_{np}^{-2}(D_i^{\rightarrow}-c_n)}\xrightarrow{\text{d}}\sum_{i=1}^\infty\varepsilon_{\Gamma_i^{-2/\alpha}}\,,\qquad n\to\infty\,,$$

where (Γ_i) *is defined in* (3.9) *and* $c_n=0$ *if* $\mathbb{E}[D^{\rightarrow}]=\infty$ *and* $c_n=\mathbb{E}[D^{\rightarrow}]$ *otherwise.*

In this result, centering is only needed for $\alpha\in[2,4)$ when $n/a_{np}^2\not\to 0$. Under the additional condition $\widetilde{C}_\beta(\alpha)$, $n/a_{np}^2\to 0$ in view of the Potter bounds; see Bingham et al. [20, p. 25]. Combining Lemma 3.4 and Corollary 3.3, we conclude that Theorem 3.1 holds for $\beta\in[0,1]$.

Extension to general β

Next we explain that it suffices to consider only the case $\beta \in [0,1]$ and how to proceed when $\beta > 1$. The main reason is that the $p \times p$ sample covariance matrix $\mathbf{X}\mathbf{X}'$ and the $n \times n$ matrix $\mathbf{X}'\mathbf{X}$ have the same rank and their non-zero eigenvalues coincide; see Bhatia [19, p. 64]. When proving limit theory for the eigenvalues of the sample covariance matrix one may switch to $\mathbf{X}'\mathbf{X}$ and vice versa, hereby interchanging the roles of p and n. By switching to $\mathbf{X}'\mathbf{X}$, one basically replaces β by β^{-1}. Since $\min(\beta, \beta^{-1}) \in [0,1]$ for any $\beta \geq 0$, one can assume without loss of generality that $\beta \in [0,1]$. This trick allows one to extend results for (p_n) satisfying $C_p(\beta)$ with $\beta \in [0,1]$ to $\beta > 1$. We illustrate this approach by providing the direct analogs of Proposition 3.2 and Corollary 3.3.

Proposition 3.5. *Assume that $\mathbf{X} = \mathbf{X}_n$ has iid entries satisfying the regular variation condition* (3.7) *for some $\alpha \in (0,4)$. If $\mathbb{E}[|X|] < \infty$ we also suppose that $\mathbb{E}[X] = 0$. Then for any sequence (p_n) satisfying $C_p(\beta)$ with $\beta > 1$ we have*

$$a_{np}^{-2} \|\mathbf{X}'\mathbf{X} - \operatorname{diag}(\mathbf{X}'\mathbf{X})\|_2 \xrightarrow{\mathbb{P}} 0\,, \qquad n \to \infty\,,$$

where $\|\cdot\|_2$ denotes the spectral norm.

Note that for $\beta > 1$ we have $\lim_{n\to\infty} p/n = \infty$. This means that $\mathbf{X}'\mathbf{X}$ has asymptotically a much smaller dimension than $\mathbf{X}\mathbf{X}'$ and therefore it is more convenient to work with $\mathbf{X}'\mathbf{X}$ when bounding the spectral norm.

Corollary 3.6. *Under the conditions of Proposition 3.5,*

$$a_{np}^{-2} \max_{i=1,\ldots,n} \big|\lambda_{(i)} - \lambda_{(i)}(\operatorname{diag}(\mathbf{X}'\mathbf{X}))\big| \xrightarrow{\mathbb{P}} 0\,, \quad n \to \infty\,.$$

Now, Theorem 3.1 for $\beta > 1$ is a consequence of Corollary 3.6.

3.3 Introducing dependence between the rows and columns

For details on the results of this section, we refer to Davis et al. [24], Heiny and Mikosch [41] and Heiny et al. [42].

3.3.1 The model

When dealing with covariance matrices of a multivariate time series $(\mathbf{X}_n)$ it is rather natural to assume dependence between the entries X_{it}. In this section we introduce a model which allows for *linear dependence* between the rows and columns of $\mathbf{X}$:

$$X_{it} = \sum_{l\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} h_{kl} Z_{i-k,t-l}\,, \qquad i,t \in \mathbb{Z}\,, \tag{3.14}$$

where $(Z_{it})_{i,t\in\mathbb{Z}}$ is a field of iid random variables and $(h_{kl})_{k,l\in\mathbb{Z}}$ is an array of real numbers. Of course, linear dependence is restrictive in some sense. However, the particular dependence structure allows one to determine those ingredients in the sample covariance matrix which contribute to its largest eigenvalues. If the series in (3.14) converges a.s. (X_{it}) constitutes a strictly stationary random field. We denote generic elements of the Z- and X-fields by Z and X, respectively. We assume that Z is regularly varying in the sense that

$$\mathbb{P}(Z > x) \sim p_+ \frac{L(x)}{x^\alpha} \quad \text{and} \quad \mathbb{P}(Z \leq -x) \sim p_- \frac{L(x)}{x^\alpha}\,, \qquad x \to \infty\,, \tag{3.15}$$

for some tail index $\alpha > 0$, constants $p_+, p_- \geq 0$ with $p_+ + p_- = 1$ and a slowly varying L. We will assume $\mathbb{E}[Z] = 0$ whenever $\mathbb{E}[Z^2] < \infty$. Moreover, we require the summability condition

$$\sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} |h_{kl}|^\delta < \infty \tag{3.16}$$

for some $\delta \in (0, \min(\alpha/2, 1))$ which ensures the a.s. absolute convergence of the series in (3.14). Under the conditions (3.15) and (3.16), the marginal and finite-dimensional distributions of the field (X_{it}) are regularly varying with index α; see Embrechts et al. [35], Appendix A3.3. Therefore we also refer to (X_{it}) and (Z_{it}) as regularly varying fields.

The model (3.14) was introduced by Davis et al. [25], assuming the rows iid, and in the present form by Davis et al. [24].

3.3.2 Sample covariance and autocovariance matrices

From the field (X_{it}) we construct the $p \times n$ matrices

$$\mathbf{X}_n(s) = (X_{i,t+s})_{i=1,\ldots,p;t=1,\ldots,n}\,, \qquad s = 0, 1, 2, \ldots\,, \tag{3.17}$$

As before, we will write $\mathbf{X} = \mathbf{X}_n(0)$. Now we can introduce the (non-normalized) *sample autocovariance matrices*

$$\mathbf{X}_n(0)\mathbf{X}_n(s)'\,, \qquad s = 0, 1, 2, \ldots\,. \tag{3.18}$$

We will refer to s as the *lag.* For $s = 0$, we obtain the *sample covariance matrix.* In what follows, we will be interested in the asymptotic behavior (of functions) of the eigen- and singular values of the sample covariance and autocovariance matrices in the heavy-tailed case. Recall that the *singular values* of a matrix A are the square roots of the eigenvalues of the non-negative definite matrix AA' and its *spectral norm* $\|A\|_2$ is its largest singular value. We notice that $\mathbf{X}_n(0)\mathbf{X}_n(s)'$ is not symmetric and therefore its eigenvalues can be complex. To avoid this situation, we use the squares

$$\mathbf{X}_n(0)\mathbf{X}_n(s)'\mathbf{X}_n(s)\mathbf{X}_n(0)' \tag{3.19}$$

whose eigenvalues are the squares of the singular values of $\mathbf{X}_n(0)\mathbf{X}_n(s)'$. The idea of using the sample autocovariance matrices and functions of their squares (3.19) originates from a paper by Lam and Yao [49] who used a model different from (3.14). This idea is quite natural in the context of time series analysis.

In Theorem 3.7 below, we provide a general approximation result for the ordered singular values of the sample autocovariance matrices in the heavy-tailed case. This result is rather technical. To formulate it we introduce further notation. As before, $p = p_n$ is any integer sequence converging to infinity.

3.3.3 More notation

Important roles are played by the quantities $(Z_{it}^2)_{i=1,\ldots,p;t=1,\ldots,n}$ and their order statistics

$$Z^2_{(1),np} \geq Z^2_{(2),np} \geq \ldots \geq Z^2_{(np),np}, \qquad n, p \geq 1\,. \tag{3.20}$$

As important are the row-sums

$$D_i^{\rightarrow} = D_i^{(n),\rightarrow} = \sum_{t=1}^n Z_{it}^2\,, \qquad i = 1, \ldots, p; \quad n = 1, 2, \ldots\,, \tag{3.21}$$

with generic element $D^{\rightarrow}$ and their ordered values

$$D^{\rightarrow}_{(1)} = D^{\rightarrow}_{L_1} \geq \cdots \geq D^{\rightarrow}_{(p)} = D^{\rightarrow}_{L_p}\,, \tag{3.22}$$

where we assume without loss of generality that $(L_1, \ldots, L_p)$ is a permutation of $(1, \ldots, p)$ for fixed n.

Finally, we introduce the column-sums

$$D^{\downarrow}_t = D^{(n),\downarrow}_t = \sum_{i=1}^{p} Z_{it}^2\,, \qquad t = 1, \ldots, n; \quad p = 1, 2, \ldots\,,$$

with generic element $D^{\downarrow}$ and we also adapt the notation from (3.22) to these quantities.

Matrix norms

For any $p \times n$ matrix $\mathbf{A} = (a_{ij})$, we will use the following norms:

- *Spectral norm:*

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{(1)}(\mathbf{A}\mathbf{A}')}\,, \tag{3.23}$$

- *Frobenius norm:*

$$\|\mathbf{A}\|_F = \Big(\sum_{i=1}^{p}\sum_{j=1}^{n} |a_{ij}|^2\Big)^{1/2}.$$

We will frequently make use of the bound $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F$. Standard references for matrix norms are [16, 19, 43, 63].

Singular values of the sample autocovariance matrices

Fix integers $n \geq 1$ and $s \geq 0$. We recycle the λ-notation for the singular values $\lambda_1(s), \ldots, \lambda_p(s)$ of the sample autocovariance matrix $\mathbf{X}_n(0)\mathbf{X}_n(s)'$, suppressing the dependence on n. Correspondingly, the order statistics are denoted by

$$\lambda_{(1)}(s) \geq \cdots \geq \lambda_{(p)}(s)\,. \tag{3.24}$$

When $s = 0$ we typically write λ_i instead of $\lambda_i(0)$.

The matrix $\mathbf{M}(s)$

We introduce some auxiliary matrices derived from the coefficient matrix $\mathbf{H} = (h_{kl})_{k,l\in\mathbb{Z}}$:

$$\mathbf{H}(s) = (h_{k,l+s})_{k,l\in\mathbb{Z}}, \qquad \mathbf{M}(s) = \mathbf{H}(0)\mathbf{H}(s)' \qquad s \geq 0\,.$$

Notice that

$$(\mathbf{M}(s))_{ij} = \sum_{l\in\mathbb{Z}} h_{i,l} h_{j,l+s}, \qquad i, j \in \mathbb{Z}. \tag{3.25}$$

We denote the ordered singular values of $\mathbf{M}(s)$ by

$$v_1(s) \geq v_2(s) \geq \cdots\,. \tag{3.26}$$

Let $r(s)$ be the rank of $\mathbf{M}(s)$ so that $v_{r(s)}(s) > 0$ while $v_{r(s)+1}(s) = 0$ if $r(s)$ is finite, otherwise $v_i(s) > 0$ for all i. We also write $r = r(0)$.

Under the summability condition (3.16) on (h_{kl}) for fixed $s \geq 0$,

$$\begin{aligned} \sum_{i=1}^{\infty}(v_i(s))^2 &= \|\mathbf{M}(s)\|_F^2 = \sum_{i,j\in\mathbb{Z}}\sum_{l_1,l_2\in\mathbb{Z}} h_{i,l_1}h_{j,l_1+s}h_{i,l_2}h_{j,l_2+s} \\ &\leq c\Big(\sum_{l_1,l_2\in\mathbb{Z}}\sum_{i\in\mathbb{Z}}|h_{i,l_1}h_{i,l_2}|\Big)^2 \leq c\sum_{l_1\in\mathbb{Z}}\sum_{i\in\mathbb{Z}}|h_{i,l_1}| < \infty\,. \end{aligned} \tag{3.27}$$

Therefore all singular values $v_i(s)$ are finite and the ordering (3.26) is justified.

Here and in what follows, we write c for any constant whose value is not of interest.

Normalizing sequence

We define (a_k) by

$$\mathbb{P}(|Z| > a_k) \sim k^{-1}\,, \qquad k \to \infty\,,$$

and choose the normalizing sequence for the singular values as (a_{np}^2) for suitable sequences $p = p_n \to \infty$.

Approximations to singular values

We will give approximations to the singular values $\lambda_i(s)$ in terms of the p largest ordered values for $s \geq 0$,

$$\begin{aligned} &\delta_{(1)}(s) \geq \cdots \geq \delta_{(p)}(s)\,, \\ &\gamma_{(1)}^{\rightarrow}(s) \geq \cdots \geq \gamma_{(p)}^{\rightarrow}(s)\,, \\ &\gamma_{(1)}^{\downarrow}(s) \geq \cdots \geq \gamma_{(n)}^{\downarrow}(s)\,, \end{aligned}$$

from the sets

$$\begin{aligned} &\big\{Z_{(i),np}^2 v_j(s)\,, i = 1,\ldots,p\,; j = 1,2,\ldots\big\}\,, \\ &\big\{D_i^{\rightarrow} v_j(s), i = 1,\ldots,p\,; j = 1,2,\ldots\big\}\,, \\ &\big\{D_t^{\downarrow} v_j(s), t = 1,\ldots,n\,; j = 1,2,\ldots\big\}\,, \end{aligned}$$

respectively.

3.3.4 Approximation of the singular values

In the following result we povide some useful approximations to the singular values of the sample autocovariance matrices of the linear model (3.14).

Theorem 3.7. *Consider the linear process* (3.14) *under*

- *the regular variation condition* (3.15) *for some* $\alpha \in (0,4)$,
- *the centering condition* $\mathbb{E}[Z] = 0$ *if* $\mathbb{E}[|Z|] < \infty$,
- *the summability condition* (3.16) *on the coefficient matrix* (h_{kl}),

- *the growth condition $C_p(\beta)$ on (p_n) for some $\beta \ge 0$.*

Then the following statements hold for $s \ge 0$:

1. *We consider two disjoint cases: $\alpha \in (0,2)$ and $\beta \in (0,\infty)$, or $\alpha \in [2,4)$ and β satisfying $\widetilde{C}_\beta(\alpha)$. Then*

$$a_{np}^{-2} \max_{i=1,\ldots,p} |\lambda_{(i)}(s) - \delta_{(i)}(s)| \stackrel{\mathbb{P}}{\to} 0, \quad n \to \infty. \tag{3.28}$$

2. *Assume $\beta \in [0,1]$. If $\alpha \in (0,2]$, $\mathbb{E}[Z^2] = \infty$ or $\alpha \in [2,4)$, $\mathbb{E}[Z^2] < \infty$ and $\beta \in (\alpha/2-1, 1]$ then*

$$a_{np}^{-2} \max_{i=1,\ldots,p} |\lambda_{(i)}(s) - \gamma^{\rightarrow}_{(i)}(s)| \stackrel{\mathbb{P}}{\to} 0, \quad n \to \infty.$$

Assume $\beta > 1$. If $\alpha \in (0,2]$, $\mathbb{E}[Z^2] = \infty$ or $\alpha \in [2,4)$, $\mathbb{E}[Z^2] < \infty$ and $\beta^{-1} \in (\alpha/2-1, 1]$. Then

$$a_{np}^{-2} \max_{i=1,\ldots,p} |\lambda_{(i)}(s) - \gamma^{\downarrow}_{(i)}(s)| \stackrel{\mathbb{P}}{\to} 0, \quad n \to \infty.$$

Remark 3.8. The proof of Theorem 3.7 is given in Heiny et al. [42]. Part (2) of this result with more restrictive conditions on the growth rate of (p_n) is contained in Davis et al. [24]. These proofs are very technical and lengthy.

Remark 3.9. If we consider a random array (h_{kl}) independent of (X_{it}) and assume that the summability condition (3.16) holds a.s. then Theorem 3.7 remains valid conditionally on (h_{kl}), hence unconditionally in $\mathbb{P}$-probability; see also [24].

3.3.5 Point process convergence

Theorem 3.7 and arguments similar to the proofs in Davis et al. [24] enable one to derive the weak convergence of the point processes of the normalized singular values. Recall the representation of the points (Γ_i) of a unit rate homogeneous Poisson process on $(0,\infty)$ given in (3.9). For $s \ge 0$, we define the point processes of the normalized singular values:

$$N_n^{\lambda,s} = \sum_{i=1}^{p} \varepsilon_{a_{np}^{-2}(\lambda_{(i)}(0),\ldots,\lambda_{(i)}(s))} . \tag{3.29}$$

Theorem 3.10. *Assume the conditions of Theorem 3.7. Then $(N_n^{\lambda,s})$ converge weakly in the space of point measures with state space $(0,\infty)^{s+1}$ equipped with the vague topology. If either $\alpha \in (0,2]$, $\mathbb{E}[Z^2] = \infty$ and $\beta \ge 0$, or $\alpha \in [2,4)$, $\mathbb{E}[Z^2] < \infty$ and $\widetilde{C}_\beta(\alpha)$ hold then*

$$N_n^{\lambda,s} \stackrel{\mathrm{d}}{\to} N = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \varepsilon_{\Gamma_i^{-2/\alpha}(v_j(0),\ldots,v_j(s))}, \qquad n \to \infty. \tag{3.30}$$

Proof. Regular variation of Z^2 is equivalent to

$$n\,p\,\mathbb{P}(a_{np}^{-2} Z^2 \in \cdot) \stackrel{\mathrm{v}}{\to} \mu(\cdot), \tag{3.31}$$

where $\stackrel{v}{\to}$ denotes vague convergence of Radon measures on $(0,\infty)$ and the measure μ is given by $\mu(x,\infty)=x^{-\alpha/2}$, $x>0$. In view of Resnick [61], Proposition 3.21, (3.31) is equivalent to the weak convergence of the following point processes:

$$\sum_{i=1}^{p}\sum_{t=1}^{n}\varepsilon_{a_{np}^{-2}Z_{it}^2}=\sum_{i=1}^{np}\varepsilon_{a_{np}^{-2}Z_{(i),np}^2}\stackrel{d}{\to}\sum_{i=1}^{\infty}\varepsilon_{\Gamma_i^{-2/\alpha}}=\widetilde{N}\,,\quad n\to\infty\,,$$

where the limit $\widetilde{N}$ is a Poisson random measure (PRM) with state space $(0,\infty)$ and mean measure μ.

Since $a_{np}^{-2}Z_{(p),np}^2\stackrel{\mathbb{P}}{\to}0$ as $n\to\infty$, the point processes $\sum_{i=1}^{p}\varepsilon_{a_{np}^{-2}Z_{(i),np}^2}$ converge weakly to the same PRM:

$$\sum_{i=1}^{p}\varepsilon_{a_{np}^{-2}Z_{(i),np}^2}\stackrel{d}{\to}\sum_{i=1}^{\infty}\varepsilon_{\Gamma_i^{-2/\alpha}}\,,\quad n\to\infty\,. \tag{3.32}$$

A continuous mapping argument together with the fact that $\sum_{i=1}^{\infty}(v_i(s))^2<\infty$ (see (3.27)) shows that

$$\sum_{j=1}^{\infty}\sum_{i=1}^{p}\varepsilon_{a_{np}^{-2}Z_{(i),np}^2(v_j(0),\dots,v_j(s))}\stackrel{d}{\to}\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}\varepsilon_{\Gamma_i^{-2/\alpha}(v_j(0),\dots,v_j(s))}\,.$$

If the assumptions of part (1) of Theorem 3.7 are satisfied an application of (3.28) (also recalling the definition of $(\delta_{(i)}(s))$) shows that (3.32) remains valid with the points $(a_{np}^{-2}Z_{(i),np}^2(v_j(0),\dots,v_j(s)))$ replaced by $(a_{np}^{-2}(\lambda_{(i)}(0),\dots,\lambda_{(i)}(s))$.

The only cases which are not covered by Theorem 3.7(1) are $\alpha\in(0,2)$, $\beta=0$ and $\alpha=2$, $\mathbb{E}[Z^2]=\infty$, $\beta\ge 0$. In these cases we get from Theorem 3.21 that

$$p\,\mathbb{P}(a_{np}^{-2}D^{\to}>x)\sim p\,n\,\mathbb{P}(Z^2>a_{np}^2x)\to\mu(x,\infty)\,,\quad x>0\,,$$

i.e., $p\,\mathbb{P}(a_{np}^{-2}D^{\to}\in\cdot)\stackrel{v}{\to}\mu(\cdot)$. It follows from Lemma 3.22 that $\sum_{i=1}^{p}\varepsilon_{a_{np}^{-2}D_i^{\to}}\stackrel{d}{\to}\widetilde{N}$. As before, a continuous mapping argument in combination with the approximation obtained in Theorem 3.7(2) justifies the replacement of the points $(a_{np}^{-2}D_{(i)}^{\to}(v_j(0),\dots,v_j(s)))$ by $(a_{np}^{-2}(\lambda_{(i)}(0),\dots,\lambda_{(i)}(s)))$ in the case $\beta\in[0,1]$. If $\beta>1$ one has to work with the quantities $(D_i^{\downarrow})_{i=1,\dots,n}$ instead of $(D_i^{\to})_{i=1,\dots,p}$ and one may follow the same argument as above. This finishes the proof. □

3.4 Some applications

3.4.1 Sample covariance matrices

The sample covariance matrix $\mathbf{X}_n(0)\mathbf{X}_n(0)'=\mathbf{X}\mathbf{X}'$ is a non-negative definite matrix. Therefore its eigenvalues and singular values coincide. Moreover, $v_j=v_j(0)$, $j\ge 1$, are the eigenvalues of $\mathbf{M}=\mathbf{M}(0)$.

Theorem 3.7(1) yields an approximation of the ordered eigenvalues $(\lambda_{(i)})$ of $\mathbf{X}\mathbf{X}'$ by the quantities $(\delta_{(i)})$ which are derived from the order statistics of (Z_{it}^2). Part (2) of this result provides an approximation of $(\lambda_{(i)})$ by the quantities $(\gamma_{(i)}^{\to/\downarrow})$ which are derived from the order statistics of the partial sums $(D_i^{\to/\downarrow})$.

In the following example we illustrate the quality of the two approximations.

Example 3.11. We choose a Pareto-type distribution for Z with density

$$f_Z(x) = \begin{cases} \frac{\alpha}{(4|x|)^{\alpha+1}}\,, & \text{if } |x| > 1/4 \\ 1\,, & \text{otherwise.} \end{cases} \tag{3.33}$$

We simulated $20,000$ matrices $\mathbf{X}_n$ for $n = 1,000$ and $p = 200$ whose iid entries have this density. We assume $\beta = 1$. Note that $\mathbf{M} = \mathbf{M}(0)$ has rank one and $v_1 = 1$. The estimated densities of the deviations $a_{np}^{-2}(\lambda_{(1)} - D^{\rightarrow}_{(1)})$ and $a_{np}^{-2}(\lambda_{(1)} - Z^2_{(1),np})$ based on the simulations are shown in Figure 3.4. The approximation error is very small indeed. According to the theory,

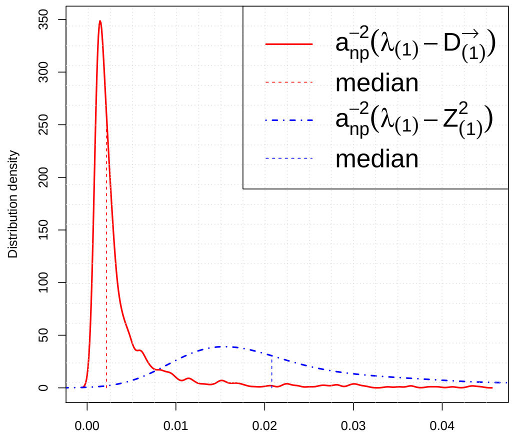


Figure 3.4: Density of the approximation errors for the eigenvalues of $a_{np}^{-2}\mathbf{X}\mathbf{X}'$. The entries of $\mathbf{X}$ are iid with density (3.33) and $\alpha = 1.6$.

$$a_{np}^{-2} \sup_i |D^{\rightarrow}_{(i)} - \lambda_{(i)}| + a_{np}^{-2} \sup_i |Z^2_{(i),np} - \lambda_{(i)}| \xrightarrow{\mathbb{P}} 0\,,$$

but for finite n the $(D^{\rightarrow}_{(i)})$ sequence yields a better approximation to $(\lambda_{(i)})$. By construction, the considered differences have a tendency to be positive but Figure 3.4 also shows that the median of the approximation error for $a_{np}^{-2}(\lambda_{(1)} - D^{\rightarrow}_{(1)})$ is almost zero.

