Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 128 (2018) 2779–2815

stochastic processes and their applications

www.elsevier.com/locate/spa

Almost sure convergence of the largest and smallest eigenvalues of high-dimensional sample correlation matrices[☆]

Johannes Heiny[a,*], Thomas Mikosch[b]

[a] Department of Mathematics, Aarhus University, Ny Munkegade 118, DK-8000 Aarhus C, Denmark

[b] Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark

Received 20 December 2016; received in revised form 19 September 2017; accepted 6 October 2017
Available online 18 October 2017

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 Bai and Zhou (2008). 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.

☆ Johannes Heiny's and Thomas Mikosch's research is partly supported by the Danish Research Council Grant DFF-4002-00435 "Large random matrices with heavy tails and dependence". Johannes Heiny thanks the Villum Foundation for the grant 11745 "Ambit fields: Probabilistic properties and statistical inference".

* Corresponding author.
E-mail addresses: johannes.heiny@math.ku.dk (J. Heiny), mikosch@math.ku.dk (T. Mikosch).
URL: http://www.math.ku.dk/∼mikosch (T. Mikosch).

https://doi.org/10.1016/j.spa.2017.10.002
0304-4149/

MSC: primary 60B20; secondary 60F05; 60F10; 60G10; 60G55; 60G70

Keywords: Sample correlation matrix; Infinite fourth moment; Largest eigenvalue; Smallest eigenvalue; Spectral distribution; Sample covariance matrix; Self-normalization; Regular variation; Combinatorics

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

$$\lambda_{(1)} \ge \cdots \ge \lambda_{(p)}, \tag{1.1}$$

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

$$\lim_{n\to\infty} \frac{p_n}{n} = \gamma \in (0, 1]. \tag{G_γ}$$

For the finite p case, we refer to [3,32,26]. 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 [30] 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 [21]. Typical applications where (G_γ) seems reasonable are discussed in [28,19].

In comparison with $(\lambda_{(i)})$, much less is known about the ordered eigenvalues

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

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

$$R_{ij} = \sum_{t=1}^{n} \frac{X_{it}X_{jt}}{\sqrt{D_i}\sqrt{D_j}} = \sum_{t=1}^{n} Y_{it}Y_{jt}, \quad i, j = 1, \dots, p. \tag{1.2}$$

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

$$D_i = D_i^{(n)} = \sum_{t=1}^{n} X_{it}^2, \qquad i = 1, \dots, p; \; n \geq 1. \tag{1.3}$$

Note that the dependence of $(\lambda_{(i)})$ and $(\mu_{(i)})$ on n is suppressed in the notation.

1