Theorem 3.10 and the continuous mapping theorem immediately yield results about the joint convergence of the largest eigenvalues of the matrices $a_{np}^{-2}\mathbf{X}_n\mathbf{X}_n'$ for $\alpha \in (0,2)$ and $\alpha \in (2,4)$ when β satisfies $\widetilde{C}_\beta(\alpha)$. For fixed $k \geq 1$ one gets

$$a_{np}^{-2}\big(\lambda_{(1)}, \ldots, \lambda_{(k)}\big) \xrightarrow{\mathrm{d}} \big(d_{(1)}, \ldots, d_{(k)}\big)\,,$$

where $d_{(1)} \geq \cdots \geq d_{(k)}$ are the k largest ordered values of the set $\{\Gamma_i^{-2/\alpha} v_j, i = 1, 2, \ldots, j = 1, \ldots, r\}$. The continuous mapping theorem yields for $k \geq 1$,

$$\frac{\lambda_{(1)}}{\lambda_{(1)} + \cdots + \lambda_{(k)}} \xrightarrow{\mathrm{d}} \frac{d_{(1)}}{d_{(1)} + \cdots + d_{(k)}}\,, \quad n \to \infty\,. \tag{3.34}$$

An application of the continuous mapping theorem to the distributional convergence of the point processes in Theorem 3.10 in the spirit of Resnick [60], Theorem 7.1, also yields the following result; see Davis et al. [24] for a proof and a similar result in the case $\alpha \in (2,4)$.

Corollary 3.12. *Assume the conditions of Theorem 3.7. If $\alpha \in (0,2]$ and $\mathbb{E}[Z^2]=\infty$, then*

$$a_{np}^{-2}\Big(\lambda_{(1)}, \sum_{i=1}^{p} \lambda_i\Big) \xrightarrow{\rm d} \Big(v_1\,\Gamma_1^{-2/\alpha}, \sum_{j=1}^{r} v_j \sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}\Big)\,,$$

where $\Gamma_1^{-2/\alpha}$ is Fréchet $\Phi_{\alpha/2}$-distributed. and $\sum_{i=1}^{\infty}\Gamma_i^{-2/\alpha}$ has the distribution of a positive $\alpha/2$-stable random variable. In particular,

$$\frac{\lambda_{(1)}}{\lambda_1+\cdots+\lambda_p} \xrightarrow{\rm d} \frac{v_1}{\sum_{j=1}^{r} v_j}\,\frac{\Gamma_1^{-2/\alpha}}{\sum_{i=1}^{\infty}\Gamma_i^{-2/\alpha}}\,,\qquad n\to\infty\,. \tag{3.35}$$

Remark 3.13. The ratio

$$\frac{\lambda_{(1)}+\cdots+\lambda_{(k)}}{\lambda_1+\cdots+\lambda_p},\qquad k\ge 1\,,$$

plays an important role in PCA. It reflects the proportion of the total variance in the data that we can explain by the first k principal components. It follows from Corollary 3.12 that for fixed $k\ge 1$,

$$\frac{\lambda_{(1)}+\cdots\lambda_{(k)}}{\lambda_1+\cdots+\lambda_p} \xrightarrow{\rm d} \frac{d_{(1)}+\cdots+d_{(k)}}{d_{(1)}+d_{(2)}+\cdots}\,.$$

Unfortunately, the limiting variable does in general not have a clean form. An exception is the case when $r=1$; see Example 3.16. Also notice that the trace of $\mathbf{XX}'$ coincides with $\lambda_1+\cdots+\lambda_p$.

To illustrate the theory we consider a simple moving average example taken from Davis et al. [24].

Example 3.14. Assume that $\alpha\in(0,2)$ and

$$X_{it} = Z_{it} + Z_{i,t-1} - 2(Z_{i-1,t} - Z_{i-1,t-1})\,,\quad i,t\in\mathbb{Z}\,. \tag{3.36}$$

In this case, the non-zero entries of $\mathbf{H}$ are

$$h_{00}=1, h_{01}=1, h_{10}=-2 \quad\text{and}\quad h_{11}=2.$$

Hence $\mathbf{M}=\mathbf{HH}'$ has the positive eigenvalues $v_1=8$ and $v_2=2$. The limit point process in (3.30) is

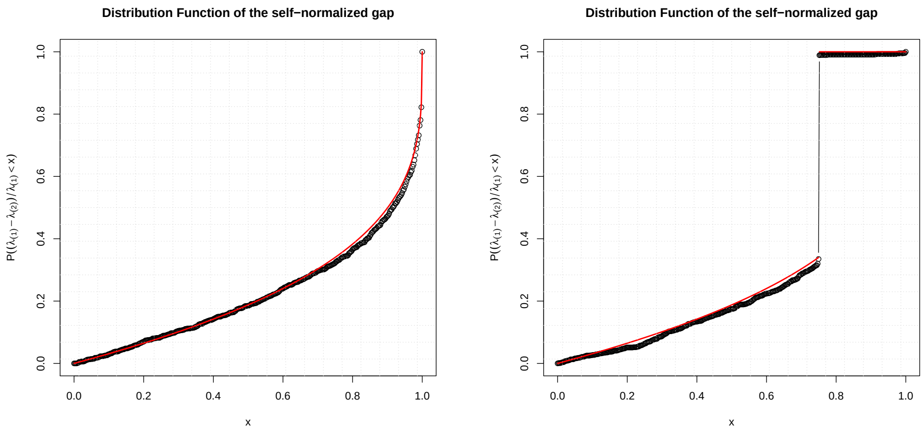
$$N=\sum_{i=1}^{\infty}\varepsilon_{8\Gamma_i^{-2/\alpha}}+\sum_{i=1}^{\infty}\varepsilon_{2\Gamma_i^{-2/\alpha}}\,,$$

so that

$$a_{np}^{-2}\left(\lambda_{(1)},\lambda_{(2)}\right) \xrightarrow{\rm d} \left(8\Gamma_1^{-2/\alpha}, 2\Gamma_1^{-2/\alpha}\vee 8\Gamma_2^{-2/\alpha}\right).$$

Using the fact that $U=\Gamma_1/\Gamma_2$ has a uniform distribution on $(0,1)$ we calculate

$$\mathbb{P}(2\Gamma_1^{-2/\alpha} > 8\Gamma_2^{-2/\alpha}) = \mathbb{P}(\Gamma_1/\Gamma_2 < 2^{-\alpha}) = 2^{-\alpha}\in(1/4,1).$$



(a) iid data (b) data from model (3.36)

Figure 3.5: Distribution function of $(\lambda_{(1)}-\lambda_{(2)})/\lambda_{(1)}$ for iid data (left) and data generated from the model (3.36) (right). In each graph we compare the empirical distribution function (dotted line, based on 1000 simulations of 200×1000 matrices with Z-distribution (3.33)) with the theoretical curve (solid line).

In particular, we have for the normalized spectral gap

$$a_{np}^{-2}\big(\lambda_{(1)} - \lambda_{(2)}\big) \stackrel{\rm d}{\to} 6\,\Gamma_1^{-2/\alpha}\mathbf{1}_{\{\Gamma_1 4^{\alpha/2}<\Gamma_2\}} + 8\,\big(\Gamma_1^{-2/\alpha} - \Gamma_2^{-2/\alpha}\big)\mathbf{1}_{\{\Gamma_1 4^{\alpha/2}>\Gamma_2\}}$$

and for the self-normalized spectral gap (see also Example 3.15 for a detailed analysis)

$$\begin{aligned}\frac{\lambda_{(1)} - \lambda_{(2)}}{\lambda_{(1)}} &\stackrel{\rm d}{\to} \frac{6}{8}\,\mathbf{1}_{\{\Gamma_1 2^{\alpha}<\Gamma_2\}} + \big(1 - (\Gamma_1/\Gamma_2)^{2/\alpha}\big)\mathbf{1}_{\{\Gamma_1 2^{\alpha}>\Gamma_2\}}\\ &= \frac{3}{4}\,\mathbf{1}_{\{U 2^{\alpha}<1\}} + \big(1 - U^{2/\alpha}\big)\mathbf{1}_{\{U2^{\alpha}>1\}} = Y\,.\end{aligned}$$

The limit distribution of the spectral gap has an atom at $3/4$ with probability $2^{-\alpha}$, i.e., $\mathbb{P}(Y = 3/4) = 2^{-\alpha}$, and

$$\mathbb{P}(Y \le x) = 1 - (1 - x)^{\alpha/2}, \qquad x \in (0, 3/4).$$

In the iid case the limit distribution of the self-normalized spectral gap has distribution function $F(x) = 1-(1-x)^{\alpha/2}$ for $x \in [0,1]$. This means that the atom disappears if the entries are iid. Figure 3.5 compares the distribution function of Y with F for $\alpha = 0.6$; the atom at $3/4$ is clearly visible.

Along the same lines, we also have

$$\big(a_{np}^{-2}\lambda_{(1)}, \lambda_{(2)}/\lambda_{(1)}\big) \stackrel{\rm d}{\to} \big(8\,\Gamma_1^{-2/\alpha}, \frac{1}{4}\,\mathbf{1}_{\{U<2^{-\alpha}\}} + U^{2/\alpha}\,\mathbf{1}_{\{U\ge 2^{-\alpha}\}}\big)$$

and hence the limit distribution of $\lambda_{(2)}/\lambda_{(1)}$ is supported on $[1/4, 1)$ with mass of $2^{-\alpha}$ at $1/4$. The histogram of the ratio $\big(\lambda_{(2)}/\lambda_{(1)}\big)^{2/\alpha}$ based on 1000 replications from the model (3.36) with noise given by a t-distribution with $\alpha = 1.5$ degrees of freedom, $n = 1000$

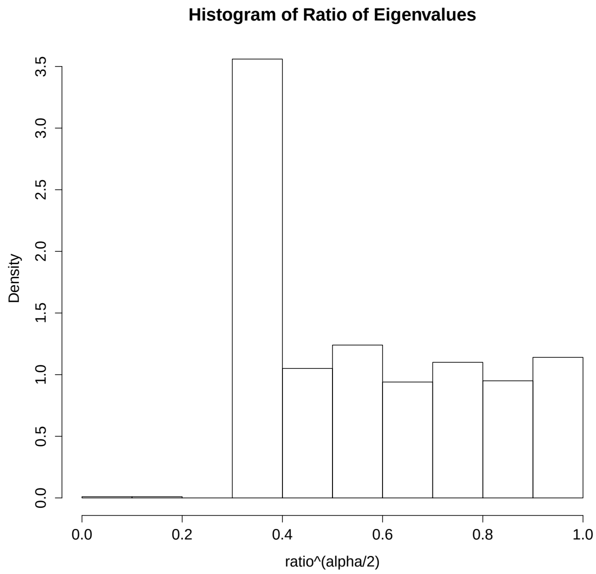


Figure 3.6: Histogram based on 1000 replications of $\big(\lambda_{(2)}/\lambda_{(1)}\big)^{2/\alpha}$ from model (3.36).

and $p = 200$ is displayed in Figure 3.6. Observing that $2^{-\alpha} = 0.3536\ldots$, the histogram is remarkably close to what one would expect from a sample from the truncated uniform distribution, $2^{-\alpha}\,\mathbf{1}_{\{U<2^{-\alpha}\}} + U\,\mathbf{1}_{\{U\ge 2^{-\alpha}\}}$. The mass of the limiting discrete component of the ratio can be much larger if one conditions on $a_{np}^{-2}\lambda_{(1)}$ being large. Specifically, for any $\epsilon \in (0, 1/4)$ and $x > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\epsilon < \lambda_{(2)}/\lambda_{(1)} \le 1/4 | \lambda_{(1)} > a_{np}^2 x) = \mathbb{P}(\Gamma_1/\Gamma_2 \le 2^{-\alpha} | \Gamma_1 < (x/8)^{-\alpha/2}) = G(x)\,.$$

The function G approaches 1 as $x \to \infty$ indicating the speed at which the two largest eigenvalues get linearly related; see Figure 3.7 for a graph of G in the case $\alpha = 1.5$. In addition, from Remark 3.13, we also have

$$\frac{\lambda_{(1)}}{\lambda_1 + \cdots + \lambda_p} \stackrel{\mathrm{d}}{\rightarrow} \frac{4}{5}\, \frac{\Gamma_1^{-2/\alpha}}{\sum_{i=1}^\infty \Gamma_i^{-2/\alpha}}\,.$$

Clearly, the limit random variable is stochastically smaller than what one would get in the iid case; see (3.35).

Example 3.15. The previous example also illustrates the behavior of the two largest eigenvalues in the general case when the rank r of the matrix $\mathbf{M}$ is larger than one. We have in general

$$\frac{\lambda_{(2)}}{\lambda_{(1)}} \stackrel{\mathrm{d}}{\rightarrow} \frac{v_2}{v_1}\,\mathbf{1}_{\{U<(v_2/v_1)^{\alpha/2}\}} + U^{2/\alpha}\,\mathbf{1}_{\{U\ge(v_2/v_1)^{\alpha/2}\}}\,.$$

In particular, the limiting *self-normalized spectral gap* has representation

$$\frac{\lambda_{(1)} - \lambda_{(2)}}{\lambda_{(1)}} \stackrel{\mathrm{d}}{\rightarrow} \frac{v_1 - v_2}{v_1}\,\mathbf{1}_{\{U<(v_2/v_1)^{\alpha/2}\}} + (1 - U^{2/\alpha})\,\mathbf{1}_{\{U\ge(v_2/v_1)^{\alpha/2}\}}\,.$$

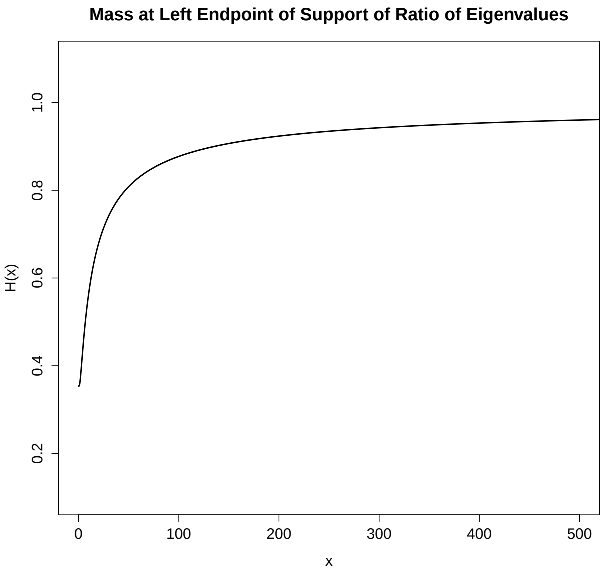


Figure 3.7: Graph of $G(x) = \mathbb{P}(\Gamma_1/\Gamma_2 \le 2^{-\alpha} | \Gamma_1 < (x/8)^{-\alpha/2})$ when $\alpha = 1.5$.

The limiting variable assumes values in $(0, 1 - v_2/v_1]$ and has an atom at the right end-point. This is in contrast to the iid case and to the case when $r = 1$ (hence $v_2 = 0$) including the case of iid rows and the separable case; see Example 3.16.

Example 3.16. We consider the separable case when $h_{kl} = \theta_k c_l$, $k, l \in \mathbb{Z}$, where (c_l), (θ_k) are real sequences such that the conditions on (h_{kl}) in Theorem 3.7 hold. In this case,

$$\mathbf{M} = \sum_{l \in \mathbb{Z}} c_l^2 \ (\theta_i \theta_j)_{i,j \in \mathbb{Z}} .$$

Note that $r = 1$ with the only non-negative eigenvalue

$$v_1 = \sum_{l \in \mathbb{Z}} c_l^2 \ \sum_{k \in \mathbb{Z}} \theta_k^2 .$$

In this case, the limiting point process in Theorem 3.10 is a PRM on $(0, \infty)$ with mean measure of (y, ∞) given by $(v_1/y)^{\alpha/2}$, $y > 0$. The normalized eigenvalues have similar asymptotic behavior as in the case of iid entries. For example, the log-spacings have the same limit as in the iid case for fixed k,

$$\big(\log \lambda_{(1)} - \log \lambda_{(2)}, \ldots, \log \lambda_{(k+1)} - \log \lambda_{(k)} \big) \xrightarrow{\mathrm{d}} -\frac{2}{\alpha} \big(\log(\Gamma_1/\Gamma_2), \ldots, \log(\Gamma_k/\Gamma_{k+1}) \big) .$$

The same observation applies to the ratio of the largest eigenvalue and the trace in the case $\alpha \in (0, 2)$:

$$\frac{\lambda_{(1)}}{\mathrm{tr}(\mathbf{X}\mathbf{X}')} = \frac{\lambda_{(1)}}{\lambda_1 + \cdots + \lambda_p} \xrightarrow{\mathrm{d}} \frac{\Gamma_1^{-2/\alpha}}{\sum_{i=1}^{\infty} \Gamma_i^{-2/\alpha}} .$$

We also mentioned in Example 3.15 that the distributional limit of the self-normalized spectral gap has no atom as in the iid case.

3.4.2 S&P 500 data

We conduct a short analysis of the largest eigenvalues of the univariate log-return time series which compose the S&P 500 stock index; see Section 3.1.2 for a description of the data. Although there is strong empirical evidence that these univariate series have power-law tails (see Figure 3.3) we do not expect that they have the same tail index. One way to proceed would be to ignore this fact because the tail indices are in a close range and the differences are due to large sampling errors for estimating such quantities. One could also collect time series with similar tail indices in the same group. In this case, the dimension p decreases. This grouping would be a rather arbitrary classification method. We have chosen a third way: to use rank transforms. This approach has its merits because it aims at standardizing the tails but it also has a major disadvantage: one destroys the covariance structure underlying the data.

Given a $p \times n$ matrix $(R_{it})_{i=1,\cdots,p;t=1,\cdots,n}$, we construct a matrix $\mathbf{X}$ via the rank transforms

$$X_{it} = -\Big[\log\Big(\frac{1}{n+1}\sum_{\tau=1}^{n}\mathbf{1}_{\{R_{i\tau}\le R_{it}\}}\Big)\Big]^{-1}, \qquad i=1,\ldots,p; t=1,\ldots,n\,.$$

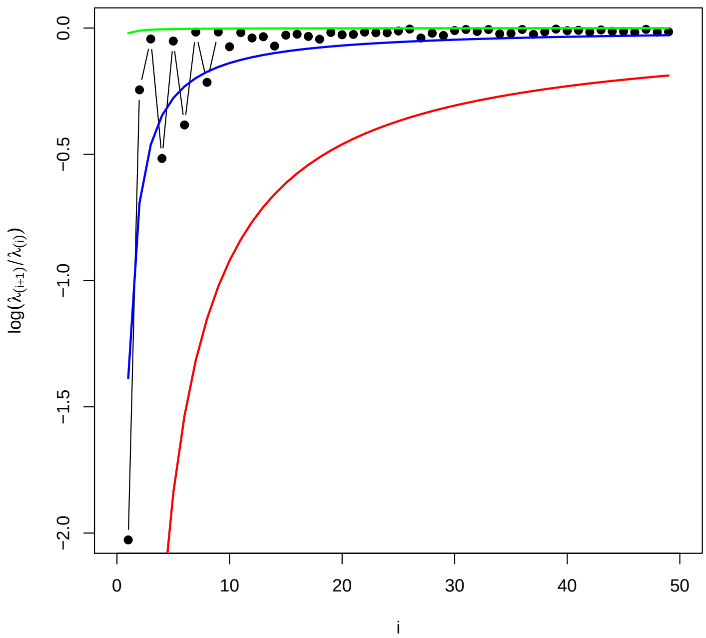


Figure 3.8: The logarithms of the ratios $\lambda_{(i+1)}/\lambda_{(i)}$ for the S&P 500 series after rank transform. We also show the 1, 50 and 99% quantiles (bottom, middle, top lines, respectively) of the variables $\log((\Gamma_i/\Gamma_{i+1})^2)$.

If the rows $R_{i1},\ldots,R_{in}$ were iid (or, more generally, stationary ergodic) with a continuous distribution then the averages under the logarithm would be asymptotically uniform

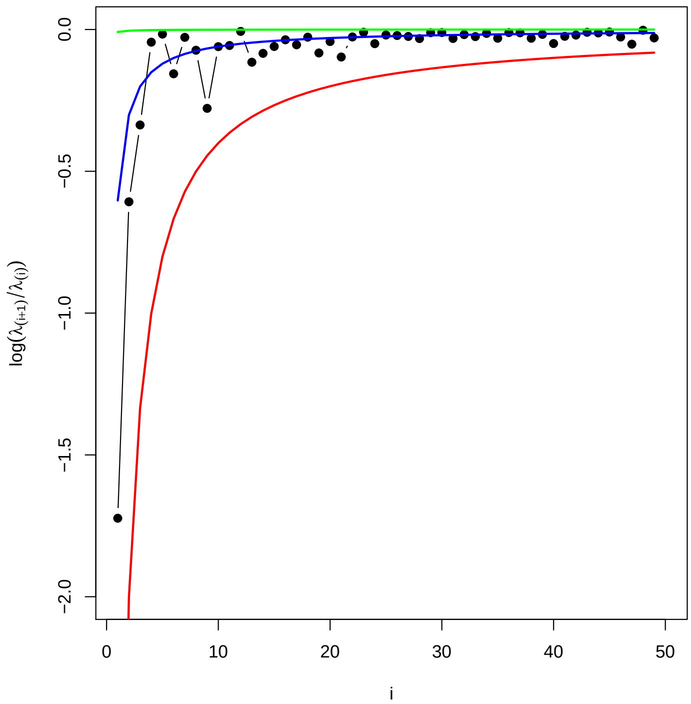


Figure 3.9: The logarithms of the ratios $\lambda_{(i+1)}/\lambda_{(i)}$ for the original (non-rank transformed) S&P 500 log-return data. We also show the 1, 50 and 99% quantiles (bottom, middle, top lines, respectively) of the variables $\log((\Gamma_i/\Gamma_{i+1})^{2/2.3})$; see also Figure 3.8 for comparison.

on $(0,1)$ as $n \to \infty$. Hence X_{it} would be asymptotically standard Fréchet Φ_1-distributed. In what follows, we assume that the aforementioned univariate time series of the S&P 500 index have undergone the rank transform and that their marginal distributions are close to Φ_1; we always use the symbol $\mathbf{X}$ for the resulting multivariate series.

In Figure 3.8 we show the ratios of the consecutive ordered eigenvalues $(\lambda_{(i+1)}/\lambda_{(i)})$ of the matrix $\mathbf{XX}'$. This graph shows the rather surprising fact that the ratios are close to one even for small values i. We also show the 1, 50 and 99 % quantiles of the variables $((\Gamma_i/\Gamma_{i+1})^{2/\alpha})$ calculated from the formula

$$\mathbb{P}\big((\Gamma_i/\Gamma_{i+1})^{2/\alpha} \le x\big) = x^{i\cdot\alpha/2}, \quad x \in (0,1)\,. \tag{3.37}$$

For increasing i, the distribution is concentrated closely to 1, in agreement with the strong law of large numbers which yields $\Gamma_i/\Gamma_{i+1} \stackrel{\text{a.s.}}{\to} 1$ as $i \to \infty$. The asymptotic distributions (3.37) correspond to the case when the matrix $\mathbf{M}$ has rank $r = 1$. It includes the iid and separable cases; see Example 3.16. The shown asymptotic quantiles are in agreement with the rank $r = 1$ hypothesis.

For comparison, in Figure 3.9 we also show the ratios $(\lambda_{(i+1)}/\lambda_{(i)})$ for the non-rank transformed S&P 500 data and the 1, 50 and 99% quantiles of the variables $\log((\Gamma_i/\Gamma_{i+1})^{2/\alpha})$, where we choose $\alpha = 2.3$ motivated by the estimated tail indices in Figure 3.3. The two graphs in Figure 3.8 and Figure 3.9 are quite similar but the smallest ratios for the original data are slightly larger than for the rank-transformed data.

3.4.3 Sums of squares of sample autocovariance matrices

In this section we consider some additive functions of the squares of $\mathbf{A}_n(s) = \mathbf{X}_n(0)\mathbf{X}_n(s)'$ given by $\mathbf{A}_n(s)\mathbf{A}_n(s)'$ for $s = 0, 1, \ldots$. By definition of the singular values of a matrix (see (3.24)), the non-negative definite matrix $\mathbf{A}_n(s)\mathbf{A}_n(s)'$ has eigenvalues $(\lambda_i^2(s))_{i=1,\ldots,p}$.

The following result is a corollary of Theorem 3.7.

Proposition 3.17. *Consider the linear process* (3.14) *under the conditions of Theorem 3.7. Then the following statements hold for* $s \geq 0$*:*

(1) *We consider two disjoint cases:* $\alpha \in (0,2)$ *and* $\beta \in (0,\infty)$*, or* $\alpha \in [2,4)$ *and* β *satisfying* $\widetilde{C}_\beta(\alpha)$*. Then*

$$a_{np}^{-4} \max_{i=1,\ldots,p} |\lambda_{(i)}^2(s) - \delta_{(i)}^2(s)| \xrightarrow{\mathbb{P}} 0, \quad n \to \infty.$$

(2) *Assume* $\beta \in [0,1]$*. If* $\alpha \in (0,2]$*,* $\mathbb{E}[Z^2] = \infty$ *or* $\alpha \in [2,4)$*,* $\mathbb{E}[Z^2] < \infty$ *and* $\beta \in (\alpha/2 - 1, 1]$*, then*

$$a_{np}^{-4} \max_{i=1,\ldots,p} |\lambda_{(i)}^2(s) - (\gamma_{(i)}^{\rightarrow}(s))^2| \xrightarrow{\mathbb{P}} 0, \quad n \to \infty.$$

Assume $\beta > 1$*. If* $\alpha \in (0,2]$*,* $\mathbb{E}[Z^2] = \infty$ *or* $\alpha \in [2,4)$*,* $\mathbb{E}[Z^2] < \infty$ *and* $\beta^{-1} \in (\alpha/2 - 1, 1]$*. Then*

$$a_{np}^{-4} \max_{i=1,\ldots,p} |\lambda_{(i)}^2(s) - (\gamma_{(i)}^{\downarrow}(s))^2| \xrightarrow{\mathbb{P}} 0, \quad n \to \infty.$$

To the best of our knowledge, sums of squares of sample autocovariance matrices were used first in the paper by Lam and Yao [49]; their time series model is quite different from ours.

Proof. Part (1). The proof follows from Theorem 3.7 if we can show that

$$a_{np}^{-2} \max_{i=1,\ldots,p} \big(\lambda_{(i)}(s) + \delta_{(i)}(s)\big) = O_{\mathbb{P}}(1) \quad n \to \infty\,.$$

We have by Theorem 3.10,

$$a_{np}^{-2} \max_{i=1\ldots,p} \lambda_{(i)}(s) = a_{np}^{-2} \lambda_{(1)}(s) \xrightarrow{\mathrm{d}} c\,\xi_{\alpha/2}\,, \tag{3.38}$$

where $\xi_{\alpha/2}$ has a $\Phi_{\alpha/2}$ distribution. In view of Theorem 3.7(1) we also have

$$a_{np}^{-2} \max_{i=1\ldots,p} \delta_{(i)}(s) \xrightarrow{\mathrm{d}} c\,\xi_{\alpha/2}\,.$$

Therefore, again using Theorem 3.7(1), we have

$$\begin{aligned} & a_{np}^{-4} \max_{i=1,\ldots,p} |\lambda_{(i)}^2(s) - \delta_{(i)}^2(s)| \\ \leq \quad & \big[a_{np}^{-2} \max_{i=1,\ldots,p} |\lambda_{(i)}(s) - \delta_{(i)}(s)|\big] \big[a_{np}^{-2} \max_{i=1,\ldots,p} \big(|\lambda_{(i)}(s)| + |\delta_{(i)}(s)|\big)\big] \xrightarrow{\mathbb{P}} 0, \quad n \to \infty. \end{aligned}$$

This proves part (1).

Part (2). Now assume $\beta \in [0,1]$ and $\alpha \in (0,2]$, $\mathbb{E}[Z^2] = \infty$ or $\alpha \in [2,4)$, $\mathbb{E}[Z^2] < \infty$ and $\beta \in (\alpha/2-1,1]$. Then (3.38) is still true and we have by Theorem 3.7(2) and Theorem 3.10

$$a_{np}^{-2} \max_{i=1\dots,p} \gamma_{\overrightarrow{(i)}}(s) \xrightarrow{\mathrm{d}} c\,\xi_{\alpha/2}\,.$$

We then have

$$\begin{aligned} & a_{np}^{-4} \max_{i=1,\dots,p} |\lambda_{(i)}^2(s) - (\gamma_{\overrightarrow{(i)}}(s))^2| \\ \le\ & \big[a_{np}^{-2} \max_{i=1,\dots,p} |\lambda_{(i)}(s) - \gamma_{\overrightarrow{(i)}}(s)|\big]\, \big[a_{np}^{-2} \max_{i=1,\dots,p} \big(\lambda_{(i)}(s) + \gamma_{\overrightarrow{(i)}}(s)\big)\big] \xrightarrow{\mathbb{P}} 0\,,\ n \to \infty. \end{aligned}$$

The proof of the remaining part is similar and therefore omitted. □

Now, using Proposition 3.17 and a continuous mapping argument, we can show limit theory for the eigenvalues

$$w_{(1)}(s_0,s_1) \ge \cdots \ge w_{(p)}(s_0,s_1)\,, \qquad 0 \le s_0 \le s_1\,,$$

of the non-negative definite random matrices

$$\sum_{s=s_0}^{s_1} \mathbf{A}_n(s)\mathbf{A}_n(s)'\,. \tag{3.39}$$

Proposition 3.18. *Assume $0 \le s_0 \le s_1$ and the conditions of Theorem 3.7 hold. If $\alpha \in (0,4)$ and $\beta \in (0,1] \cap (\alpha/2-1,1]$ then*

$$a_{np}^{-4} \max_{i=1,\dots,p} |w_{(i)}(s_0,s_1) - \omega_{(i)}(s_0,s_1)| \xrightarrow{\mathbb{P}} 0, \quad n \to \infty,$$

where $\omega_{(i)}(s_0,s_1)$ are the ordered values of the set $\{Z_{(i),np}^4 v_j(s_0,s_1), i = 1,\dots,p; j = 1,2,\dots\}$ and $(v_j(s_0,s_1))$ are the ordered eigenvalues of $\sum_{s=s_0}^{s_1} \mathbf{M}(s)\mathbf{M}(s)'$.

Example 3.19. Recall the separable case from Example 3.16, i.e., $h_{kl} = \theta_k c_l$, $k,l \ge 0$, where (c_l), (θ_k) are real sequences such that the conditions on (h_{kl}) in Theorem 3.7 hold. Write $\Theta_{ij} = \theta_i\theta_j$. It is symmetric and has rank one; the only non-zero eigenvalue is $\gamma_\theta(0) = \sum_{k=0}^\infty \theta_k^2$. Hence Θ is non-negative definite. We get from (3.25) that

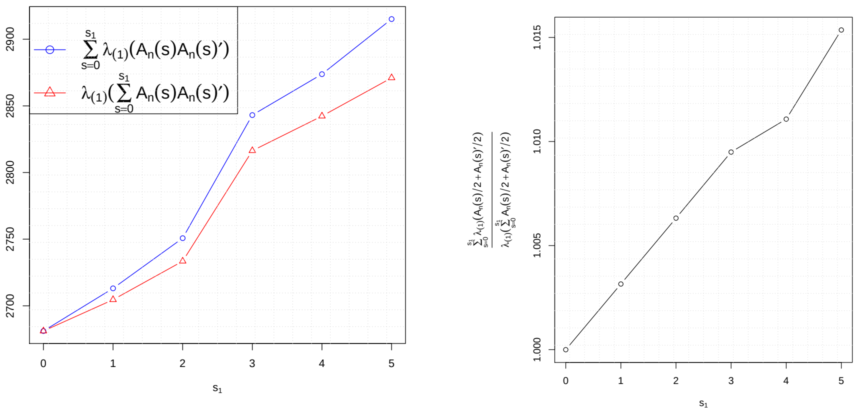
$$\mathbf{M}(s) = \gamma_c(s)\,\Theta, \quad s \ge 0\,,$$

where

$$\gamma_c(s) = \sum_{l=0}^\infty c_l c_{l+s}\,, \qquad s \ge 0\,.$$

The matrix $\mathbf{M}(s)$ has the only non-zero eigenvalue $\gamma_c(s)\gamma_\theta(0)$. The factors $(\gamma_c(s))$ can be positive or negative; they constitute the autocovariance function of a stationary linear process with coefficients (c_l). Accordingly, $\mathbf{M}(s)$ is either non-negative or non-positive definite. Now we consider the non-negative definite matrix

$$\sum_{s=s_0}^{s_1} \mathbf{M}(s)\,\mathbf{M}(s)' = \sum_{s=s_0}^{s_1} \gamma_c^2(s)\,\Theta\Theta'\,.$$



(a) (b)

Figure 3.10: The largest eigenvalues of the sums of the squared autocovariance matrices compared with the sums of the largest eigenvalues of these matrices for the S&P 500 data for different values s_1. The two values are surprisingly close to each other; mind the scale of the y-axis. We also show their ratios.

This matrix has rank 1 and its largest eigenvalue is given by

$$C_{c,\theta}(s_0,s_1)=\sum_{s=s_0}^{s_1}\gamma_c^2(s)\,\gamma_\theta^2(0)\,.$$

An application of Proposition 3.18 yields that the ordered eigenvalues of the matrix $a_{np}^{-4}\sum_{s=s_0}^{s_1}\mathbf{A}_n(s)\mathbf{A}_n(s)'$ are uniformly approximated by the quantities

$$a_{np}^{-4}Z_{(i),np}^4 C_{c,\theta}(s_0,s_1)\,,\qquad i=1,\ldots,p\,. \tag{3.40}$$

Since

$$C_{c,\theta}(s_0,s_1)=\sum_{i=s_0}^{s_1}C_{c,\theta}(i,i)$$

one gets the remarkable property that

$$\begin{aligned}&a_{np}^{-4}\max_{i=1,\ldots,p}\Big|\lambda_{(i)}\Big(\sum_{s=s_0}^{s_1}\mathbf{A}_n(s)\mathbf{A}_n(s)'\Big)-Z_{(i),np}^4C_{c,\theta}(s_0,s_1)\Big|\\ =\;&a_{np}^{-4}\max_{i=1,\ldots,p}\Big|\sum_{s=s_0}^{s_1}\lambda_{(i)}(\mathbf{A}_n(s)\mathbf{A}_n(s)')-Z_{(i),np}^4C_{c,\theta}(s_0,s_1)\Big|+o_P(1)\,.\end{aligned}$$

In particular, for $s_1\ge s_0$ we get the weak convergence of the point processes towards a PRM:

$$\begin{aligned}&\sum_{i=1}^{p}\varepsilon_{a_{np}^{-4}\Big(\lambda_i\big(\sum_{s=s_0}^{s_0}\mathbf{A}_n(s)\mathbf{A}_n(s)'\big),\ldots,\lambda_i\big(\sum_{s=s_0}^{s_1}\mathbf{A}_n(s)\mathbf{A}_n(s)'\big)\Big)}\\ &\stackrel{\mathrm{d}}{\to}\sum_{i=1}^{\infty}\varepsilon_{\Gamma_i^{-4/\alpha}\big(C_{c,\theta}(s_0,s_0),\ldots,C_{c,\theta}(s_0,s_1)\big)}\,,\qquad n\to\infty\,.\end{aligned}$$

Example 3.20. In Figure 3.10 we calculate the largest eigenvalues $\lambda_{(1)}\big(\sum_{s=0}^{s_1}\mathbf{A}_n(s)\mathbf{A}_n(s)'\big)$ for $s_1 = 0,\dots,5$ as well as the sums of the largest eigenvalues $\sum_{s=0}^{s_1}\lambda_{(1)}\big(\mathbf{A}_n(s)\mathbf{A}_n(s)'\big)$ the log-return series from the S&P 500 index described in Section 3.1.2. The data are not rank-transformed. We notice that the two values are surprisingly close across the values $s_0 = 0,\dots,5$. This phenomenon could be explained by the structure of the eigenvalues in Example 3.19. Also note that the largest eigenvalue $\mathbf{A}_n(0)\mathbf{A}_n(0)'$ makes a major contribution to the values in Figure 3.10; the contribution of the squares $\mathbf{A}_n(s)\mathbf{A}_n(s)'$, $s = 1,\dots,5$, to the largest eigenvalue of the sum of squares is less substantial.

3.5 Auxiliary results

Let (Z_i) be iid copies of Z whose distribution satisfies

$$\mathbb{P}(Z > x) \sim p_+\frac{L(x)}{x^\alpha} \quad \text{and} \quad \mathbb{P}(Z \le -x) \sim p_-\frac{L(x)}{x^\alpha}\,, \quad x \to \infty\,,$$

for some tail index $\alpha > 0$, where $p_+, p_- \ge 0$ with $p_+ + p_- = 1$ and L is a slowly varying function. If $\mathbb{E}[|Z|] < \infty$ also assume $\mathbb{E}[Z] = 0$. The product Z_1Z_2 is regular varying with the same index α and $\mathbb{P}(|Z_1Z_2| > x) = x^{-\alpha}L_1(x)$, where L_1 is slowly varying function different from L; see Embrechts and Goldie [34]. Write

$$S_n = Z_1 + \dots + Z_n\,, \quad n \ge 1,$$

and consider a sequence (a_n) such that $\mathbb{P}(|Z| > a_n) \sim n^{-1}$.

3.5.1 Large deviation results

The following theorem can be found in Nagaev [55] and Cline and Hsing [22] for $\alpha > 2$ and $\alpha \le 2$, respectively; see also Denisov et al. [27].

Theorem 3.21. *Under the assumptions on the iid sequence (Z_t) given above the following relation holds*

$$\sup_{x \ge c_n}\left|\frac{\mathbb{P}(S_n > x)}{n\mathbb{P}(|Z| > x)} - p_+\right| \to 0\,,$$

where (c_n) is any sequence satisfying $c_n/a_n \to \infty$ for $\alpha \le 2$ and $c_n \ge \sqrt{(\alpha-2)n\log n}$ for $\alpha > 2$.

3.5.2 A point process convergence result

Assume that the conditions at the beginning of Appendix 3.5 hold. Consider a sequence of iid copies $(S_n^{(t)})_{t=1,2,\dots}$ of S_n and the sequence of point processes

$$N_n = \sum_{t=1}^{p} \varepsilon_{a_{np}^{-1}S_n^{(t)}}\,, \quad n = 1,2,\dots\,,$$

for an integer sequence $p = p_n \to \infty$. We assume that the state space of the point processes N_n is $\overline{\mathbb{R}}_0 = [\mathbb{R} \cup \{\pm\infty\}]\backslash\{0\}$.

Lemma 3.22. *Assume $\alpha \in (0,2)$ and the conditions of Appendix 3.5 on the iid sequence (Z_t) and the normalizing sequence (a_n). Then the limit relation $N_n \stackrel{\mathrm{d}}{\to} N$ holds in the space of point measures on $\overline{\mathbb{R}}_0$ equipped with the vague topology (see [61, 60]) for a Poisson random measure N with state space $\overline{\mathbb{R}}_0$ and intensity measure $\mu_\alpha(dx) = \alpha|x|^{-\alpha-1}(p_+\mathbf{1}_{\{x>0\}} + p_-\mathbf{1}_{\{x<0\}})dx$.*

Proof. According to Resnick [61], Proposition 3.21, we need to show that $p\,\mathbb{P}(a_{np}^{-1}S_n \in \cdot) \stackrel{\mathrm{v}}{\to} \mu_\alpha$, where $\stackrel{\mathrm{v}}{\to}$ denotes vague convergence of Radon measures on $\overline{\mathbb{R}}_0$. Observe that we have $a_{np}/a_n \to \infty$ as $n \to \infty$. This fact and $\alpha \in (0,2)$ allow one to apply Theorem 3.21:

$$\frac{\mathbb{P}(S_n > x a_{np})}{n\,\mathbb{P}(|Z| > a_{np})} \to p_+ x^{-\alpha} \quad \text{and} \quad \frac{\mathbb{P}(S_n \le -x a_{np})}{n\,\mathbb{P}(|Z| > a_{np})} \to p_-\, x^{-\alpha}\,, \quad x > 0\,.$$

On the other hand, $n\,\mathbb{P}(|Z| > a_{np}) \sim p^{-1}$ as $n \to \infty$. This proves the lemma. □

Chapter 4

Almost sure convergence of the largest and smallest eigenvalues of high-dimensional sample correlation matrices under infinite fourth moment

Johannes Heiny & Thomas Mikosch (2016)

Abstract

In this paper, we show that the largest and smallest eigenvalues of a sample correlation matrix stemming from n independent observations of a p-dimensional time series with iid components converge almost surely to $(1+\sqrt{\gamma})^2$ and $(1-\sqrt{\gamma})^2$, respectively, as $n \to \infty$, if $p/n \to \gamma \in (0,1]$ and the truncated variance of the entry distribution is "almost slowly varying", a condition we describe via moment properties of self-normalized sums. Moreover, the empirical spectral distributions of these sample correlation matrices converge weakly, with probability 1, to the Marčenko–Pastur law, which extends a result in [7]. We compare the behavior of the eigenvalues of the sample covariance and sample correlation matrices and argue that the latter seems more robust, in particular in the case of infinite fourth moment. We briefly address some practical issues for the estimation of extreme eigenvalues in a simulation study.

In our proofs we use the method of moments combined with a Path-Shortening Algorithm, which efficiently uses the structure of sample correlation matrices, to calculate precise bounds for matrix norms. We believe that this new approach could be of further use in random matrix theory.

Keywords: Sample correlation matrix, infinite fourth moment, largest eigenvalue, smallest eigenvalue, spectral distribution, sample covariance matrix, self-normalization, regular variation, combinatorics.

4.1 Introduction and notation

In modern statistical analyses one is often faced with large data sets where both the dimension of the observations and the sample size are large. The dramatic increase and improvement of computing power and data collection devices have triggered the necessity to study and interpret the sometimes overwhelming amounts of data in an efficient and tractable way. Huge data sets arise naturally in wireless communication, finance, natural sciences and genetic engineering. For such data one commonly studies the dependence structure via covariances and correlations which can be estimated by their sample analogs. *Principal component analysis*, for example, uses an orthogonal transformation of the data such that only a few of the resulting vectors explain most of the variation in the data. The empirical variances of these so-called *principal component vectors* are the largest eigenvalues of the *sample covariance or correlation matrix*.

Throughout this paper we consider the $p \times n$ data matrix

$$\mathbf{X} = \mathbf{X}_n = (X_{it})_{i=1,\dots,p;t=1,\dots,n}$$

of identically distributed entries (X_{it}) with generic element X, where we assume $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = 1$ if the first and second moments of X are finite, respectively. A column of $\mathbf{X}$ represents an observation of a p-dimensional time series.

Random matrix theory provides a great variety of results on the ordered eigenvalues

$$\lambda_{(1)} \geq \cdots \geq \lambda_{(p)} , \tag{4.1}$$

of the (non-normalized) *sample covariance matrix* $\mathbf{XX}'$. Here we will only discuss the case $p = p_n \to \infty$ and, unless stated otherwise, we assume the growth condition

$$\lim_{n\to\infty} \frac{p_n}{n} \to \gamma \in (0,1] . \tag{G_γ}$$

For the finite p case, we refer to [3, 54, 45]. When studying the asymptotic properties of estimators under (G_γ) one often obtains results that dramatically differ from the standard p fixed, $n \to \infty$ case, in which the spectrum of $(n^{-1}\mathbf{XX}')$ converges to its population covariance spectrum. In 1967, Marčenko and Pastur [51] observed that even in the case of iid entries (X_{it}) with $\mathbb{E}[X^2] = 1$ the eigenvalues $(\lambda_{(i)}/n)$ do not concentrate around 1. For more examples, see [6, Chapter 1] and [32]. Typical applications where (G_γ) seems reasonable are discussed in [48, 29].

In comparison with $(\lambda_{(i)})$, much less is known about the ordered eigenvalues

$$\mu_{(1)} \geq \cdots \geq \mu_{(p)}$$

of the *sample correlation matrix* $\mathbf{R} = \mathbf{YY}'$ with entries

$$R_{ij} = \sum_{t=1}^{n} \frac{X_{it}X_{jt}}{\sqrt{D_i}\sqrt{D_j}} = \sum_{t=1}^{n} Y_{it}Y_{jt} , \quad i,j = 1,\dots,p . \tag{4.2}$$

In this paper we will often make use of the notation $\mathbf{Y} = (Y_{it}) = (X_{it}/\sqrt{D_i})$ and

$$D_i = D_i^{(n)} = \sum_{t=1}^{n} X_{it}^2 , \qquad i = 1,\dots,p;\ n \geq 1 . \tag{4.3}$$

Note that the dependence of $(\lambda_{(i)})$ and $(\mu_{(i)})$ on n is suppressed in the notation.

4.1.1 The case (X_{it}) iid, $\mathbb{E}[X^4] < \infty$ and $\mathbb{E}[X^2] = 1$

In this case the behavior of the eigenvalues of the sample covariance matrix $\mathbf{XX}'$ and the sample correlation matrix $\mathbf{R}$ are closely intertwined.

For any random $p \times p$ matrix $\mathbf{A}$ with real eigenvalues $\lambda_1(\mathbf{A}), \ldots, \lambda_p(\mathbf{A})$ the *empirical spectral distribution* is defined by

$$F_{\mathbf{A}}(x) = \frac{1}{p} \sum_{i=1}^{p} \mathbf{1}_{\{\lambda_i(\mathbf{A}) \le x\}}, \qquad x \in \mathbb{R}.$$

Many functionals of the eigenvalues $\lambda_1(\mathbf{A}), \ldots, \lambda_p(\mathbf{A})$ can be expressed in terms of $F_{\mathbf{A}}$ [5], for instance

$$\det \mathbf{A} = \prod_{i=1}^{p} \lambda_i(\mathbf{A}) = \exp\Big(p \int_0^\infty \log x \; \mathrm{d}F_{\mathbf{A}}(x)\Big).$$

A major problem in random matrix theory is to find the weak limit of $(F_{\mathbf{A}_n})$ for suitable sequences $(\mathbf{A}_n)$; see for example [6, 78] for more details. By weak convergence of a sequence of probability distributions $(F_{\mathbf{A}_n})$ to a probability distribution F, we mean $\lim_{n\to\infty} F_{\mathbf{A}_n}(x) = F(x)$ a.s. for all continuity points of F. In this context a useful tool is the *Stieltjes transform* of the empirical spectral distribution $F_{\mathbf{A}}$:

$$s_{\mathbf{A}}(z) = \int_{\mathbb{R}} \frac{1}{x-z} \, \mathrm{d}F_{\mathbf{A}}(x) = \frac{1}{p} \operatorname{tr}(\mathbf{A} - z\mathbf{I})^{-1}, \quad z \in \mathbb{C}^+,$$

where $\mathbb{C}^+$ denotes the complex numbers with positive imaginary part. Weak convergence of $(F_{\mathbf{A}_n})$ to F is equivalent to $s_{F_{\mathbf{A}_n}}(z) \to s_F(z)$ a.s. for all $z \in \mathbb{C}^+$.

Under the growth condition (G_γ), the sequence of empirical spectral distributions of the normalized sample covariance matrix $n^{-1}\mathbf{XX}'$ converges weakly to the Marčenko–Pastur law with density

$$f_\gamma(x) = \begin{cases} \frac{1}{2\pi x\gamma}\sqrt{(b-x)(x-a)}\,, & \text{if } a \le x \le b, \\ 0\,, & \text{otherwise,} \end{cases} \tag{4.4}$$

where $\gamma \in (0,1]$, $a = (1-\sqrt{\gamma})^2$ and $b = (1+\sqrt{\gamma})^2$. This classical result is sometimes referred to as Marčenko–Pastur theorem [51]. Informally, the histogram of $(\lambda_{(i)}/n)$ is asymptotically non-random and the limiting shape depends only on the fraction p/n. For an illustration, see Figure 4.1.

The Marčenko–Pastur law has k-th moment

$$\beta_k = \beta_k(\gamma) = \int_a^b x^k f_\gamma(x)\,\mathrm{d}x = \sum_{r=1}^{k} \frac{1}{r}\binom{k}{r-1}\binom{k-1}{r-1}\gamma^{r-1}, \quad k \ge 1, \tag{4.5}$$

and Stieltjes transform

$$s(z) = \int_{\mathbb{R}} \frac{1}{x-z} f_\gamma(x)\,\mathrm{d}x = \frac{1-\gamma-z+\sqrt{(1+\gamma-z)^2-4\gamma}}{2\gamma z}; \tag{4.6}$$

see [6, Chapter 3] or [5, 78].

The a.s. behavior of the extreme eigenvalues is more involved and therefore it has received significant attention in the literature. From the Marčenko–Pastur theorem one can infer

$$\limsup_{n\to\infty} n^{-1}\lambda_{(p)} \le (1-\sqrt{\gamma})^2 \le (1+\sqrt{\gamma})^2 \le \liminf_{n\to\infty} n^{-1}\lambda_{(1)} \quad \text{a.s.} \tag{4.7}$$

The finiteness of the fourth moment of X is necessary for the almost sure convergence of $\lambda_{(1)}/n$; see [9]. If $\mathbb{E}[X^4] < \infty$, one has (see [6])

$$n^{-1}\lambda_{(1)} \to (1+\sqrt{\gamma})^2 \quad \text{and} \quad n^{-1}\lambda_{(p)} \to (1-\sqrt{\gamma})^2 \quad \text{a.s.} \tag{4.8}$$

The minimal moment requirement for the convergence of the normalized smallest eigenvalue, however, was an open question for a long time. Recently, it was proved in [70] that $n^{-1}\lambda_{(p)} \to (1-\sqrt{\gamma})^2$ a.s. only requires a finite second moment. Under suitable moment assumptions $\lambda_{(1)}$ and $\lambda_{(p)}$ possess *Tracy–Widom* fluctuations around their almost sure limits. For instance, the paper [48] complemented (4.8) by the corresponding central limit theorem in the special case of iid standard normal entries:

$$n^{2/3}\, \frac{(\sqrt{\gamma})^{1/3}}{\big(1+\sqrt{\gamma}\big)^{4/3}} \Big(\frac{\lambda_{(1)}}{n} - \big(1+\sqrt{\tfrac{p}{n}}\big)^2\Big) \xrightarrow{\mathrm{d}} \xi\,, \tag{4.9}$$

where the limiting random variable has a *Tracy–Widom distribution* of order 1. Notice that the centering $\big(1+\sqrt{\tfrac{p}{n}}\big)^2$ can in general not be replaced by $(1+\sqrt{\gamma})^2$. This distribution is ubiquitous in random matrix theory. Its distribution function F_1 is given by

$$F_1(s) = \exp\Big\{-\frac{1}{2}\int_s^\infty [q(x)+(x-s)q^2(x)]\,\mathrm{d}x\Big\}\,,$$

where $q(x)$ is the unique solution to the Painlevé II differential equation

$$q''(x) = xq(x) + 2q^3(x)\,,$$

where $q(x) \sim \mathrm{Ai}(x)$ as $x \to \infty$ and $\mathrm{Ai}(\cdot)$ is the Airy kernel; see Tracy and Widom [71] for details.

Sometimes practitioners would like to know "to which extent the random matrix results would hold if one were concerned with sample correlation matrices and not sample covariance matrices [32]". A partial answer is that the aforementioned results also hold for the sample correlation matrix $\mathbf{R}$ and its eigenvalues $\mu_{(1)} \ge \cdots \ge \mu_{(p)}$. With $\mathbf{F} = \mathrm{diag}(1/D_1,\ldots,1/D_p)$, we have $\mathbf{R} = \mathbf{F}^{1/2}\mathbf{X}\mathbf{X}'\mathbf{F}^{1/2}$ which has the same eigenvalues as $\mathbf{X}\mathbf{X}'\mathbf{F}$. Weyl's inequality (see [19]) yields

$$\begin{aligned}\max_{i=1,\ldots,p} |\mu_{(i)} - n^{-1}\lambda_{(i)}| &\le \|\mathbf{X}\mathbf{X}'\mathbf{F} - n^{-1}\mathbf{X}\mathbf{X}'\|_2 \\ &\le n^{-1}\|\mathbf{X}\mathbf{X}'\|_2 \|n\mathbf{F} - \mathbf{I}\|_2 \\ &= n^{-1}\lambda_{(1)} \max_{i=1,\ldots,p}\Big|\frac{n}{D_i} - 1\Big|\,,\end{aligned} \tag{4.10}$$

where for any matrix $\mathbf{A}$, $\|\mathbf{A}\|_2$ denotes its spectral norm, i.e., its largest singular value.

Lemma 2 in [10] implies that $\mathbb{E}[X^4] < \infty$ is equivalent to

$$\max_{i=1,\ldots,p}\Big|\frac{n}{D_i} - 1\Big| \xrightarrow{\text{a.s.}} 0\,,$$

while $n^{-1}\lambda_{(1)} \to (1+\sqrt{\gamma})^2$ a.s. Hence, $\max_{i=1,\ldots,p}|\mu_{(i)} - n^{-1}\lambda_{(i)}| \to 0$ a.s. This approach was used in [46, 77] to derive

$$\mu_{(1)} \to (1+\sqrt{\gamma})^2 \quad \text{and} \quad \mu_{(p)} \to (1-\sqrt{\gamma})^2 \quad \text{a.s.} \tag{4.11}$$

If the assumption $\mathbb{E}[X^4] < \infty$ is weakened to $\lim_{n\to\infty} n\,\mathbb{P}(X^4 > n) = 0$, the paper [10] proves that $n^{-1}\lambda_{(1)} \xrightarrow{\mathbb{P}} (1+\sqrt{\gamma})^2$ and $\max_{i=1,\dots,p} \big|n/D_i - 1\big| \xrightarrow{\mathbb{P}} 0$. As a consequence, the limit results for $\mu_{(1)}$ and $\mu_{(p)}$ hold in probability instead of a.s.

Distributional limit results have been derived for the appropriately centered and normalized eigenvalues of sample correlation matrices. The authors of [14] assumed iid, symmetric entries X_{it} and that there exist positive constants C, C' such that $\mathbb{P}(|X| \ge t^C) \le \mathrm{e}^{-t}, t \ge C'$. They showed (4.9) with $\lambda_{(1)}/n$ replaced by $\mu_{(1)}$. A similar limit result holds for $\mu_{(p)}$.

4.1.2 The case (X_{it}) iid and $\mathbb{E}[X^4] = \infty$

Asymptotic theory for the eigenvalues of $\mathbf{XX}'$ in the case of an entry distribution with infinite fourth moment was studied in [65, 66, 4] in the cases when $p/n \to \gamma \in (0,\infty)$, while the authors of [23, 41] allowed nearly arbitrary growth of the dimension p. In their model, the entries of $\mathbf{X}$ are regularly varying with index $\alpha > 0$, implying that

$$\mathbb{P}(|X| > x) = x^{-\alpha} L(x)\,, \tag{4.12}$$

for a slowly varying function L. For $\alpha \in (0,4)$, which implies an infinite fourth moment, they showed that $(a_{np}^{-2}\lambda_{(1)})$ converges to a Fréchet distributed random variable $\eta_{\alpha/2}$ with parameter $\alpha/2$ while $a_{np}^{-2}\lambda_{(p)} \xrightarrow{\mathbb{P}} 0$. Here the normalizing sequence (a_n) is defined via $\mathbb{P}(|X| > a_n) \sim n^{-1}$, hence $n/a_{np}^2 \to 0$.

To illustrate the stark contrast between the cases $\alpha > 4$ and $\alpha < 4$, assume (G_γ) and $\mathbb{E}[X] = 0$ if $\mathbb{E}[|X|] < \infty$. Then it follows from (4.8) that

$$\begin{aligned}
\frac{\lambda_{(p)}}{\lambda_{(1)}} &\xrightarrow{\text{a.s.}} \frac{(1-\sqrt{\gamma})^2}{(1+\sqrt{\gamma})^2} \quad \text{if } \alpha > 4\,,\\
\frac{a_{np}^2}{n}\frac{\lambda_{(p)}}{\lambda_{(1)}} &\xrightarrow{\text{d}} \frac{(1-\sqrt{\gamma})^2}{\eta_{\alpha/2}} \quad \text{if } \alpha \in (2,4)\,,\\
\frac{a_{np}^2}{n}\frac{\lambda_{(p)}}{\lambda_{(1)}} &\xrightarrow{\text{a.s.}} 0 \quad \text{if } \alpha \in (0,2)\,,
\end{aligned} \tag{4.13}$$

where the rate $a_{np}^2/n \to \infty$ in the last line can even be increased. To the best of our knowledge, a suitable normalization (b_n) such that $(b_n\lambda_{(p)})$ has a nontrivial limit is not available when $\alpha \in (0,2)$.

Under (G_γ) the asymptotic behavior of the eigenvalues of sample correlation matrices can be very different from that of sample covariance matrices, especially for an entry distribution with infinite fourth moment. If $\alpha \in (2,4)$, the Marčenko–Pastur theorem and Theorem 2.3 in [7] assert that $(F_{n^{-1}\mathbf{XX}'})$ and $(F_{\mathbf{R}})$ converge weakly to the Marčenko–Pastur law. From [9] it is known that $\limsup_n \lambda_{(1)}/n = \infty$ a.s.

For $\mathbb{E}[X^4] = \infty$, the approach to sample correlation matrices from (4.10) fails. No limit results for $\mu_{(1)}$ or $\mu_{(p)}$ seem to be available in the literature at this point, although Theorem 2.3 in [7] ensures the weak convergence of the empirical spectral distribution $F_{\mathbf{R}}$ to the Marčenko–Pastur law if X is in the domain of attraction of the normal distribution. Analogously to (4.7), the weak limit of $(F_{\mathbf{R}})$ provides a first idea what the limits of the extreme eigenvalues might be.

4.1.3 (X_{it}) identically distributed, but dependent

For practical purposes it is important to work with arbitrary population covariance matrices and not just $n^{-1}\mathbb{E}[\mathbf{X}\mathbf{X}'] = \mathbf{I}$. Based on well understood results in the iid case, numerous generalizations and estimation techniques have been developed. For many models the limiting spectral distribution can only be characterized in terms of an integral equation (=Marčenko–Pastur equation) for its Stieltjes transform. Explicit solutions are more involved; see the monographs [6, 5, 78]. Over the last couple of years siginificant progress on limiting spectral distributions for dependent time series was achieved; see for example [13, 12, 11]. Since the sample covariance matrix is a poor estimator for the population covariance matrix in high dimension, a different approach to the fundamental problem of estimating population eigenvalues is needed. In [33] the authors find that the bootstrap works for the top eigenvalues if they are sufficiently separated from the bulk. Among others, El Karoui [31] proposed to use the Marčenko–Pastur equation, which basically requires more insight into the empirical spectral distribution and its support. This was achieved in [28], where an algorithm for calculating the spectral distribution based on certain approximate integral equations for its Stieltjes transform was presented.

In view of [25, 24, 23] the behavior of the top eigenvalues is reasonably well understood in the case of linear dependence among the X_{it} and $\mathbb{E}[X^4] = \infty$. If $\mathbb{E}[X^4] < \infty$, similar arguments to (4.10) can be developed to show that methods for sample covariance matrices can be applied to sample correlation matrices; see for example [32]. Theorem 1 in [32] proves that if the spectral norm of the population correlation matrix is uniformly bounded and $\mathbb{E}[X^4(\log X)^{2+\varepsilon}] < \infty$, then the spectral properties of $\mathbf{R}$ and $n^{-1}\mathbf{X}\mathbf{X}'$ are asymptotically the same. In particular, if $\lambda_{(1)}/n \stackrel{\text{a.s.}}{\to} c$, then $\mu_{(1)} \stackrel{\text{a.s.}}{\to} c$.

For the sake of completeness we mention that the study of non-asymptotic high-dimensional sample covariance matrices was subject to an intense line of research in the last years. Good references are [67, 1, 2, 78].

4.1.4 About this paper

In Section 4.2 we introduce the basic assumptions of this paper and discuss their meaning. The main results are given in Section 4.3. We show that the limiting spectral distribution of the sample correlation matrices is the Marčenko–Pastur law (Theorem 4.3) and that the extreme eigenvalues converge a.s. to the endpoints of the limiting support (Theorem 4.5) provided $\mathbf{X}$ has iid entries such that their truncated variance is "almost slowly varying". In this sense, the limiting spectral distribution of sample correlation matrices is universal. A similar kind of universality holds for the limiting spectral distribution of sample covariance matrices given a finite variance, while the asymptotic behavior of their extreme eigenvalues is totally different if the fourth moment is infinite. Thus the eigenvalues of sample correlation matrices exhibit a "more robust" behavior than their sample covariance analogs. This is perhaps not surprising in view of the *self-normalizing property* of sample correlations. Self-normalization also has the advantage that one does not have to worry about the correct normalization. This is a crucial problem in the study of sample covariance matrices in the case of an infinite fourth moment where one needs a normalization stronger than the classical one. We conclude Section 4.3 with a small simulation study which shows that the asymptotic results work nicely.

We continue with some technical results in Section 4.4. These are of independent interest because they provide a *Path-Shortening Algorithm* for the calculation of bounds for the very high moments of $\mu_{(1)}$. We believe that this technique is novel and will be of further use for proving results in random matrix theory. The proofs of our main results

Theorems 4.5 and 4.3 are given in Sections 4.5 and 4.6, respectively. Both proofs heavily depend on the techniques developed in Section 4.4. We conclude with an Appendix which contains some auxiliary analytical results.

Condition (C_q) is crucial for the proof of Theorem 4.3. In Section 4.2 we discuss this condition and find out that it is very close to condition (4.15) which in turn is very close (but not equivalent) to membership of the distribution of X in the domain of attraction of the Gaussian law. We conjecture that the statement of Theorem 4.3 may be proved only under (4.15).

4.2 Assumptions

In this section we will present some distributional assumptions and discuss their meaning. We assume that (X_{it}) is an iid field with generic element X. Recall the notation

$$Y_{it} = \frac{X_{it}}{\sqrt{D_i}}\,, \quad i = 1, \dots, p;\ t = 1, \dots, n\,. \tag{4.14}$$

For ease of notation we will sometimes write $(Y_1, \dots, Y_n) = (Y_{11}, \dots, Y_{1n})$, $Y = Y_1$ and $D = D_1$.

4.2.1 Domain of attraction type-condition for the Gaussian law

One of the basic assumptions in this paper is

$$\mathbb{E}\big[Y_1 Y_2\big] = o(n^{-2}) \quad \text{and} \quad \mathbb{E}\big[Y_1^4\big] = o(n^{-1})\,, \qquad n \to \infty\,. \tag{4.15}$$

In [39] it was proved that condition (4.15) holds if the distribution of X is in the domain of attraction of the normal law, which is equivalent to $\mathbb{E}[X^2 \mathbf{1}_{\{|X| \le x\}}]$ being slowly varying.

The converse implication is not valid. Indeed, let $h(\cdot)$ be a positive function such that $0 < c_1 = \liminf_{x\to\infty} h(x) < \limsup_{x\to\infty} h(x) = c_2 < \infty$ and consider a symmetric random variable X with tail $\mathbb{P}(X > x) = \mathbb{P}(X < -x) = x^{-2} h(|x|)/2$ for x sufficiently large. Then we have

$$c_1 = \liminf_{x\to\infty} \frac{\mathbb{P}(|X| > x)}{x^2} < \limsup_{x\to\infty} \frac{\mathbb{P}(|X| > x)}{x^2} = c_2\,,$$

and therefore $\mathbb{E}[X^2 \mathbf{1}_{\{|X| \le x\}}]$ is not slowly varying, or, equivalently, the distribution of X is not in the domain of attraction of the normal law, but (4.15) is valid as a domination argument shows.

4.2.2 Condition (C_q)

This condition will be crucial for the proofs in this paper:

There exists a sequence $q = q_n \to \infty$ such that for some integer sequence $k = k_n$ with $k/\log n \to \infty$ we have $(k^3 q)/n \to 0$, and the moment inequality

$$\mathbb{E}[Y_1^{2m_1} \cdots Y_r^{2m_r}] \le \frac{q_n}{n}\, \mathbb{E}[Y_1^{2m_1} \cdots Y_{r-1}^{2m_{r-1}} Y_r^{2m_r - 2}] \tag{C_q}$$

holds for $1 \le r \le \ell - 1$ and any positive integers $m_1, \dots, m_r$ satisfying $m_1 + \cdots + m_r = \ell$, where $\ell \le k$.

Next we shed some light on this condition. It turns out to be closely related to (4.15). Indeed, assume (C_q). Iteration of (C_q) for any fixed ℓ yields

$$\mathbb{E}[Y_1^{2m_1}\cdots Y_r^{2m_r}] \le \big(\frac{q_n}{n}\big)^{\ell-r}\mathbb{E}[Y_1^2\cdots Y_r^2] \sim \frac{q_n^{\ell-r}}{n^\ell}, \quad n\to\infty\,.$$

In particular, $n\,\mathbb{E}[Y_1^4] \le q_n/n \le (\log n)^{-3}$. Thus, (C_q) provides some precise rate at which $n\,\mathbb{E}[Y_1^4]$ converges to zero.

Moreover, (C_q) does not hold if $\varepsilon = \liminf_{n\to\infty} n\,\mathbb{E}[Y_1^4] > 0$. If (C_q) were valid we would have for large n,

$$\varepsilon/2 \le n\,\mathbb{E}[Y_1^4] \le \frac{q_n}{n-1}\, n(n-1)\mathbb{E}[Y_1^2Y_2^2] \le \frac{q_n}{n-1} \to 0\,.$$

For example, Proposition 1 in [52] asserts that the distribution of X^2 is in the domain of attraction of an $\alpha/2$-stable distribution with $0<\alpha<2$ if and only if

$$\lim_{n\to\infty} n\,\mathbb{E}[Y_1^4] = 1-\frac{\alpha}{2}\,, \tag{4.16}$$

hence (C_q) does not hold if $|X|$ has a regularly varying tail with index $0<\alpha<2$.

The expectations in (C_q) can be calculated by using the following formula due to [39]:

$$\mathbb{E}[Y_1^{2m_1}\cdots Y_r^{2m_r}] = \frac{1}{(k-1)!}\int_0^\infty \lambda^{k-1}(\mathbb{E}[\mathrm{e}^{-\lambda X^2}])^{n-r}\prod_{j=1}^r \mathbb{E}[X^{2m_j}\,\mathrm{e}^{-\lambda X^2}]\,\mathrm{d}\lambda\,, \tag{4.17}$$

where $1\le r\le k$, $m_1+\cdots+m_r=k$ and $m_i\ge 1$.

We present some examples of distributions of X which satisfy (C_q).

Example 4.1 (Standard normal distribution)**.** Assume $X_i \sim N(0,1)$. We calculate $\mathbb{E}[Y_1^{2m_1}\cdots Y_r^{2m_r}]$ for the standard normal distibution via (4.17). Since X_1^2 has χ^2-distribution we know for $\lambda\ge 0$ that $\mathbb{E}[\mathrm{e}^{-\lambda X^2}] = (1+2\lambda)^{-1/2}$. We have

$$\frac{\mathrm{d}^m}{\mathrm{d}\lambda^m}\,\mathrm{e}^{-\lambda X^2} = (-X^2)^m\,\mathrm{e}^{-\lambda X^2}\,.$$

Calculation yields

$$(-1)^m\mathbb{E}[X^{2m}\,\mathrm{e}^{-\lambda X^2}] = (-1)^n(2m-1)!!(1+2\lambda)^{-1/2-m}\,. \tag{4.18}$$

By (4.17) and (4.18), we have for $\ell\le k$

$$\begin{aligned}\mathbb{E}[Y_1^{2m_1}\cdots Y_r^{2m_r}] &= \frac{1}{(\ell-1)!}\int_0^\infty \lambda^{\ell-1}(\mathbb{E}[\mathrm{e}^{-\lambda X^2}])^{n-r}\prod_{j=1}^r \mathbb{E}[X^{2m_j}\,\mathrm{e}^{-\lambda X^2}]\,\mathrm{d}\lambda\\ &= \frac{1}{(\ell-1)!}\int_0^\infty \lambda^{\ell-1}(1+2\lambda)^{-(n+2\ell)/2}\,\mathrm{d}\lambda\prod_{j=1}^r(2m_j-1)!!\,.\end{aligned}$$

Since

$$\int_0^\infty \lambda^{\ell-1}(1+2\lambda)^{-(n+2\ell)/2}\,\mathrm{d}\lambda = \frac{\Gamma(n/2)\Gamma(\ell)}{2^\ell\Gamma(n/2+\ell)}\,,$$

one obtains

$$\mathbb{E}[Y_1^{2m_1}\cdots Y_r^{2m_r}] = \frac{\Gamma(n/2)}{2^\ell\Gamma(n/2+\ell)}\prod_{j=1}^r(2m_j-1)!!\,, \tag{4.19}$$

which allows one to conclude that

$$\frac{\mathbb{E}[Y_1^{2m_1}\cdots Y_r^{2m_r}]}{\mathbb{E}[Y_1^{2m_1}\cdots Y_r^{2m_r-2}]}=\frac{2m_r-1}{n+2\ell-2}\le\frac{2k}{n}\,,$$

where we used $m_r \le \ell \le k$. Hence (C_q) holds with $q_n = 2k_n$.

Example 4.2 (Gamma distribution)**.** Assume $X^2 \sim \text{Gamma}(\alpha,\beta)$, $\alpha,\beta > 0$. In this case

$$\frac{\mathrm{d}^m}{\mathrm{d}\lambda^m}\mathbb{E}[\mathrm{e}^{-\lambda X^2}]=\frac{\mathrm{d}^m}{\mathrm{d}\lambda^m}\Big(1+\frac{\lambda}{\beta}\Big)^{-\alpha}=\frac{\Gamma(1-\alpha)}{\Gamma(1-\alpha-m)}\beta^{-n}\Big(1+\frac{\lambda}{\beta}\Big)^{-\alpha-n}.$$

For $\ell \le k$ one can calculate

$$\begin{aligned}\mathbb{E}[Y_1^{2m_1}\cdots Y_r^{2m_r}]&=\frac{1}{(\ell-1)!}\int_0^\infty \lambda^{\ell-1}(\mathbb{E}[\mathrm{e}^{-\lambda X^2}])^{n-r}\prod_{j=1}^r(-1)^{m_j}\frac{\mathrm{d}^{m_j}}{\mathrm{d}\lambda^{m_j}}\mathbb{E}[\mathrm{e}^{-\lambda X^2}]\,\mathrm{d}\lambda\\&=\frac{\Gamma(\alpha n)(-1)^\ell}{\Gamma(\alpha n+\ell)}\prod_{j=1}^r\frac{\Gamma(1-\alpha)}{\Gamma(1-\alpha-m_j)}\,.\end{aligned}$$

Similarly to the previous example, (C_q) holds with $q_n = (k_n+\alpha)/\alpha$.

4.3 Main results

Our first result identifies the limit of the empirical spectral distribution $F_{\mathbf{R}}$ of the sample correlation matrix $\mathbf{R}$ for iid random fields (X_{it}) with generic element X.

Theorem 4.3 (Limiting spectral distribution)**.** *Assume the condition* (G_γ)*.*

(1) If X is centered and (4.15) *holds then the sequence* $(F_{\mathbf{R}})$ *converges weakly to the Marčenko–Pastur law given in* (4.4)*.*

(2) If X is symmetric and (4.15) *does not hold, i.e.,* $\liminf_{n\to\infty} n\,\mathbb{E}[Y^4]>0$*, then*

$$\liminf_{n\to\infty}\mathbb{E}\Big[\int x^k F_{\mathbf{R}}(dx)\Big]>\beta_k(\gamma)\,,\qquad k\ge 1\,,$$

where $\beta_k(\gamma)$ is the k-th moment of the Marčenko–Pastur law given in (4.5)*.*

The proof of parts (1) and (2) will be given in Sections 4.6.1 and 4.6.2, repectively.

Remark 4.4. Part (1) with condition (4.15) replaced by $\mathbb{E}[X^2]<\infty$ was proved in [46]. Later, in [7] the finite variance assumption was replaced by the weaker condition that the distribution of X belongs to the domain of attraction of the normal law. We discussed in the previous section that (4.15) holds under the latter condition. Part (2) shows that (4.15) is the minimal condition for part (1). By Lemma B.1 in [6], $\lim_{n\to\infty}\mathbb{E}\big[\int x^k F_{\mathbf{R}}(dx)\big]=\beta_k(\gamma)\,, k\ge 1$, implies weak convergence of $F_{\mathbf{R}}$ to the Marčenko–Pastur distribution as the latter is uniquely determined by its moments $(\beta_k(\gamma))_{k\ge1}$.

If X is symmetric, $n\,\mathbb{E}[Y^4]=o(1)$ and $p/n\to 0$, a slight modification of the proof of part (2) combined with the method of moments yields $F_{\mathbf{R}}\to \mathbf{1}_{[1,\infty)}$ weakly. Consequently, for any $\varepsilon\in(0,1)$ the number of eigenvalues outside $(1-\varepsilon,1+\varepsilon)$ is $o(p)$ a.s. In particular, if p is fixed, then $\mu_{(1)}$ and $\mu_{(p)}$ converge to 1 a.s. In view of part (2), one concludes that $n\mathbb{E}[Y^4] = o(1)$ is a necessary and sufficient condition for the a.s. convergence of the eigenvalues $(\mu_{(i)})$ if X is symmetric and p fixed.

When $p \to \infty$ one has to deal with the potentially $o(p)$ eigenvalues outside the support of the limiting spectral distribution. We develop a method to overcome this problem at the expense of strengthening the assumption $n\,\mathbb{E}[Y^4] = o(1)$ to (C_q).

A Borel–Cantelli argument to obtain an upper bound for $\limsup_n \mu_{(1)}$ requires an adequate bound on $\mathbb{E}[\mu_{(1)}^{k_n}]$, where $k_n \to \infty$. To this end, we use the inequality

$$\mathbb{E}[\mu_{(1)}^{k_n}] \le \mathbb{E}[\operatorname{tr} \mathbf{R}^{k_n}] = \sum_{i_1,\ldots,i_{k_n}=1}^{p} \sum_{t_1,\ldots,t_{k_n}=1}^{n} \mathbb{E}[Y_{i_1 t_{k_n}} Y_{i_1 t_1} Y_{i_2 t_1} Y_{i_2 t_2} \cdots Y_{i_{k_n} t_{k_n-1}} Y_{i_{k_n} t_{k_n}}]$$

and determine those summands on the right-hand side which are largest when weighted by their multiplicities. Using our *Path-Shortening Algorithm*, which is a novel technique that efficiently uses the inherent structure of sample correlation matrices, their contribution is calculated explicitly. The other summands can –with considerable technical effort– be controlled by (C_q). Note that because of the identity $\mathbb{E}[\operatorname{tr} \mathbf{R}^{k_n}] = p\,\mathbb{E}\big[\int x^{k_n} F_{\mathbf{R}}(dx)\big]$ the behavior of moments of the empirical spectral distribution is closely linked to the above upper bound.

The following result provides general conditions for the a.s. convergence of the largest and smallest eigenvalues $\mu_{(1)}$ and $\mu_{(p)}$ of $\mathbf{R}$ to the endpoints of the Marčenko–Pastur law. The proof of this result is given in Section 4.5.

Theorem 4.5 (Limit of extreme eigenvalues)**.** *Assume* (G_γ).

(1) If $\mathbb{E}[X^4] < \infty$ *and* $\mathbb{E}[X] = 0$

(2) or X *is symmetric and satisfies condition* (C_q),

then

$$\mu_{(1)} \to (1+\sqrt{\gamma})^2 \quad \text{a.s.}, \tag{4.20}$$

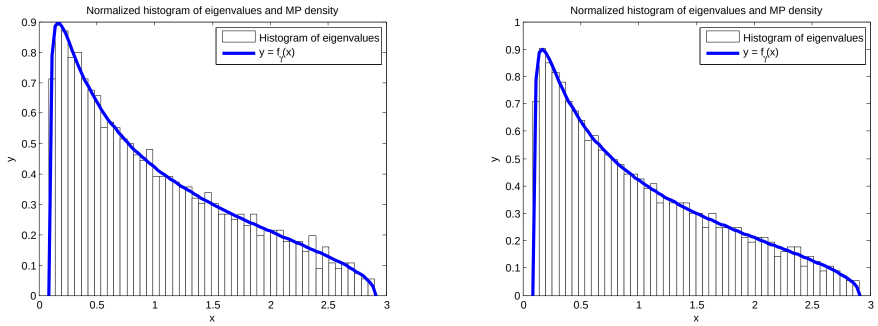
$$\mu_{(p)} \to (1-\sqrt{\gamma})^2 \quad \text{a.s.} \tag{4.21}$$

Remark 4.6. Part (1) was proved in [46, 77]; see the discussion in Section 4.1. Theorem 4.5 indicates that the a.s. convergence of the extreme eigenvalues of $\mathbf{R}$ does not depend on the finiteness of the fourth or even second moment. This is in stark contrast to the a.s. behavior of $n^{-1}\lambda_{(1)}$, the largest eigenvalue of the sample covariance matrix $n^{-1}\mathbf{X}\mathbf{X}'$. Note that there is a phase transition of the a.s. asymptotic behavior of the extreme eigenvalues at the border between finite and infinite fourth moment of X, while such a transition occurs for the empirical spectral distribution at the border between finite and infinite variance.

4.3.1 Simulation study

In this subsection we simulate a large data matrix $\mathbf{X}$ of iid entries. We compare the spectra of $\mathbf{X}\mathbf{X}'/n$ and $\mathbf{R}$ to the limiting Marčenko–Pastur spectral density with appropriate parameter γ; see Theorem 4.3. We simulate from different distributions of X and choose various values for p and n to cover Marčenko–Pastur distributions of several shapes. In what follows, we assume $\mathbb{E}[X^2] = 1$, whenever the second moment is finite.

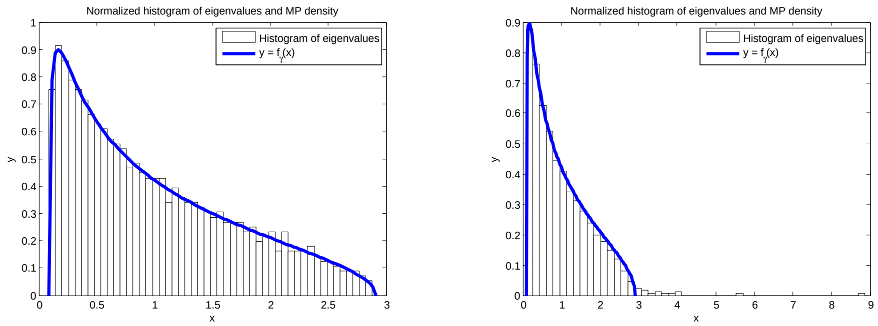
In Figure 4.1 we simulated a 1000×2000 data matrix $\mathbf{X}$ with iid entries drawn from a t_6-distribution which we renormalized to meet the requirement $\mathbb{E}[X^2] = 1$. To illustrate



(a) Sample correlation: $\mu_{(p)} = 0.086, \mu_{(1)} = 2.898$.

(b) Sample covariance: $\lambda_{(p)}/(n\mathbb{E}[X^2]) = 0.085, \lambda_{(1)}/(n\mathbb{E}[X^2]) = 2.908$.

Figure 4.1: Histogram and Marčenko–Pastur density for $X \sim t_6$, $n = 2000, p = 1000$. $\gamma = 0.5, (1-\sqrt{\gamma})^2 = 0.085, (1+\sqrt{\gamma})^2 = 2.914$.



(a) Sample correlation: $\mu_{(p)} = 0.088, \mu_{(1)} = 2.880$.

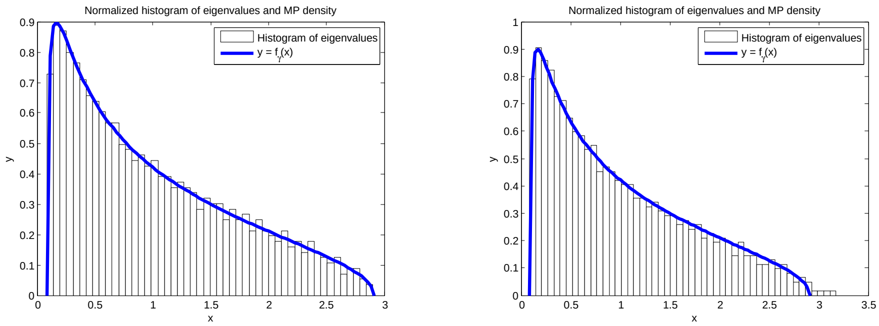
(b) Sample covariance: $\lambda_{(p)}/(n\mathbb{E}[X^2]) = 0.083$, $\lambda_{(1)}/(n\mathbb{E}[X^2]) = 8.870$.

Figure 4.2: Histogram and Marčenko–Pastur density for $X \sim t_3$, $n = 2000, p = 1000$. Here $\gamma = p/n = 0.5, (1-\sqrt{\gamma})^2 = 0.085, (1+\sqrt{\gamma})^2 = 2.914$.

the weak convergence of $(F_{\mathbf{R}})$ and $(F_{\mathbf{XX}'/n})$ we plot the histograms of the eigenvalues $(\mu_{(i)})$ and $(\lambda_{(i)}/n)$ and compare them to the Marčenko–Pastur distribution with $\gamma = 1/2$. As expected in the case $\mathbb{E}[X^4] < \infty$, the values $n^{-1}\lambda_{(1)} = 2.9086$ and $n^{-1}\lambda_{(p)} = 0.0855$ are very close to their theoretical almost sure limits 2.9142 and 0.0858, respectively. The same is valid for $\mu_{(1)}$ and $\mu_{(p)}$.

In Figure 4.2 we simulate X from a renormalized t_3-distribution with unit variance. The histograms of $(\mu_{(i)})$ and $(\lambda_{(i)}/n)$ resemble the corresponding Marčenko–Pastur density $f_{1/2}$. Note that $\lambda_{(1)}/n$ can be larger than the right endpoint $(1+\sqrt{\gamma})^2$ since it has a different limit behavior than in the case $\mathbb{E}[X^4] < \infty$, while $\mu_{(p)}$ and $\mu_{(1)}$ are close to the endpoints $(1-\sqrt{\gamma})^2$ and $(1+\sqrt{\gamma})^2$, respectively, for which Theorem 4.5 provides a formal justification.

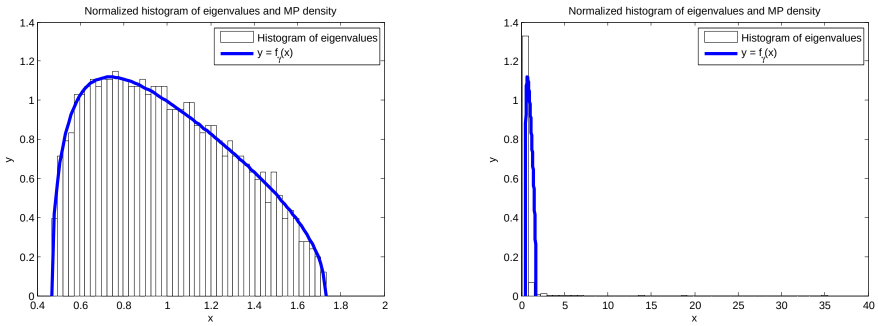
In Figures 4.3 and 4.4 we simulated from distributions with infinite fourth moment. We drew from a symmetrized Pareto distribution with parameter 3.99 to create the



(a) Sample correlation: $\mu_{(p)} = 0.086, \mu_{(1)} = 2.902$.

(b) Sample covariance: $\lambda_{(p)}/(n\mathbb{E}[X^2]) = 0.083, \lambda_{(1)}/(n\mathbb{E}[X^2]) = 3.176$.

Figure 4.3: Histogram and Marčenko–Pastur density: $X \stackrel{\mathrm{d}}{=} Z_1 - Z_2$ for $Z_i \sim \text{Pareto}(3.99)$, $n = 2000, p = 1000$. Here $\gamma = p/n = 0.5, (1-\sqrt{\gamma})^2 = 0.085, (1+\sqrt{\gamma})^2 = 2.914$.



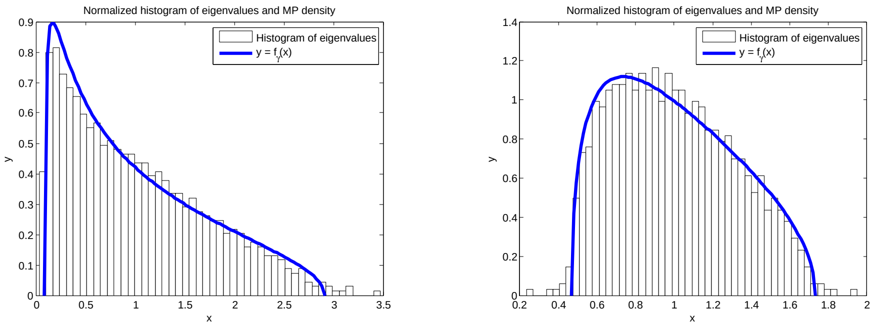
(a) Sample correlation: $\mu_{(p)} = 0.469, \mu_{(1)} = 1.731$.

(b) Sample covariance: $\lambda_{(p)}/(n\mathbb{E}[X^2]) = 0.159, \lambda_{(1)}/(n\mathbb{E}[X^2]) = 35.319$.

Figure 4.4: Histogram and Marčenko–Pastur density for $X \sim t_{2.1}$, $n = 10000, p = 1000$. Here $\gamma = p/n = 0.1, (1-\sqrt{\gamma})^2 = 0.467, (1+\sqrt{\gamma})^2 = 1.732$.

plots in Figure 4.3. Note that in this case $\mathbb{E}[X^{3.99}] = \infty$, while $\mathbb{E}[X^{3.99-\varepsilon}] < \infty$ for any $\varepsilon > 0$, i.e., we are at the "border" between finite and infinite fourth moment. The extreme eigenvalues in the sample correlation case are very close to their theoretical limits stated in Theorem 4.5, whereas the largest eigenvalues of the sample covariance matrix cease to lie within the support of the Marčenko–Pastur distribution. Note that the assumption $\mathbb{E}[X^2] = 1$ is superfluous for the sample correlation plots due to self-normalization. For the histogram of $(\lambda_{(i)}/n)$ the knowledge of the correct value $\mathbb{E}[X^2]$ is crucial since, for instance, $\lambda_{(1)}/n \to (1+\sqrt{\gamma})^2\mathbb{E}[X^2]$ a.s. In applications, $\mathbb{E}[X^2]$ needs to be estimated first and estimation errors might significantly alter the conclusion. One can easily imagine that Figure 4.1(b) with a misspecified variance of the data could have resembled Figure 4.3(b). In this respect sample correlations are more robust.

In Figure 4.4, we choose X from the standardized $t_{2.1}$-distribution, moving closer to the infinite variance case. The histogram of $(\mu_{(i)})$ fits the Marčenko–Pastur density very



(a) $X \overset{\mathrm{d}}{=} Z^2 - \mathbb{E}Z^2$ for $Z \sim t_{1.5}$, $n = 2000, p = 1000, \gamma = 0.5$.

(b) $X \sim t_{1.8}$, $n = 10000, p = 1000, \gamma = 0.1$

Figure 4.5: Histogram of $(\mu_{(i)})$ and Marčenko–Pastur density

well and the extreme eigenvalues are located in a close proximity of the endpoints of the Marčenko–Pastur support. The sample covariance case in (b) does not look particularly appealing due to the fact that there are a few relatively large eigenvalues. For example, $\lambda_{(1)}/n = 35.3196$ while $(1+\sqrt{\gamma})^2$ is only 1.7325. By [4, 41, 23], the properly normalized $\lambda_{(1)}$ converges to a Fréchet distributed random variable. The correct normalization is roughly $n^{4/2.1}$ and hence it is expected that $\lambda_{(1)}/n$ is separated from the bulk, whose behavior ultimately determines the limiting spectral distribution, which is the Marčenko–Pastur law with parameter $\gamma = 0.1$. However, due to the separation between the top eigenvalues and the bulk, it is not obvious from a histogram with only 50 classes that the Marčenko–Pastur law provides a good fit to the spectral distribution in (b). This different behavior of sample correlations and covariances is an additional argument for the higher stability of results obtained from an analysis of the sample correlation matrix.

Finally, we present two histograms of $(\mu_{(i)})$ with $\mathbb{E}[X^2] = \infty$ in Figure 4.5. In (a), we choose the non-symmetric $X \overset{\mathrm{d}}{=} Z^2 - \mathbb{E}Z^2$ for $Z \sim t_{1.5}$. In (b), the simulated X is standardized $t_{1.8}$. The plots look surprisingly stable, given that the empirical spectral distribution does not weakly converge to the Marčenko–Pastur law; see Theorem 4.3(2). The extreme eigenvalues $\mu_{(1)}$ and $\mu_{(p)}$ are much further away from $(1-\sqrt{p/n})^2$ and $(1+\sqrt{p/n})^2$, respectively, than in all the other sample correlation histograms we have seen so far.

4.3.2 A remark on the centered sample correlation matrix

We presented results for the matrices $\mathbf{R}$ and $\mathbf{XX}'$, assuming that $\mathbb{E}[X] = 0$ when $\mathbb{E}[|X|] < \infty$. In practice, the expectation of X typically has to be estimated. We discuss what has to be changed in the aforementioned theory in this case. We consider the matrix $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}'$, where

$$\widetilde{X}_{it} = X_{it} - \overline{X}_i \quad \text{and} \quad \overline{X}_i = \frac{1}{n}\sum_{t=1}^{n} X_{it}\,.$$

and the corresponding correlation matrix $\widetilde{\mathbf{R}} = \widetilde{\mathbf{F}}^{1/2}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}'\widetilde{\mathbf{F}}^{1/2}$, where $\widetilde{\mathbf{F}}$ is the $p \times p$ diagonal matrix with entries

$$\widetilde{F}_{ii} = \frac{1}{(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}')_{ii}}\,, \quad i = 1, \ldots, p.$$

In contrast to (4.10) an application of Weyl's inequality [6] yields

$$n^{-1}|\lambda_{(1)}(\mathbf{X}\mathbf{X}') - \lambda_{(1)}(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}')| \le n^{-1}\|\mathbf{X}\mathbf{X}' - \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}'\|_2\,, \tag{4.22}$$

where, in general, the right-hand side does not converge to zero. However, since $\mathbf{X} - \widetilde{\mathbf{X}}$ is a rank 1 matrix, it is known from [6] that $n^{-1}\mathbf{X}\mathbf{X}'$ and $n^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}'$ share the same limiting spectral distribution (if it exists) with right endpoint b say. Therefore we have

$$\liminf_{n\to\infty} \frac{\lambda_{(1)}(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}')}{n} \ge b\,.$$

Following [32], we let $\mathbf{H} = \mathbf{I}_n - n^{-1}\mathbf{1}\mathbf{1}'$, where $\mathbf{1} = (1, \ldots, 1)'$. Then one can write $\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{H}$ and since $\mathbf{H}$ is a symmetric matrix with $(n-1)$ eigenvalues equal to 1 and one eigenvalue equal to 0 we see that

$$\lambda_{(1)}(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}') \le \lambda_{(1)}(\mathbf{X}\mathbf{X}')\,.$$

We conclude

$$\lim_{n\to\infty} \frac{\lambda_{(1)}(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}')}{n} = b \quad \text{a.s.}$$

whenever $\lambda_{(1)}(\mathbf{X}\mathbf{X}')/n \to b$ a.s. Therefore the a.s. behavior of the largest eigenvalues of $\mathbf{X}\mathbf{X}'$ and $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}'$ are closely related.

Due to the shift and scale invariance of sample correlations, the aforementioned arguments remain valid for the ordered eigenvalues

$$\widetilde{\mu}_{(1)} \ge \cdots \ge \widetilde{\mu}_{(p)}$$

of $\widetilde{\mathbf{R}}$ if $\mathbb{E}[X^4] < \infty$ and $\mathbb{E}[X] = c$ (not necessarily zero), as shown in [46]. Then we have $\widetilde{\mu}_{(1)} \to (1+\sqrt{\gamma})^2$ a.s. and $\widetilde{\mu}_{(p)} \to (1-\sqrt{\gamma})^2$ a.s. In Theorem 2 of [46] it was proven that if $\mathbb{E}[X^2] < \infty$ and $p/n \to \gamma \in (0, \infty)$, the empirical spectral distribution of $\widetilde{\mathbf{R}}$ converges weakly to the Marčenko–Pastur law.

4.4 Technical results

In this section we provide most technical results required for the proofs of the main theorems. We develop a new approach which efficiently uses the structure of sample correlation matrices. The goal of this section is to prove Proposition 4.11.

Throughout (X_{it}) are iid symmetric. We will study the moments

$$\sum_{t_1,\ldots,t_k=1}^{n} \mathbb{E}[Y_{i_1t_k}Y_{i_1t_1}Y_{i_2t_1}Y_{i_2t_2}Y_{i_3t_2}Y_{i_3t_3}\cdots Y_{i_kt_{k-1}}Y_{i_kt_k}].$$

for $k \ge 1$ and various choices of *paths* $I = (i_1, i_2, \ldots, i_k) \in \{1, \ldots, p\}^k$. In this case, $\text{length}(I) = k$ is the *length of the path.* We say that a *path* $(i_1, i_2, \ldots, i_k)$ is an *r-path* if it contains exactly r distinct components. A path is *canonical* if $i_1 = 1$ and $i_l \le$

$\max\{i_1,\dots,i_{l-1}\}+1, l \geq 2$. A canonical r-path satisfies $\{i_1, i_2, \dots, i_k\} = \{1,\dots,r\}$. Two paths are *isomorphic* if one becomes the other by a suitable permutation on $(1,\dots,p)$. Each *isomorphism class* contains exactly one canonical path. For $k \geq 1$, define

$$f(I,T) = \mathbb{E}[Y_{i_1 t_k} Y_{i_1 t_1} Y_{i_2 t_1} Y_{i_2 t_2} Y_{i_3 t_2} Y_{i_3 t_3} \cdots Y_{i_k t_{k-1}} Y_{i_k t_k}]\,, \quad I,T \in \{1,\dots,k\}^k\,.$$

Finally, define $F(\emptyset) = n$ and

$$F(i_1,\dots,i_k) = F_n(i_1,\dots,i_k) = \sum_{t_1,\dots,t_k=1}^{n} f((i_1,\dots,i_k),(t_1,\dots,t_k))\,.$$

Note that $F(I_1) = F(I_2)$ if I_1, I_2 lie in the same isomorphism class. Therefore, whenever we are interested in $F(I)$ we can assume without loss of generality that I is canonical.

In what follows, we will consider transformations of the path I leading to a new path $S(I)$. For ease of notation, we will also assume $S(I)$ canonical. If it is not canonical, we can always work with its *canonical representative*, the unique canonical path in its isomorphism class.

When calculating values of F, the path-shortening function PS will be useful. Let $I = (i_1,\dots,i_k) \in \{1,\dots,k\}^k$. $PS(I)$ is the output of the following algorithm.

Path-Shortening Algorithm $PS(I)$.

Input: Path $I = (i_1,\dots,i_k)$. Set $J = I$ and $R = 0, \text{runs} = 0$.

Step 0: Set $l = \text{length}(I)$. Go to Step 1.

Step 1: Erase runs.

- If $i_j = i_{j+1}$ for some $1 \leq j \leq l$, where we interpret i_{l+1} as i_1, erase element i_j from the path. Set $I = (i_1,\dots,i_{j-1},i_{j+1},\dots,i_l)$, runs $=$ runs $+1$ and return to Step 0.
- Otherwise proceed with Step 2.

Step 2: Let R_1 be the number of elements of the path I which appear exactly once. Set $R := R + R_1$. Then define I to be the resulting (possibly shorter) path which is obtained by deleting those R_1 elements from the path I. Go to Step 3.

Step 3:
- If $J = I$, then return (I, R, runs) as output.
- If $J \neq I$, set $J := I$ and return to Step 0.

Definition 4.7. *The path-shortening function PS is the output $(S(I), R(I), \text{runs}(I))$ of the Path-Shortening Algorithm (PSA) where $S(I)$ is the resulting shortened path and $R(I)$ is the total number of elements that were removed in Step 2 of the PSA. We write $PS(I) = (S(I), R(I), \text{runs}(I))$.*

Properties of $PS(I)$.

Clearly, $\text{length}(S(I)) \leq \text{length}(I)$. If $I = (1,\dots,r)$ then $S(I) = \emptyset$, which shows that $S(I)$ can have length zero. Furthermore, all elements in $S(I)$ appear at least twice. If I is an r-path then $R(I) \leq r$.

Lemma 4.8. *For any $I \in \{1,\dots,k\}^k$, we have $F(I) = F(S(I))\, n^{-R(I)}$.*

Proof. We shall look at the changes made to I in Steps 1 and 2 of the PSA separately. *Assume we are in Step 1.*

- If $i_j = i_{j+1}$ for some $1 \le j \le l$, where we interpret i_{l+1} as i_1, erase element i_j from the path. Set $S_1(I) = (i_1, \dots, i_{j-1}, i_{j+1}, \dots, i_l)$.
- Otherwise, $S_1(I) = I$.

Since Step 1 does not influence the value of R it suffices to show $F(I) = F(S_1(I))$. If $S_1(I) = I$ there is nothing to show. Therefore assume $i_j = i_{j+1}$ for some j. In this case, we have

$$\begin{aligned} F(I) &= \sum_{t_1,\dots,t_{j-1},t_{j+1},\dots,t_k=1}^{n} \mathbb{E}[Y_{i_1t_k}Y_{i_1t_1}\cdots Y_{i_{j-1}t_{j-2}}Y_{i_{j-1}t_{j-1}}Y_{i_jt_{j-1}}\sum_{t_j=1}^{n} Y_{i_jt_j}^2 \\ &\qquad Y_{i_jt_{j+1}}Y_{i_{j+2}t_{j+1}}Y_{i_{j+2}t_{j+2}}\cdots Y_{i_kt_{k-1}}Y_{i_kt_k}] \\ &= \sum_{t_1,\dots,t_{j-1},t_{j+1},\dots,t_k=1}^{n} \mathbb{E}[Y_{i_1t_k}Y_{i_1t_1}\cdots Y_{i_{j-1}t_{j-2}}Y_{i_{j-1}t_{j-1}}Y_{i_jt_{j-1}}Y_{i_jt_{j+1}} \\ &\qquad Y_{i_{j+2}t_{j+1}}Y_{i_{j+2}t_{j+2}}\cdots Y_{i_kt_{k-1}}Y_{i_kt_k}] = F(S_1(I)), \end{aligned} \tag{4.23}$$

where we used $\sum_{t=1}^n Y_{it}^2 = 1$. This proves that Step 1 poses no problem.

Next we turn to Step 2. Without loss of generality we can assume that I does not contain any runs. If all elements of I appear at least twice there is nothing to prove. Therefore assume the jth element i_j appears only once and $R_1 = 1$. Let $S_2(I)$ denote the path I with the jth element removed. Thus we have to show $F(I) = F(S_2(I))n^{-1}$. In this case, we have

$$\begin{aligned} F(I) &= \sum_{t_1,\dots,t_{j-1},t_{j+1},\dots,t_k=1}^{n} \sum_{t_j=1}^{n} \mathbb{E}[Y_{i_jt_{j-1}}Y_{i_jt_j}] \\ &\qquad \times\mathbb{E}[Y_{i_1t_k}Y_{i_1t_1}\cdots Y_{i_{j-1}t_{j-2}}Y_{i_{j-1}t_{j-1}}Y_{i_{j+1}t_j}Y_{i_{j+1}t_{j+1}}\cdots Y_{i_kt_{k-1}}Y_{i_kt_k}] \\ &= \sum_{t_1,\dots,t_{j-1},t_{j+1},\dots,t_k=1}^{n} n^{-1}\mathbb{E}[Y_{i_1t_k}Y_{i_1t_1}\cdots Y_{i_{j-1}t_{j-2}}Y_{i_{j-1}t_{j-1}}Y_{i_{j+1}t_{j-1}} \\ &\qquad \times Y_{i_{j+1}t_{j+1}}\cdots Y_{i_kt_{k-1}}Y_{i_kt_k}] \\ &= F(S_2(I))n^{-1}. \end{aligned} \tag{4.24}$$

Here we used that $t_{j-1} = t_j$ is necessary for $\mathbb{E}[Y_{i_jt_{j-1}}Y_{i_jt_j}]$ to be non-zero. If $R_1 > 1$ we can apply the above argument iteratively to obtain $F(I) = F(S_2 \circ \cdots \circ S_2(I))n^{-R_1}$. The proof is complete. ☐

Define for $k \ge 1$, a function g by $g(\emptyset) = 1$ and

$$g(I) = \max_{T\in\{1,\dots,k\}^k} \{|T| : f(I,T) > 0\}, \qquad I \in \{1,\dots,k\}^k. \tag{4.25}$$

From now, on we assume the I-paths to be canonical.

Lemma 4.9. *Let I be a canonical r-path of length k. For any $T \in \{1,\dots,k\}^k$ such that $f(I,T) > 0$ we have $|T| \le k - r + 1$.*

Proof. Without loss of generality we may assume that T is canonical. We shall sometimes refer to the t_i's as t-indices. In the beginning one should think of the t-indices as pairwise distinct whenever possible. Their actual values are not relevant for the value $f(I,T)$. In all cases, except $r=1$, there are certain t-indices that have to coincide such that $f(I,T)$ can be positive: $t_i = t_j$ for some $i,j,\ i \neq j$. This is due to the symmetry of X. We will see that in some cases these i,j are not unique. More precisely, it may happen that there is a set $\{t_{i_1}, t_{i_2}, \ldots, t_{i_{2a}}\}$ with $i_1, \ldots, i_{2a}$ distinct such that $|\{t_{i_1}, t_{i_2}, \ldots, t_{i_{2a}}\}| \le a$ is necessary for $f(I,T) > 0$. In these cases, the cardinality of T is less than k provided $f(I,T) > 0$.

We start with the two simplest cases. If $r = 1$, we have $f(I,T) > 0$ for any T, and hence $|T| \le k = k - r + 1$. Moreover, if $r = k$, we have $f(I,T) > 0$ if and only if $t_1 = \cdots = t_k$, and hence $|T| = 1 = k - r + 1$.

Now we assume $1 < r < k$. Our arguments will rely on the proof of Lemma 4.8. Clearly, $1 \le g(I) \le k$. From the definition of runs in the PSA we have $\text{runs}(I) \le k - r$. From (4.23) and (4.24) one infers

$$\text{runs}(I) + 1 \le g(I) \le k - R(I)\,. \tag{4.26}$$

First, we analyze paths I with $S(I) = \emptyset$, or equivalently $\text{length}(S(I)) = 0$. This implies $\text{runs}(I) = k - r$; otherwise the path-shortening function stops earlier and $S(I) \neq \emptyset$. Therefore, we get from the identity

$$g(I) = g(S(I)) + \text{runs}(I) \tag{4.27}$$

that $g(I) = k - r + 1$, which finishes the proof in the case $S(I) = \emptyset$. Note that (4.27) holds for all I. This follows from the proof of Lemma 4.8.

Next, assume $S(I) \neq \emptyset$. In this case, we immediately see

$$\text{length}(S(I)) = k - R(I) - \text{runs}(I)\,. \tag{4.28}$$

Since each element in $S(I)$ has to appear at least twice and $r \ge 2$ we have $\text{length}(S(I)) \ge 4$. Moreover, $S(I)$ has $r - R(I) \ge 2$ distinct components. As a consequence, it must hold

$$\text{runs}(I) \le k - r - (r - R(I)) = k - 2r + R(I)\,. \tag{4.29}$$

In view of (4.27), we have to bound $g(S(I))$. Without loss of generality $S(I)$ may be assumed canonical: there is exactly one canonical path in the isomorphism class of $S(I)$ and every path in an isomorphism class has the same g-value. If, for example, $S(I)$ happens to be $(3,4,3,4)$, then we will work with the canonical representative $(1,2,1,2)$. Write $S(I) = (s_1, \ldots, s_{\text{length}(S(I))})$. Since $S(I)$ is canonical, we have

$$\{s_1, \ldots, s_{\text{length}(S(I))}\} = \{1, \ldots, r - R(I)\}\,.$$

For $i = 1, \ldots, r - R(I)$ define $N_i := |\{1 \le j \le \text{length}(S(I)) : s_j = i\}|$. If we now let L_i be the set of all u such that (i, t_u) appears as an index in $f(S(I),T)$, then $|L_i| = 2N_i$. Finally, define

$$T_i := \{t_j \,:\, j \in L_i\} \quad \text{and} \quad \widetilde{T}_i := (t_j \,:\, j \in L_i)\,, \quad i = 1, \ldots, r - R(I)\,.$$

For example, if $I = (1,2,1,2,3,3)$, then $k = 6, r = 3$, $N_1 = N_2 = 2$ and we have $(S(I), R(I), \text{runs}(I)) = ((1,2,1,2),1,1)$ and $L_1 = L_2 = \{1,2,3,4\}$.

By construction, $f(S(I),T)$ can only be positive if $|T_i| \le N_i$. More precisely every t-index in the vector $\widetilde{T}_i$ needs to coincide with at least 1 other t-index of this vector.

Otherwise, $\mathbb{E}\big[\prod_{u\in L_i} Y_{i,t_u}\big]=0$ which would imply $f(S(I),T)=0$. The quantity $g(S(I))$ is the maximum number of distinct t-indices such that $f(S(I),T)>0$. Hence, there can be at most length$(\widetilde{T}_i)/2$ distinct t-indices in $\widetilde{T}_i$. Since each t_j appears exactly twice in $(\widetilde{T}_1,\ldots,\widetilde{T}_{r-R(I)})$,

$$g(S(I)) \le 0.5 \text{ length}(S(I))\,. \tag{4.30}$$

Now we are ready to finish the proof of the lemma. By (4.27), (4.30), (4.28), (4.29), in this order, one obtains

$$\begin{aligned} g(I) = g(S(I)) + \text{runs}(I) &\le \frac{\text{length}(S(I)}{2} + \text{runs}(I)\\ &= \frac{k-R(I)-\text{runs}(I)}{2} + \text{runs}(I)\\ &\le \frac{k-R(I)+k-2r+R(I)}{2} = k-r\,. \end{aligned}$$

□

Remark 4.10. The above proof reveals that $g(I)=k-r+1$ if and only if $S(I)=\emptyset$. For r-paths I of length k with $S(I)\neq\emptyset$, the bound $g(I)\le k-r$ is sharp. Consider for instance $I=(1,2,1,2)$, where

$$f(I,T) = (\mathbb{E}[Y_{t_1}Y_{t_2}Y_{t_3}Y_{t_4}])^2\,.$$

From this relation it is easily deduced that the only canonical representatives $T=(t_1,t_2,t_3,t_4)$ leading to $f(I,T)>0$ are $(1,1,2,2),(1,2,1,2),(1,2,2,1)$ and $(1,1,1,1)$. The first three of them have the highest number of distinct values. We conclude $g(I)=2$. In general, the canonical paths T for which the maximum in (4.25) is attained are not unique, whenever $g(I)\le k-r$. On the other hand, if $g(I)=k-r+1$ there exists exactly one canonical $(k-r+1)$-path T of length k for which the maximum is obtained. This is an immediate consequence of the above proofs. In [6], Bai and Silverstein present a way to describe this T.

For a canonical r-path I of length k let

$$d(I) = k-r+1-g(I)\,. \tag{4.31}$$

The function d satisfies $0\le d(I)\le k-r$ and $d(S(I))=d(I)$. The set of canonical r-paths of length k, denoted by $\mathcal{I}_{r,k}$, can be written as a disjoint union

$$\mathcal{I}_{r,k} = \bigcup_{u=0}^{k-r} \mathcal{I}_{r,k}(u)\,,$$

where $\mathcal{I}_{r,k}(u)$ contains those I with $d(I)=u$.

Lemma 3.4 in [6] determines the cardinality of $\mathcal{I}_{r,k}(0)$: for $k\in\mathbb{N}$ and $r\le k$,

$$|\mathcal{I}_{r,k}(0)| = \frac{1}{r}\binom{k}{r-1}\binom{k-1}{r-1}\,. \tag{4.32}$$

Proposition 4.11. *Assume condition (C_q). Then the following statements hold for any r-path I of length $k\ge 1$ and $1\le r\le k$:*

(1) If $S(I)=\emptyset$, then $F(I)=n^{1-r}$.

(2) In general, we have

$$F(I) \leq 2\, n^{1-r-d(I)}(2k)^{d(I)} q^{d(I)}\,. \tag{4.33}$$

Proof. $S(I) = \emptyset$ is equivalent to $R(I) = r$. By Lemma 4.8,

$$F(I) = F(S(I))n^{-R(I)} = F(\emptyset)n^{-r} = n^{1-r}\,.$$

Therefore, we only have to prove (4.33) for paths I with $d(I) \geq 1$. Without loss of generality we assume $S(I)$ is a canonical $(r-R(I))$-path. We use the notation for paths with $S(I) \neq \emptyset$ developed in the proof of Lemma 4.9. We know that

$$S(I) = (\pi_1, \ldots, \pi_{\mathrm{length}(S(I))})\,,$$

where $\pi_1, \ldots, \pi_{\mathrm{length}(S(I))}$ is a permutation of the path

$$I_0 = (\underbrace{1, \ldots, 1}_{N_1}, \underbrace{2, \ldots, 2}_{N_2}, \ldots, \underbrace{r-R(I), \ldots, r-R(I)}_{N_{r-R(I)}})\,.$$

Clearly, $I_0 \in \mathcal{I}_{r-R(I),\mathrm{length}(S(I))}(0) = \mathcal{I}_{r-R(I),k-R(I)-\mathrm{runs}(I)}(0)$. By Lemma 4.9,

$$g(I_0) = (k - R(I) - \mathrm{runs}(I)) - (r - R(I)) + 1 = k - r - \mathrm{runs}(I) + 1$$

and by definition of the function $d(\cdot)$,

$$g(S(I)) = k - r - \mathrm{runs}(I) + 1 - d(S(I) = k - r - \mathrm{runs}(I) + 1 - d(I)\,.$$

The main idea will be to compare $F(S(I))$ to $F(I_0)$. Both of them are sums of expressions of the type

$$\prod_{i=1}^{r-R(I)} \mathbb{E}\Big[Y_{i1}^{2m_{i,1}} \cdots Y_{is_i}^{2m_{i,s_i}}\Big] = \prod_{i=1}^{r-R(I)} \mathbb{E}\Big[Y_1^{2m_{i,1}} \cdots Y_{s_i}^{2m_{i,s_i}}\Big]\,, \tag{4.34}$$

where for all $i = 1, \ldots, r-R(I)$, $1 \leq s_i \leq N_i$, $m_{i,j} \geq 1$ for all $j \geq 1$ and $m_{i,1}+\cdots+m_{i,s_i} = N_i$. We write

$$\mathbf{s} = (s_1, \ldots, s_{r-R(I)}) \quad \text{and} \quad \mathbf{m}_i = (m_{i,1}, \ldots, m_{i,s_i})\,, \quad i = 1, \ldots, r-R(I)\,. \tag{4.35}$$

Observe that in

$$F(I_0) = \sum_{t_1,\ldots,t_{N_1+\cdots+N_{r-R(I)}}=1}^{n} \mathbb{E}\Big[Y_{t_{N_1+\cdots+N_{r-R(I)}}} Y_{t_1}^2 \cdots Y_{t_{N_1-1}}^2 Y_{t_{N_1}}\Big] \cdots$$
$$\mathbb{E}\Big[Y_{t_{N_1+\cdots+N_{r-R(I)-1}}} Y^2_{t_{N_1+\cdots+N_{r-R(I)-1}+1}} \cdots Y^2_{t_{N_1+\cdots+N_{r-R(I)}-1}} Y_{,t_{N_1+\cdots+N_{r-R(I)}}}\Big]$$

the non-zero summands have to satisfy $t_{N_1} = t_{N_2} = \cdots = t_{N_1+\cdots+N_{r-R(I)}}$. Hence, the above sum is effectively a sum only over $g(I_0)$ t-indices. The point we want to stress is that there is never a choice, in the sense that even though there are $g(I_0)$ distinct t-indices, something like $t_{N_1} = t_1 \neq t_2 = t_{N_1+N_2}$ is never possible. The reason is that the associated canonical $g(I_0)$-path for the t-indices is unique.

For $S(I) \neq \emptyset$, however, the associated canonical $g(S(I))$-path for the t-indices is not unique, as mentioned in Remark 4.10. Depending on the sets L_i there are several possibilities. For instance, for $S(I) = (1,2,1,2)$ we have $L_1 = L_2 = \{1,2,3,4\}$, $d(I) = 1$

and $\text{length}(S(I)) = 4$. To produce a positive summand one needs $|\{t_1, t_2, t_3, t_4\}| \le 2$ with every t-index appearing at least twice. In this case, t_1 has to take the same value as one of the other three t-indices. Then there are two t-indices left which all have to appear at least twice. In this specific example, there are three canonical paths of t-indices which are listed in Remark 4.10 above.

We are interested in the general case. How many distinct canonical $g(S(I))$-paths T of length $\text{length}(S(I))$ with $f(S(I), T) > 0$ can exist? With the reasoning which lead to (4.30) one can show that this number is at most $(2d(I)+1)!!$. This bound is attained if $N_1 = N_2 = \text{length}(S(I))/2$ which implies $L_1 = L_2 = \{1, \dots, \text{length}(S(I)\}$.

For our purpose we will use a much larger bound, namely

$$(2d(I)+1)!! = (2d(I)+1)(2d(I)-1)\cdots 3 \le (2d(I)+1)^{d(I)} \le (2k)^{d(I)} . \tag{4.36}$$

Now let us compare $F(S(I))$ and $F(I_0)$, which look very similar at first sight. The main difference is the dimension of the index sets in the summation. While the sum for $F(I_0)$ contains $n^{g(I_0)}$ positive elements, the sum for $F(S(I))$ has at most $(2k)^{d(I)} n^{g(I_0)-d(I)}$ positive elements. Let $Q_{S(I)}$ denote the set of all canonical T for which the maximum in (4.25) is attained. By the above considerations, $|Q_{S(I)}| \le (2k)^{d(I)}$. Each element in $Q_{S(I)}$ corresponds to a different configuration of t-indices in $F(S(I))$, i.e., it tells us which t-indices have to be equal. Therefore, we have

$$F(S(I)) \le \sum_{Q \in Q_{S(I)}} F_Q(S(I)) , \tag{4.37}$$

where $F_Q(S(I))$ is defined as follows. Write $Q = (q_1, \dots, q_{\text{length}(S(I))})$. By construction, $\{q_1, \dots, q_{\text{length}(S(I))}\} = \{1, \dots, g(S(I))\}$. Set $K_j = \{1 \le i \le \text{length}(S(I)) : q_i = j\}$. Then

$$F_Q(S(I)) = \sum_{\substack{t_1,\dots,t_{\text{length}(S(I))}=1 \\ t_l = t_m \, \forall l,m \in K_j \,, 1 \le j \le g(S(I))}}^{n} f(S(I), (t_1, \dots, t_{\text{length}(S(I))})) . \tag{4.38}$$

We will show later that

$$F_Q(S(I)) \le 2\, q^{d(I)} n^{-d(I)} F(I_0) , \quad Q \in Q_{S(I)} . \tag{4.39}$$

Then it follows from (4.37) and (4.39) that

$$\begin{aligned} F(S(I)) &\le \sum_{Q \in Q_{S(I)}} F_Q(S(I)) \\ &\le (2k)^{d(I)} 2\, q^{d(I)} n^{-d(I)} F(I_0) = 2\,(2k)^{d(I)} q^{d(I)} n^{-d(I)} n^{1-r+R(I)} . \end{aligned}$$

Finally, an application of Lemma 4.8 gives

$$F(I) = n^{-R(I)} F(S(I)) \le 2\,(2k)^{d(I)} q^{d(I)} n^{-d(I)} n^{1-r} ,$$

which completes the proof.

Next, we show (4.39) by matching each of the $n^{g(I_0)-d(I)}$ positive summands in (4.38) with $n^{d(I)}$ of the $n^{g(I_0)}$ positive summands in $F(I_0)$, where we recall that

$$F(I_0) = \sum_{\substack{t_1,\dots,t_{\text{length}(S(I))}=1 \\ t_{N_1} = t_{N_2} = \cdots = t_{N_1 + \cdots + N_{r-R(I)}}}}^{n} f(I_0, (t_1, \dots, t_{\text{length}(S(I))})) . \tag{4.40}$$

By *matching* we mean the following. Assume we want to prove

$$\sum_{i=1}^{n} A_i \le \sum_{j=1}^{m} B_j \tag{4.41}$$

for nonnegative A_i, B_i and $m \ge n$. If for every $i = 1, \ldots, n$ there exists a $j_i \in \{1, \ldots, m\}$ such that $A_i \le B_{j_i}$ and the j_i's are distinct, then (4.41) holds. In this case, we say that each A_i is matched by some B_{j_i}.

We say that $f(S(I), (t_1, \ldots, t_{\mathrm{length}(S(I))}))$ and $f(I_0, (t_1, \ldots, t_{\mathrm{length}(S(I))}))$ are in class $y = \sum_{i=1}^{r-R(I)} s_i$ if they can be written in the form

$$\prod_{i=1}^{r-R(I)} \mathbb{E}\big[Y_1^{2m_{i,1}} \cdots Y_{s_i}^{2m_{i,s_i}}\big]\,. \tag{4.42}$$

By construction, y takes values in the set $\{r - R(I), \ldots, \mathrm{length}(S(I))\}$. A summand in class y is fully determined by the vector $(\mathbf{s}, \mathbf{m}_1, \ldots, \mathbf{m}_{r-R(I)}) =: (\mathbf{s}, \mathbf{m})$; see (4.35) for this notation. Hence, we call this summand of type $(\mathbf{s}, \mathbf{m})$ and denote it $f_{\mathbf{s},\mathbf{m}}$. Note that the class y is comprised of all elements of type $(\mathbf{s}, \mathbf{m})$ such that $\sum_{i=1}^{r-R(I)} s_i = y$ and $\mathbf{m}$ satisfies the restriction stated below equation (4.34).

Let $\mathcal{T}_0(y)$ and $\mathcal{T}_Q(y)$ be index sets which contain the exact type of all summands (counted with multiplicity) of class y in (4.40) and (4.38), respectively. As mentioned before, we must have

$$\sum_{y=r-R(I)}^{\mathrm{length}(S(I))} |\mathcal{T}_0(y)| = n^{g(I_0)} \quad \text{and} \quad \sum_{y=r-R(I)}^{\mathrm{length}(S(I))} |\mathcal{T}_Q(y)| = n^{g(I_0)-d(I)}\,.$$

With this notation we can write

$$2\,F(I_0) = 2 \sum_{y=r-R(I)}^{\mathrm{length}(S(I))} \sum_{(\mathbf{s},\mathbf{m})\in\mathcal{T}_0(y)} f_{\mathbf{s},\mathbf{m}}\,, \tag{4.43}$$

$$n^{d(I)}\,F_Q(S(I)) = n^{d(I)} \sum_{y=r-R(I)}^{\mathrm{length}(S(I))} \sum_{(\mathbf{s},\mathbf{m})\in\mathcal{T}_Q(y)} f_{\mathbf{s},\mathbf{m}}\,. \tag{4.44}$$

We show (4.39) by a *matching argument*. We start by matching summands in class $\mathrm{length}(S(I))$. From (4.42) we see that elements of class $\mathrm{length}(S(I))$ are necessarily of the form

$$\prod_{i=1}^{r-R(I)} \mathbb{E}\big[Y_1^2 \cdots Y_{N_i}^2\big]\,,$$

in other words they are all equal. Note that

$$|\mathcal{T}_0(\mathrm{length}(S(I)))| = n(n-1)\cdots(n-g(I_0)+1)\,,$$
$$|\mathcal{T}_Q(\mathrm{length}(S(I)))| = n(n-1)\cdots(n-g(I_0)+d(I)+1)\,.$$

Therefore, we have

$$n^{d(I)} \sum_{(\mathbf{s},\mathbf{m})\in\mathcal{T}_Q(\mathrm{length}(S(I)))} f_{\mathbf{s},\mathbf{m}} \le 2 \sum_{(\mathbf{s},\mathbf{m})\in\mathcal{T}_0(\mathrm{length}(S(I)))} f_{\mathbf{s},\mathbf{m}}\,.$$

This shows that for each summand in class length$(S(I))$ on the left-hand side of (4.39) we can find at least one summand of the same type on the right-hand side of (4.39).

Since a large number of summands of the class length$(S(I))$ have not been used for matching of summands from the same class, we can use (C_q) to match them with summands of classes $\text{length}(S(I))-1, \text{length}(S(I))-2, \ldots, \max(\text{length}(S(I))-d(I), r-R(I))$.

Applying (C_q) to a summand of class length$(S(I))-1$ we obtain that it is bounded by q times a summand in class length$(S(I))$. Hence, we can perform the matching in (4.39) also between different classes. Clearly, for $y \in \{r-R(I)+1, \ldots, \text{length}(S(I))\}$ the index set $\mathcal{T}_0(y)$ is much larger than $\mathcal{T}_0(y-1)$. In fact, we have $|\mathcal{T}_0(y)| = n\, c\, |\mathcal{T}_0(y-1)|$ for some constant $c>1$. Note that $y \mapsto |\mathcal{T}_Q(y)|$ is not a strictly increasing function since some $\mathcal{T}_Q(y)$ can be empty.

The matching is performed as follows: first match the class length$(S(I))$ summands on the left-hand side of (4.39). Then match the class length$(S(I))-1$ summands on the left-hand side of (4.39) with the remaining class length$(S(I))$ summands on the right-hand side which have not been used for the matching yet.

Let $r-R(I) \le u \le \text{length}(S(I))$. The general strategy is to match class u summands on the left-hand side with class $u, \ldots, \min(u+d(I), \text{length}(S(I)))$ summands on the right-hand side. During the matching one tries to use the (still available) class $\min(u+d(I), \text{length}(S(I)))$ summands on the right-hand side first, then turns to class $\min(u+d(I), \text{length}(S(I)))-1$, and so forth. Whenever a matching between different classes is performed an application of (C_q) is necessary to ensure that the expression on the left-hand side is bounded by whatever we have matched it with on the right-hand side. This leads to powers of q and since $q^{d(I)}$ is the highest possible power we have explained the factor $q^{d(I)}$ in (4.39).

Note that the factor 2 in (4.39) is there to guarantee

$$|\mathcal{T}_Q(\text{length}(S(I)))| < 2\, n^{-d(I)}\, |\mathcal{T}_0(\text{length}(S(I)))|$$

for sufficiently large n, but it is of no central importance.

The last step in the procedure is the matching of the summands with the highest possible powers on the left-hand side of (4.39), which appear when all t-indices are equal. They are elements of the class $r-R(I)$. We have

$$|\mathcal{T}_0(r-R(I))| = |\mathcal{T}_Q(r-R(I))| = n\,,$$

which is a simple explanation why matching of (4.43) and (4.44) with summands in the same class cannot work in general. Using (C_q) $d(I)$ times, we can bound class $r-R(I)$ summands by class $r-R(I)+d(I)$ summands of which we originally have $|\mathcal{T}_0(r-R(I))| \approx n^{d(I)} |\mathcal{T}_0(r-R(I))|$, which explains the factor $n^{d(I)}$ in (4.39). The general matching strategy applies and the proof of (4.39) is complete. □

4.5 Proof of Theorem 4.5

The following proposition contains our main technical novelty. Its proof is given after the proof of Theorem 4.5.

Proposition 4.12. *Assume (G_γ) and that the iid symmetric field (X_{it}) satisfies (C_q). Then the following limit results hold for the largest and smallest eigenvalues $\mu_{(1)}$ and*

$\mu_{(p)}$ *of* $\mathbf{R}$*:*

$$\limsup_{n\to\infty} \mu_{(1)} \quad \le \quad (1+\sqrt{\gamma})^2 \quad \text{a.s.} \tag{4.45}$$

$$\liminf_{n\to\infty} \mu_{(p)} \quad \ge \quad (1-\sqrt{\gamma})^2 \quad \text{a.s.} \tag{4.46}$$

Proof of Theorem 4.5. (1) If $\mathbb{E}[X^4] < \infty$, (4.20) and (4.21) hold for any mean zero distribution as seen in (4.11).
(2) Now assume (C_q). The convergence of $F_{\mathbf{R}}$ to a deterministic distribution supported on a compact interval implies that the number of the eigenvalues outside this interval is $o(p)$. Since the right and left endpoints of the Marčenko–Pastur law are $(1+\sqrt{\gamma})^2$ and $(1-\sqrt{\gamma})^2$, respectively, we conclude from Theorem 4.3(1) that

$$\liminf_{n\to\infty} \mu_{(1)} \ge (1+\sqrt{\gamma})^2 \quad \text{a.s.} \quad \text{and} \quad \limsup_{n\to\infty} \mu_{(p)} \le (1-\sqrt{\gamma})^2 \quad \text{a.s.}\,;$$

see [6] for details. Together with Proposition 4.12 this completes the proof. □

Proof of equation (4.45) in Proposition 4.12

Following [38], we prove (4.45) by showing

$$\sum_{n=1}^{\infty} \mathbb{E}\Big[\Big(\frac{\mu_{(1)}}{z}\Big)^k\Big] < \infty\,, \tag{4.47}$$

where $z > (1+\sqrt{\gamma})^2$ and $k = k_n$ satisfies $k/\log n \to \infty$ and $(k^3 q)/n \to 0$, which exists by condition (C_q). We use that $\mathbb{E}[\mu_{(1)}^k] \le \mathbb{E}[\operatorname{tr}(\mathbf{R})^k]$ and

$$\begin{aligned}\mathbb{E}[\operatorname{tr}(\mathbf{R})^k] &= \sum_{i_1,\ldots,i_k=1}^{p} \sum_{t_1,\ldots,t_k=1}^{n} \mathbb{E}[Y_{i_1t_k}Y_{i_1t_1}Y_{i_2t_1}Y_{i_2t_2}Y_{i_3t_2}Y_{i_3t_3}\cdots Y_{i_kt_{k-1}}Y_{i_kt_k}]\\ &= \sum_{i_1,\ldots,i_k=1}^{p} F(i_1,\ldots,i_k)\,.\end{aligned}$$

We rewrite $\mathbb{E}[\operatorname{tr}(\mathbf{R})^k]$ by sorting according to the number of distinct components in the path $(i_1,\ldots,i_k)$. Any r-path of length k is an element in the disjoint union $\mathcal{J}_{r,k}(0) \cup \cdots \cup \mathcal{J}_{r,k}(k-r)$, where $\mathcal{J}_{r,k}(u)$ is the set of all r-paths I of length k with $d(I) = u$; see (4.31) for the definition of $d(I)$. Hence we have

$$\{1,\ldots,p\}^k = \bigcup_{r=1}^{k} \bigcup_{u=0}^{k-r} \mathcal{J}_{r,k}(u)\,. \tag{4.48}$$

Given a path $I \in \mathcal{J}_{r,k}(u)$ we can look at the positions where the r distinct components appear for the first time. There are r such positions. The first such position is always 1, in general i_1 can take p different values. For the second such position there are $(p-1)$ possibilities; the original p minus the one from the first position. In total there are $p(p-1)\cdots(p-r+1)$ ways to assign values to these r positions. For this reason

$$|\mathcal{J}_{r,k}(u)| = p(p-1)\cdots(p-r+1)|\mathcal{I}_{r,k}(u)|\,, \tag{4.49}$$

where $\mathcal{I}_{r,k}(u)$ is the set of all canonical r-paths I of length k with $d(I) = u$. The only difference between the definitions of $\mathcal{J}_{r,k}(u)$ and $\mathcal{I}_{r,k}(u)$ is that the elements of the latter are canonical. Note that $\mathcal{I}_{k,k}(u) = \emptyset$ for all $u \ge 1$.

In view of (4.48) and (4.49) we obtain

$$\begin{aligned}\mathbb{E}[\mathrm{tr}(\mathbf{R})^k] &= \sum_{r=1}^{k}\sum_{u=0}^{k-r}\sum_{I\in\mathcal{J}_{r,k}(u)} F(I)\\ &= \sum_{r=1}^{k} p(p-1)\cdots(p-r+1)\sum_{u=0}^{k-r}\sum_{I\in\mathcal{I}_{r,k}(u)} F(I) \qquad (4.50)\\ &\le \sum_{r=1}^{k} p^r \sum_{I\in\mathcal{I}_{r,k}(0)} F(I) + \sum_{r=1}^{k-1} p^r \sum_{u=1}^{k-r} |\mathcal{I}_{r,k}(u)| \max_{I\in\mathcal{I}_{r,k}(u)} F(I) =: S_1 + S_2\,.\end{aligned}$$

By Proposition 4.11, (4.32) and since $|\mathcal{I}_{r,k}(0)| \le \binom{k-1}{r-1}^2$, we have

$$S_1 \le \sum_{r=1}^{k} p^r \binom{k-1}{r-1}^2 n^{1-r} = p\sum_{r=1}^{k}\binom{k-1}{r-1}^2\Big(\frac{p}{n}\Big)^{r-1}. \qquad (4.51)$$

Next we bound S_2. Consider $1 \le u \le k-r$. We will see how elements of $\mathcal{I}_{r,k}(u)$ can be constructed by modifying elements of $\mathcal{I}_{r,k}(0)$. Let $I \in \mathcal{I}_{r,k}(u)(N_1,\dots,N_r)$ be the subset of $\mathcal{I}_{r,k}(u)$ for whose elements the integer i appears exactly N_i times as a component. Here $N_i, i = 1,\dots,r$ are positive integers satisfying $N_1 + \cdots + N_r = k$. Obviously it is possible to obtain I by permuting the components of any $I_0 \in \mathcal{I}_{r,k}(0)(N_1,\dots,N_r)$. Consider the following permutation of I_0: two components of I_0 exchange places, all other remain untouched. We denote such a *switching permutation* by SP. The number of such permutations is bounded by $k^2/2$. Indeed, the first component can switch places with the remaining $k-1$ components, the second with $k-2$ components, etc. In total there are

$$(k-1)+(k-2)+\cdots+1 = \sum_{j=1}^{k-1} j = \frac{(k-1)k}{2} \le \frac{k^2}{2}$$

ways how two components can switch positions.

Let $u = 1$. For any $I \in \mathcal{I}_{r,k}(u)(N_1,\dots,N_r)$ there exists at least one path $I_0 \in \mathcal{I}_{r,k}(0)(N_1,\dots,N_r)$ and a switching permutation SP such that $I = SP(I_0)$. Here SP and I_0 are in general not unique. This is a consequence of the proof of Lemma 4.9. This implies

$$|\mathcal{I}_{r,k}(u)| \le |\mathcal{I}_{r,k}(0)|\frac{k^2}{2}\,.$$

Similarly, for $1 \le u \le k-r$ and $I \in \mathcal{I}_{r,k}(u)(N_1,\dots,N_r)$ there exist $I_0 \in \mathcal{I}_{r,k}(0)(N_1,\dots,N_r)$ and switching permutations $SP_1,\dots,SP_u$ such that $I = SP_1\circ\cdots\circ SP_u(I_0)$, which shows

$$|\mathcal{I}_{r,k}(u)| \le |\mathcal{I}_{r,k}(0)|\Big(\frac{k^2}{2}\Big)^u. \qquad (4.52)$$

Now we are ready to bound S_2. From Proposition 4.11 we get

$$\max_{I\in\mathcal{I}_{r,k}(u)} F(I) \le 2n^{1-r-u}(2k)^u q^u$$

and therefore,

$$\begin{aligned}
S_2 &\le \sum_{r=1}^{k-1} p^r \sum_{u=1}^{k-r} \binom{k-1}{r-1}^2 \Big(\frac{k^2}{2}\Big)^u 2n^{1-r-u}(2k)^u q^u \\
&= p\sum_{r=1}^{k-1} \binom{k-1}{r-1}^2 \Big(\frac{p}{n}\Big)^{r-1} 2\sum_{u=1}^{k-r} \Big(\frac{k^3 q}{n}\Big)^u \\
&\le p\sum_{r=1}^{k-1} \binom{k-1}{r-1}^2 \Big(\frac{p}{n}\Big)^{r-1} 2\Big[\Big(1-\frac{k^3 q}{n}\Big)^{-1} - 1\Big].
\end{aligned}$$

Finally, we have the bound

$$\begin{aligned}
\mathbb{E}[\mathrm{tr}(\mathbf{R})^k] \le S_1 + S_2 &\le p\sum_{r=1}^{k} \binom{k-1}{r-1}^2 \Big(\frac{p}{n}\Big)^{r-1} \Big(1 + 2\Big[\Big(1-\frac{k^3 q}{n}\Big)^{-1} - 1\Big]\mathbf{1}_{\{r<k\}}\Big) \\
&\le p\sum_{r=1}^{k} \binom{2k-2}{2r-2} \Big(\frac{p}{n}\Big)^{r-1} \Big(1 + 2\Big[\Big(1-\frac{k^3 q}{n}\Big)^{-1} - 1\Big]\mathbf{1}_{\{r<k\}}\Big) \\
&\le p\sum_{r=0}^{2k-2} \binom{2k-2}{r} \Big(\sqrt{\frac{p}{n}}\Big)^r \Big(2\Big(1-\frac{k^3 q}{n}\Big)^{-1} - 1\Big)^{2k-2-r} \\
&= \Big[p^{1/(k-1)}\Big(2\Big(1-\frac{k^3 q}{n}\Big)^{-1} - 1 + \sqrt{\frac{p}{n}}\Big)^2\Big]^{k-1} \le \eta^k ,
\end{aligned}$$

where η is a constant satisfying $(1+\sqrt{\gamma})^2 < \eta < z$. The last inequality follows from $p^{1/(k-1)} \to 1$ and

$$\lim_{n\to\infty} \Big(2\Big(1-\frac{k^3 q}{n}\Big)^{-1} - 1 + \sqrt{\frac{p}{n}}\Big)^2 = (1+\sqrt{\gamma})^2 .$$

This shows (4.47) which concludes the proof.

4.5.1 Proof of equation (4.46) in Proposition 4.12

We start with the following result.

Proposition 4.13. *Assume* (G_γ). *If the iid entries* (X_{it}) *are symmetric and satisfy condition* (C_q) *then*

$$\limsup_{n\to\infty} \|\mathbf{R} - (1+\gamma)\mathbf{I}\|_2 \le 2\sqrt{\gamma} \quad \text{a.s.} \tag{4.53}$$

Proof. The general idea is the same as in the proof of equation (4.45) in Proposition 4.12: we will bound the spectral norm of $\mathbf{R} - (1+\gamma)\mathbf{I}$ by the trace of high powers of this matrix and then take an appropriate root. To this end we choose an integer sequence $k = k_n \to \infty$ such that $k/\log n \to \infty$ and $(k^3 q)/n \to 0$, which exists by condition (C_q). Since the matrices $\mathbf{R}$ and $(1+\gamma)\mathbf{I}$ commute we have

$$(\mathbf{R} - (1+\gamma)\mathbf{I})^{2k} = \sum_{i=0}^{2k} \binom{2k}{i} \mathbf{R}^i (-1)^i (1+\gamma)^{2k-i}\mathbf{I} .$$

By linearity of the trace,

$$\mathbb{E}[\operatorname{tr}(\mathbf{R}-(1+\gamma)\mathbf{I})^{2k}] = p\,(1+\gamma)^{2k}\Big[1+p^{-1}\sum_{i=1}^{2k}\binom{2k}{i}\Big(\frac{-1}{1+\gamma}\Big)^{i}\mathbb{E}[\operatorname{tr}\mathbf{R}^i]\Big]. \tag{4.54}$$

From (4.50) combined with (4.32) we know that for n sufficiently large

$$\mathbb{E}[\operatorname{tr}\mathbf{R}^i] \ge p\sum_{r=1}^{i}\frac{(p-1)(p-2)\cdots(p-r+1)}{n^{r-1}}\frac{1}{r}\binom{i}{r-1}\binom{i-1}{r-1} = p\,\beta_i(\gamma)\,(1-\delta_n)\,, \tag{4.55}$$

where $\delta_n = O(1/n)$. Additionally, we established

$$\mathbb{E}[\operatorname{tr}\mathbf{R}^i] \le p\,\beta_i(\gamma)\,\Big(1+\frac{2k^3q}{n}\Big)(1+\delta_n)\,. \tag{4.56}$$

Hence, by (4.54), Lemma 4.17, and noting that f_k is continuous on $\mathbb{R}$, and $p/n\to\gamma\in(0,1]$, we have for n sufficiently large,

$$\begin{aligned}\mathbb{E}[\operatorname{tr}(\mathbf{R}-(1+\gamma)\mathbf{I})^{2k}] &= p(1+\gamma)^{2k}\Big[1+\sum_{i=1}^{2k}\binom{2k}{i}\Big(\frac{-1}{1+\gamma}\Big)^{i}\beta_i(\gamma)\Big]\big(1+O(2k^3q_n/n)\big)\\ &= p(1+\gamma)^{2k}f_k(\gamma)\big(1+O(2k^3q_n/n)\big)\\ &\le p\,(1+\gamma)(4\gamma)^k\big(1+O(2k^3q_n/n)\big) < z^{2k}\,,\end{aligned}$$

for any $z>2\sqrt{\gamma}$. The last inequality follows from

$$\lim_{n\to\infty} p^{1/(2k)}\big(1+2k^3q_n/n\big)^{1/(2k)}(1+\gamma)^{1/(2k)} = 1\,.$$

Using the same Borel-Cantelli argument as in the proof of (4.45), one obtains the desired relation

$$\limsup_{n\to\infty}\|\mathbf{R}-(1+\gamma)\mathbf{I}\|_2 \le 2\sqrt{\gamma}\quad \text{a.s.}$$

□

With Proposition 4.13 we can finish the proof of (4.46). We have

$$\|\mathbf{R}-(1+\gamma)\mathbf{I}\|_2 = \max\{\mu_{(1)}-(1+\gamma), -\mu_{(p)}+(1+\gamma)\}\,.$$

From (4.53) we conclude

$$\begin{aligned}\limsup_{n\to\infty}\mu_{(1)} &\le 2\sqrt{\gamma}+1+\gamma = (1+\sqrt{\gamma})^2\quad \text{a.s.}\,,\\ \liminf_{n\to\infty}\mu_{(p)} &\ge -2\sqrt{\gamma}+1+\gamma = (1-\sqrt{\gamma})^2\quad \text{a.s.}\end{aligned}$$

4.6 Proof of Theorem 4.3

4.6.1 Proof of Theorem 4.3(1)

We appeal to the proof of Theorem 2.3 in [7]. The following lemma is a version of Corollary 1.1 in [7].

Lemma 4.14. *Let $\mathbf{B} = \mathbf{B}_n = (B_{jk)}$ be a non-random $n \times n$ matrix with bounded norm and*

$$\mathcal{S} = \{(i_1, j_1, i_2, j_2) : 1 \le i_1, j_1, i_2, j_2 \le n\} \\ \setminus \{(i_1, j_1, i_2, j_2) : i_1 = i_2, j_1 = j_2 \text{ or } i_1 = j_2 \ne i_2 = j_1\}.$$

If $\mathbb{E}[Y^4] = o(n^{-1})$,

$$n \operatorname{var}(Y_1 Y_2) \to 0, \tag{4.57}$$

$$V_n = n^2 \sum_{\mathcal{S}} \big(\operatorname{cov}(Y_{i_1} Y_{j_1}, Y_{i_2} Y_{j_2})\big)^2 \to 0, \tag{4.58}$$

then Condition 1 of Theorem 1.1 in [7] holds, i.e.,

$$\mathbb{E}[|\mathbf{Y}_0 \mathbf{B} \mathbf{Y}_0' - \operatorname{tr}(\mathbf{B}\, \mathbb{E}[\mathbf{Y}_0 \mathbf{Y}_0'])|^2] = o(1),$$

where $\mathbf{Y}_0 = (Y_1, \dots, Y_n)$.

Proof. We have for some constant $c > 0$,

$$\begin{aligned}
&\mathbb{E}[|\mathbf{Y}_0 \mathbf{B} \mathbf{Y}_0' - \operatorname{tr}(\mathbf{B} \mathbb{E}[\mathbf{Y}_0 \mathbf{Y}_0'])|^2] \\
&= \mathbb{E}\Big[\Big| \sum_{i_1, j_1 = 1}^{n} B_{i_1 j_1} (Y_{i_1} Y_{j_1} - \mathbb{E}[Y_{i_1} Y_{j_1}])\Big|^2\Big] \\
&= \sum_{i_1, j_1 = 1}^{n} \sum_{i_2, j_2 = 1}^{n} B_{i_1 j_1} B_{i_2 j_2} \operatorname{cov}(Y_{i_1} Y_{j_1}, Y_{i_2} Y_{j_2}) \\
&\le c\, \big[n \operatorname{var}(Y_1^2) + n \operatorname{var}(Y_{11} Y_{12})\big] + \sum_{\mathcal{S}} B_{i_1 j_1} B_{i_2 j_2} \operatorname{cov}(Y_{i_1} Y_{j_1}, Y_{i_2} Y_{j_2}).
\end{aligned}$$

By assumption, $n \operatorname{var}(Y^2) = n\, (\mathbb{E}[Y^4] - n^{-2}) \to 0$. The second summand converges to zero by (4.57). It is shown in [7] that the last summand is bounded by

$$c\, n \Big(\sum_{\mathcal{S}} \big(\operatorname{cov}(Y_{i_1} Y_{j_1}, Y_{1 i_2} Y_{1 j_2})\big)^2 \Big)^{1/2}$$

which converges to zero by (4.58). □

Remark 4.15. Lemma 4.14 corrects the proof of Theorem 2.3 and Corollary 1.1 in [7]. In the latter paper it is claimed that

$$V_n' = n^2 \sum_{\mathcal{S}'} \big(\operatorname{cov}(Y_{i_1} Y_{j_1}, Y_{i_2} Y_{j_2}\big)^2 \to 0,$$

where

$$\mathcal{S}' = \{(i_1, j_1, i_2, j_2) : 1 \le i_1, j_1, i_2, j_2 \le n\} \\ \setminus \{(i_1, j_1, i_2, j_2) : i_1 = i_2 \ne j_1 = j_2 \text{ or } i_1 = j_2 \ne i_2 = j_1\}.$$

However, $\mathcal{S}'$ contains the quadruples (i, i, i, i). Hence

$$V_n' \ge np^2 \big(\operatorname{var}(Y^2)\big)^2 = n^{-1} p^2\, (n\, \mathbb{E}[Y^4])^2 - 2\, \frac{p^2}{n^2}\, (n\, \mathbb{E}[Y^4]) + \frac{p^2}{n^3},$$

which does not necessarily converge to zero since $n\, \mathbb{E}[Y^4]$ may converge to zero arbitrarily slowly.

Now we are ready for the proof of Theorem 4.3(1). If the distribution of X is in the domain of attraction of the normal law the claim follows from Theorem 2.3 in [7], using our Lemma 4.14.

Now assume the alternative condition (4.15). We will apply Theorem 2.2 in [7] and our Lemma 4.14. Our goal is to find the limiting spectral distribution of $\mathbf{R} = \mathbf{Y}\mathbf{Y}'$ via the limit of the Stieltjes transform of $\mathbf{Y}'\mathbf{Y}$, using the fact that $\mathbf{Y}\mathbf{Y}'$ and $\mathbf{Y}'\mathbf{Y}$ have the same non-zero eigenvalues. Since $\lambda_{(i)} = 0$ for any of these matrices whenever $i > n \vee p$ we obtain a connection between the two spectral distributions:

$$F_{\mathbf{Y}'\mathbf{Y}} = \Big(1 - \frac{p}{n}\Big)\mathbf{1}_{[0,\infty)} + \frac{p}{n} F_{\mathbf{Y}\mathbf{Y}'} \,.$$

Hence

$$\begin{aligned} s_{\mathbf{R}}(z) &= \int \frac{1}{x-z} \,\mathrm{d}F_{\mathbf{R}}(x) \\ &= \int \frac{1}{x-z} \,\mathrm{d}\Big(\frac{n}{p}F_{\mathbf{Y}'\mathbf{Y}} - \Big(\frac{n}{p} - 1\Big)\mathbf{1}_{[0,\infty)}\Big)(x) \\ &= \frac{n}{p} s_{\mathbf{Y}'\mathbf{Y}}(z) - \Big(\frac{n}{p} - 1\Big)\frac{1}{-z}\,, \quad z \in \mathbb{C}^+\,, \end{aligned} \tag{4.59}$$

where we used that for a constant $c \neq 0$ we have $s_{c\mathbf{A}}(z) = c^{-1} s_{\mathbf{A}}(cz)$.

We introduce the $n \times n$ matrix $\mathbf{T} = (T_{ij}) = (p\,\mathbb{E}[Y_i Y_j])$ which is a a circulant matrix whose eigenvalues can be determined as $T_{11} + (n-1)T_{12}$ and $T_{11} - T_{12}$ where the latter appears with multiplicity $n-1$. By assumption (4.15), we have $T_{ij} = o(n^{-1})$ for $i \neq j$ and hence $\|\mathbf{T}\|_2$ is bounded. The empirical spectral distribution

$$F_{\mathbf{T}}(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{\{\lambda_j(\mathbf{T}) \le x\}} = \frac{\mathbf{1}_{\{T_{11}+(n-1)T_{12} \le x\}}}{n} + \frac{n-1}{n}\mathbf{1}_{\{T_{11}-T_{12} \le x\}}$$

converges to the degenerate distribution H_γ with all mass at $\lim_{n\to\infty}(T_{11} - T_{12}) = \lim_{n\to\infty} p/n = \gamma$.

Next we verify the assumptions of Lemma 4.14. We have

$$\begin{aligned} n\,\mathbb{E}[\mathrm{var}(Y_1Y_2)] &= n\left(\mathbb{E}[(Y_1Y_2)^2] - (\mathbb{E}[Y_1Y_2])^2\right) \\ &\le n\Big(\frac{1}{n(n-1)} - o(n^{-2})\Big) \to 0\,, \qquad n \to \infty\,. \end{aligned}$$

This implies (4.57).

Now we turn to V_n in (4.58). If we distinguish between the types of indices in $\mathcal{S}$ we find that either possible structure for the summands $(Y_{i_1}Y_{j_1} - \mathbb{E}[Y_{i_1}Y_{j_1}])(Y_{i_2}Y_{j_2} - \mathbb{E}[Y_{i_2}Y_{j_2}])$ in V_n is of the type $Y_1^3Y_2, Y_1^2Y_2Y_3$ or $Y_1Y_2Y_3Y_4$. Keeping this in mind, we conclude that for some constant $c > 0$,

$$\begin{aligned} V_n &\le c\Big(n^4\left(\mathrm{cov}(Y_1^2, Y_1Y_2)\right)^2 + n^5\left(\mathrm{cov}(Y_1^2, Y_2Y_3)\right)^2 + n^5\left(\mathrm{cov}(Y_1Y_2, Y_2Y_3)\right)^2 \\ &\quad + n^6\left(\mathrm{cov}(Y_1Y_2, Y_3Y_4)\right)^2\Big) \\ &\le c\Big(n^4\left(\mathbb{E}[Y_1^3Y_2] - (1/n)\mathbb{E}[Y_1Y_2]\right)^2 + n^5\left(\mathbb{E}[Y_1^2Y_2Y_3] - (1/n)\mathbb{E}[Y_1Y_2]\right)^2 \\ &\quad + n^5\left(\mathbb{E}[Y_1Y_2^2Y_3] - (\mathbb{E}[Y_1Y_2])^2\right)^2 + n^6\left(\mathbb{E}[Y_1Y_2Y_3Y_4] - (\mathbb{E}[Y_1Y_2])^2\right)^2\Big)\,. \end{aligned}$$

The right-hand side converges to zero in view of assumption (4.15) and because (see [39])

$$\mathbb{E}[Y_1^3 Y_2] = O(n^{-2}), \quad \mathbb{E}[Y_1^2 Y_2 Y_3] = O(n^{-3}) \quad \text{and} \quad \mathbb{E}[Y_1 Y_2 Y_3 Y_4] = O(n^{-4})\,.$$

Applications of Theorem 2.2 in [7] and our Lemma 4.14 yield for $s = \lim_{n\to\infty} s_{\mathbf{Y}'\mathbf{Y}}$,

$$\begin{aligned} s(z) &= \int \frac{1}{\omega(1-\gamma^{-1}-\gamma^{-1}zs(z)) - z}\, \mathrm{d}H_\gamma(\omega) \\ &= \frac{1}{\gamma(1-\gamma^{-1}-\gamma^{-1}zs(z)) - z}\,, \end{aligned}$$

Thus $s = s(z)$ is the solution of the quadratic equation

$$s^2 z + s(1+z-\gamma) + 1 = 0.$$

By convention of [6], the square root of a complex number is the one with a positive imaginary part. Hence

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

Writing m for the limiting Stieltjes transform of $F_{\mathbf{Y}\mathbf{Y}'}$, we conclude from (4.59) and since $n/p \to \gamma^{-1}$ that

$$\begin{aligned} m(z) &= \gamma^{-1}s(z) + \frac{\gamma^{-1}-1}{z} \\ &= \frac{1-\gamma-z+\sqrt{(1+\gamma-z)^2-4\gamma}}{2\gamma z}\,, \end{aligned}$$

which we recognize as the Stieltjes transform of the Marčenko–Pastur law in (4.4); see (4.6). The proof is complete.

4.6.2 Proof of Theorem 4.3(2)

Assume $\liminf_{n\to\infty} n\,\mathbb{E}[Y^4] = \delta > 0$. For $k \ge 1$, the expected moments of the empirical spectral distribution $F_{\mathbf{R}}$ are

$$\widetilde{\beta}_k = \mathbb{E}\Big[\int x^k\, \mathrm{d}F_{\mathbf{R}}(x)\Big] = p^{-1}\mathbb{E}[\mathrm{tr}\, \mathbf{R}^k] = p^{-1} \sum_{i_1,\ldots,i_k=1}^{p} F(i_1,\ldots,i_k)\,. \tag{4.60}$$

From (4.50) we know that

$$p^{-1}\mathbb{E}[\mathrm{tr}(\mathbf{R})^k] \ge \sum_{r=1}^{k} (p-1)(p-2)\cdots(p-r+1)\Big(\sum_{I\in\mathcal{I}_{r,k}(0)} + \sum_{I\in\mathcal{I}_{r,k}(1)}\Big)F(I) =: S_3 + S_4\,.$$

By Proposition 4.11 and (4.32), we have

$$\lim_{n\to\infty} S_3 = \sum_{r=1}^{k} \frac{1}{r}\binom{k}{r-1}\binom{k-1}{r-1}\gamma^{r-1} = \beta_k(\gamma)\,, \tag{4.61}$$

which we recognize from (4.5) as the k-th moment of the Marčenko–Pastur law.

Next, observe that for $k \ge 4$ and $2 \le r \le k-2$, $\mathcal{I}_{r,k}(1)$ contains the element

$$I_r = (1,2,1,2,\underbrace{2,\ldots,2}_{k-r-2},3,\ldots,r)\,.$$

One checks that $R(I_r) = r-2$ and $S(I_r) = (1,2,1,2)$; consult the PSA and Definition 4.7 for the definitions of $R(\cdot)$ and $S(\cdot)$. Moreover, by symmetry of Y_{it} we have

$$\begin{aligned} F(1,2,1,2) &= \sum_{t_1,\ldots,t_4=1}^{n} \mathbb{E}[Y_{1t_1}Y_{1t_2}Y_{1t_3}Y_{1t_4}Y_{2t_1}Y_{2t_2}Y_{2t_3}Y_{2t_4}] \\ &= \sum_{t_1,\ldots,t_4=1}^{n} (\mathbb{E}[Y_{t_1}Y_{t_2}Y_{t_3}Y_{t_4}])^2 \\ &= \sum_{t_1=1}^{n} (\mathbb{E}[Y_{t_1}^4])^2 + 3 \sum_{t_1\neq t_2=1}^{n} (\mathbb{E}[Y_{t_1}^2 Y_{t_2}^2])^2 \ge \frac{1}{n}(n\mathbb{E}[Y^4])^2\,. \end{aligned}$$

By Lemma 4.8 we have

$$F(I_r) = n^{2-r}F(1,2,1,2) \ge n^{1-r}(n\mathbb{E}[Y^4])^2$$

and consequently

$$\begin{aligned} \liminf_{n\to\infty} S_4 &\ge \liminf_{n\to\infty} \sum_{r=2}^{k-2} (p-1)(p-2)\cdots(p-r+1)F(I_r) \\ &\ge \liminf_{n\to\infty} \sum_{r=2}^{k-2} (p-1)(p-2)\cdots(p-r+1)n^{1-r}(n\mathbb{E}[Y^4])^2 = \delta^2 \sum_{r=2}^{k-2} \gamma^{r-1}\,. \end{aligned}$$

This together with (4.61) proves $\liminf_{n\to\infty} \widetilde{\beta}_k > \beta_k(\gamma)$, as desired.

4.7 Appendix

In this section we provide some auxiliary tools for the proofs of the main results.

Lemma 4.16. *Let $k \in \mathbb{N}$ and $1 \le j \le k$. Then*

$$-\sum_{i=2j-1}^{2k} (-1)^i \binom{2k}{i}\binom{i-1}{2j-2} = 1 \tag{4.62}$$

Proof. For $1 \le j \le k$ we rewrite (4.62) as

$$-\sum_{i=2j-1}^{2k} (-1)^i \binom{2k}{i} \frac{(i-1)!}{(i+1-2j)!} = (2j-2)!\,. \tag{4.63}$$

We define the functions

$$u(x) = (x-1)^{2k} - 1\,, \quad v(x) = \sum_{i=1}^{2k} \binom{2k}{i} (-1)^i x^{i-1}\,, \quad \text{and} \quad w(x) = \frac{1}{x}\,.$$

Then $v(x) = u(x)w(x)$ and since

$$\frac{(i-1)!}{(i+1-2j)!} = (i-1)(i-2)\cdots(i-2j+2)\,,$$

equation (4.63) is equivalent to an equation for the $(2j-2)$-th derivative of v evaluated at 1,

$$v^{(2j-2)}(1) = -(2j-2)!\,.$$

By Leibniz's rule for differentiation, one gets

$$v^{(2j-2)}(x) = (uw)^{(2j-2)}(x) = \sum_{\ell=0}^{2j-2}\binom{2j-2}{\ell}u^{(\ell)}(x)w^{(2j-2-\ell)}(x)\,.$$

Observe that $u^{(0)}(1) = -1$ and $u^{(\ell)}(1) = 0$ for $1 \le \ell \le 2j-2$. Furthermore we have $w^{(2j-2-\ell)}(1) = (2j-2-\ell)!$. Hence, we conclude

$$v^{(2j-2)}(1) = \sum_{\ell=0}^{2j-2}\binom{2j-2}{\ell}u^{(\ell)}(1)w^{(2j-2-\ell)}(1) = -(2j-2)!\,,$$

completing the proof. □

For $k \in \mathbb{N}$ and $x \in [0,1]$, define the function

$$f_k(x) = 1 + \sum_{i=1}^{2k}\binom{2k}{i}\Big(\frac{-1}{1+x}\Big)^i \sum_{r=1}^{i}\frac{1}{r}\binom{i}{r-1}\binom{i-1}{r-1}x^{r-1}\,. \tag{4.64}$$

The following is our key lemma.

Lemma 4.17. *We have for $k \in \mathbb{N}$ and $x \in [0,1]$*

$$f_k(x) = 1 - \sum_{j=1}^{k}\frac{x^{j-1}}{(1+x)^{2j-1}}\frac{(2j-2)!}{j!(j-1)!} \le \frac{(4x)^k}{(1+x)^{2k-1}}\,.$$

Proof. From [6, page 41] we know that

$$\sum_{r=1}^{i}\frac{1}{r}\binom{i}{r-1}\binom{i-1}{r-1}x^{r-1} = \sum_{r=0}^{\lfloor (i-1)/2\rfloor} x^r(1+x)^{i-1-2r}\frac{(i-1)!}{(i-1-2r)!r!(r+1)!}\,.$$

Changing the order of summation one obtains

$$\begin{aligned} f_k(x) - 1 &= \sum_{r=0}^{k-1}\sum_{i=2r+1}^{2k}\binom{2k}{i}(-1)^i x^r(1+x)^{-1-2r}\frac{(i-1)!}{(i-1-2r)!r!(r+1)!} \\ &= \sum_{j=1}^{k}\frac{x^{j-1}}{(1+x)^{2j-1}}\frac{1}{j!(j-1)!}\sum_{i=2j-1}^{2k}\binom{2k}{i}(-1)^i\frac{(i-1)!}{(i+1-2j)!} \\ &= -\sum_{j=1}^{k}\frac{x^{j-1}}{(1+x)^{2j-1}}\frac{(2j-2)!}{j!(j-1)!}\,, \end{aligned}$$

where the last equality followed from Lemma 4.16 and its equivalent formulation (4.63).

For $j \in \mathbb{N}$ define

$$g_j(x) = \frac{(1+x)^{2j-1}}{x^j} f_j(x) \,. \tag{4.65}$$

We have $g_1(x) = 1$ and $g_2(x) = 2 + x$. A straightforward induction proves the recursion

$$g_j(x) = \frac{(1+x)^2 g_{j-1}(x) - g_{j-1}(0)}{x} \,, \quad j \geq 2 \,. \tag{4.66}$$

From this recursive construction one deduces that $g_j(x)$ is a polynomial of degree $j-1$ with positive coefficients.

Next we show $g_k(x) \leq 4^k$. Clearly we have $g_1(x) \leq 4$ and $g_2(x) \leq 4^2$. Therefore assume

$$\|g_{k-1}\|_{[0,1]} := \sup_{y \in [0,1]} |g_{k-1}(y)| \leq 4^{k-1} \,.$$

Then for $x \in [0,1]$,

$$\begin{aligned} g_k(x) &= \frac{x(2+x)g_{k-1}(x) + g_{k-1}(x) - g_{k-1}(0)}{x} \\ &\leq (2+x)g_{k-1}(x) + g_{k-1}(1) \leq (2+x+1)\|g_{k-1}\|_{[0,1]} \\ &\leq (3+x)4^{k-1} \leq 4^k \,. \end{aligned}$$

In view of (4.65), this finishes the proof. □

Bibliography

[1] ADAMCZAK, R., LITVAK, A. E., PAJOR, A., AND TOMCZAK-JAEGERMANN, N. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc. 23*, 2 (2010), 535–561.

[2] ADAMCZAK, R., LITVAK, A. E., PAJOR, A., AND TOMCZAK-JAEGERMANN, N. Sharp bounds on the rate of convergence of the empirical covariance matrix. *C. R. Math. Acad. Sci. Paris 349*, 3-4 (2011), 195–200.

[3] ANDERSON, T. W. Asymptotic theory for principal component analysis. *Ann. Math. Statist. 34* (1963), 122–148.

[4] AUFFINGER, A., BEN AROUS, G., AND PÉCHÉ, S. Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Ann. Inst. Henri Poincaré Probab. Stat. 45*, 3 (2009), 589–610.

[5] BAI, Z., FANG, Z., AND LIANG, Y.-C. *Spectral Theory of Large Dimensional Random Matrices and Its Applications to Wireless Communications and Finance Statistics: Random Matrix Theory and Its Applications.* World Scientific, 2014.

[6] BAI, Z., AND SILVERSTEIN, J. W. *Spectral Analysis of Large Dimensional Random Matrices*, second ed. Springer Series in Statistics. Springer, New York, 2010.

[7] BAI, Z., AND ZHOU, W. Large sample covariance matrices without independence structures in columns. *Statist. Sinica 18*, 2 (2008), 425–442.

[8] BAI, Z. D. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica 9*, 3 (1999), 611–677. With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author.

[9] BAI, Z. D., SILVERSTEIN, J. W., AND YIN, Y. Q. A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal. 26*, 2 (1988), 166–168.

[10] BAI, Z. D., AND YIN, Y. Q. Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab. 21*, 3 (1993), 1275–1294.

[11] BANNA, M. Limiting spectral distribution of Gram matrices associated with functionals of β-mixing processes. *J. Math. Anal. Appl. 433*, 1 (2016), 416–433.

[12] BANNA, M., AND MERLEVÈDE, F. Limiting spectral distribution of large sample covariance matrices associated with a class of stationary processes. *J. Theoret. Probab. 28*, 2 (2015), 745–783.

[13] BANNA, M., MERLEVÈDE, F., AND PELIGRAD, M. On the limiting spectral distribution for a large class of symmetric random matrices with correlated entries. *Stochastic Process. Appl. 125*, 7 (2015), 2700–2726.

[14] BAO, Z., PAN, G., AND ZHOU, W. Tracy-Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab. 17* (2012), no. 88, 32.

[15] Belinschi, S., Dembo, A., and Guionnet, A. Spectral measure of heavy tailed band and covariance random matrices. *Comm. Math. Phys. 289*, 3 (2009), 1023–1055.

[16] Belitskiĭ, G. R., and Lyubich, Y. I. *Matrix Norms and their Applications*, vol. 36 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, Basel, 1988.

[17] Ben Arous, G., and Guionnet, A. The spectrum of heavy tailed random matrices. *Comm. Math. Phys. 278*, 3 (2008), 715–751.

[18] Benaych-Georges, F., and Péché, S. Localization and delocalization for heavy tailed band matrices. *Ann. Inst. Henri Poincaré Probab. Stat. 50*, 4 (2014), 1385–1403.

[19] Bhatia, R. *Matrix Analysis*, vol. 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[20] Bingham, N. H., Goldie, C. M., and Teugels, J. L. *Regular Variation*, vol. 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1987.

[21] Chakrabarty, A., Hazra, R. S., and Roy, P. Maximum eigenvalue of symmetric random matrices with dependent heavy tailed entries. *Available at* `http://arxiv.org/abs/1309.1407` (2013).

[22] Cline, D. B. H., and Hsing, T. Large deviation probabilities for sums of random variables with heavy or subexponential tails. *Technical report. Statistics Dept., Texas A&M University.* (1998).

[23] Davis, R. A., Heiny, J., Mikosch, T., and Xie, X. Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series. *Extremes 19*, 3 (2016), 517–547.

[24] Davis, R. A., Mikosch, T., and Pfaffel, O. Asymptotic theory for the sample covariance matrix of a heavy-tailed multivariate time series. *Stochastic Process. Appl. 126*, 3 (2016), 767–799.

[25] Davis, R. A., Pfaffel, O., and Stelzer, R. Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails. *Stochastic Process. Appl. 124*, 1 (2014), 18–50.

[26] de Haan, L., and Ferreira, A. *Extreme Value Theory: An Introduction.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.

[27] Denisov, D., Dieker, A. B., and Shneer, V. Large deviations for random walks under subexponentiality: the big-jump domain. *Ann. Probab. 36*, 5 (2008), 1946–1991.

[28] Dobriban, E. Efficient computation of limit spectra of sample covariance matrices. *Random Matrices Theory Appl. 4*, 4 (2015), 1550019, 36.

[29] Donoho, D. High-dimensional data analysis: the curses and blessings of dimensionality. *Technical Report, Stanford University* (2000).

[30] El Karoui, N. On the largest eigenvalue of Wishart matrices with identity covariance when n,p and p/n tend to infinity. *Available at* `http://arxiv.org/abs/math/0309355` (2003).

[31] El Karoui, N. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist. 36*, 6 (2008), 2757–2790.

[32] El Karoui, N. Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab. 19*, 6 (2009), 2362–2405.

[33] El Karoui, N., and Purdom, E. The bootstrap, covariance matrices and pca in moderate and high-dimensions. *Technical Report, University of California* (2016).

[34] Embrechts, P., and Goldie, C. M. On closure and factorization properties of subexponential and related distributions. *J. Austral. Math. Soc. Ser. A 29*, 2 (1980), 243–256.

[35] Embrechts, P., Klüppelberg, C., and Mikosch, T. *Modelling Extremal Events for Insurance and Finance*, vol. 33 of *Applications of Mathematics (New York).* Springer, Berlin, 1997.

[36] Embrechts, P., and Veraverbeke, N. Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance Math. Econom. 1*, 1 (1982), 55–72.

[37] Feller, W. *An Introduction to Probability Theory and its Applications. Vol. II.* John Wiley & Sons, Inc., New York-London-Sydney, 1966.

[38] Geman, S. A limit theorem for the norm of random matrices. *Ann. Probab. 8*, 2 (1980), 252–261.

[39] Giné, E., Götze, F., and Mason, D. M. When is the Student *t*-statistic asymptotically standard normal? *Ann. Probab. 25*, 3 (1997), 1514–1531.

[40] Heinrich, C., and Podolskij, M. On spectral distribution of high dimensional covariation matrices. Workingpaper, Institut for Økonomi, Aarhus Universitet, 12 2014.

[41] Heiny, J., and Mikosch, T. Eigenvalues and eigenvectors of heavy-tailed sample covariance matrices with general growth rates: the iid case. *Stochastic Process. Appl.* (2016), 29.

[42] Heiny, J., Mikosch, T., and Davis, R. A. Limit theory for the singular values of the sample autocovariance matrix function of multivariate time series. *Work in progress* (2016).

[43] Horn, R. A., and Johnson, C. R. *Matrix Analysis*, second ed. Cambridge University Press, Cambridge, 2013.

[44] Hult, H., and Samorodnitsky, G. Tail probabilities for infinite series of regularly varying random vectors. *Bernoulli 14*, 3 (2008), 838–864.

[45] Janssen, A., Mikosch, T., Mohsen, R., and Xiaolei, X. The eigenvalues of the sample covariance matrix of a multivariate heavy-tailed stochastic volatility model. *To appear in Bernoulli.*

[46] JIANG, T. The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā 66*, 1 (2004), 35–48.

[47] JOHANSSON, K. Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices. *Comm. Math. Phys. 215*, 3 (2001), 683–705.

[48] JOHNSTONE, I. M. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist. 29*, 2 (2001), 295–327.

[49] LAM, C., AND YAO, Q. Factor modeling for high-dimensional time series: inference for the number of factors. *Ann. Statist. 40*, 2 (2012), 694–726.

[50] MA, Z. Accuracy of the Tracy-Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli 18*, 1 (2012), 322–359.

[51] MARČENKO, V. A., AND PASTUR, L. A. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.) 72 (114)* (1967), 507–536.

[52] MASON, D. M., AND ZINN, J. Acknowledgment of priority: "When does a randomly weighted self-normalized sum converge in distribution?" [Electron. Comm. Probab. **10** (2005), 70–81 (electronic); mr2133894]. *Electron. Comm. Probab. 10* (2005), 297 (electronic).

[53] MIKOSCH, T., AND WINTENBERGER, O. A large deviations approach to limit theory for heavy-tailed time series. *Probab. Theory Related Fields 166*, 1-2 (2016), 233–269.

[54] MUIRHEAD, R. J. *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, Inc., New York, 1982. Wiley Series in Probability and Mathematical Statistics.

[55] NAGAEV, S. V. Large deviations of sums of independent random variables. *Ann. Probab. 7*, 5 (1979), 745–789.

[56] O'ROURKE, S., VU, V., AND WANG, K. Eigenvectors of random matrices: a survey. *J. Combin. Theory Ser. A 144* (2016), 361–442.

[57] PAUL, D., AND AUE, A. Random matrix theory in statistics: a review. *J. Statist. Plann. Inference 150* (2014), 1–29.

[58] PÉCHÉ, S. Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields 143*, 3-4 (2009), 481–516.

[59] PFAFFEL, O., AND SCHLEMM, E. Eigenvalue distribution of large sample covariance matrices of linear processes. *Probab. Math. Statist. 31*, 2 (2011), 313–329.

[60] RESNICK, S. I. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007.

[61] RESNICK, S. I. *Extreme Values, Regular Variation and Point Processes.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2008. Reprint of the 1987 original.

[62] RUDELSON, M., AND VERSHYNIN, R. Delocalization of eigenvectors of random matrices with independent entries. *Duke Math. J. 164*, 13 (2015), 2507–2538.

[63] SHORES, T. S. *Applied Linear Algebra and Matrix Analysis.* Undergraduate Texts in Mathematics. Springer, New York, 2007.

[64] SILVERSTEIN, J. W. Some limit theorems on the eigenvectors of large-dimensional sample covariance matrices. *J. Multivariate Anal. 15*, 3 (1984), 295–324.

[65] SOSHNIKOV, A. Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails. *Electron. Comm. Probab. 9* (2004), 82–91 (electronic).

[66] SOSHNIKOV, A. Poisson statistics for the largest eigenvalues in random matrix ensembles. In *Mathematical physics of quantum mechanics*, vol. 690 of *Lecture Notes in Phys.* Springer, Berlin, 2006, pp. 351–364.

[67] SRIVASTAVA, N., AND VERSHYNIN, R. Covariance estimation for distributions with $2+\varepsilon$ moments. *Ann. Probab. 41*, 5 (2013), 3081–3111.

[68] TAO, T., AND VU, V. Random matrices: universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys. 298*, 2 (2010), 549–572.

[69] TAO, T., AND VU, V. Random covariance matrices: universality of local statistics of eigenvalues. *Ann. Probab. 40*, 3 (2012), 1285–1315.

[70] TIKHOMIROV, K. The limit of the smallest singular value of random matrices with i.i.d. entries. *Adv. Math. 284* (2015), 1–20.

[71] TRACY, C. A., AND WIDOM, H. Distribution functions for largest eigenvalues and their applications. In *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)* (2002), Higher Ed. Press, Beijing, pp. 587–596.

[72] WIGNER, E. P. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2) 62* (1955), 548–564.

[73] WIGNER, E. P. Characteristic vectors of bordered matrices with infinite dimensions. II. *Ann. of Math. (2) 65* (1957), 203–207.

[74] WISHART, J. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* (1928), 32–52.

[75] XIA, N., AND ZHENG, X. On the inference about the spectra of high-dimensional covariance matrix based on noisy observations-with applications to integrated covolatility matrix inference in the presence of microstructure noise. *arXiv:1409.2121* (2014).

[76] XIA, N., AND ZHENG, X. On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations. *Available at SSRN 2764654* (2016).

[77] XIAO, H., AND ZHOU, W. Almost sure limit of the smallest eigenvalue of some sample correlation matrices. *J. Theoret. Probab. 23*, 1 (2010), 1–20.

[78] YAO, J., ZHENG, S., AND BAI, Z. *Large sample covariance matrices and high-dimensional data analysis.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York, 2015.

[79] YIN, Y. Q., BAI, Z. D., AND KRISHNAIAH, P. R. On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields 78*, 4 (1988), 509–521.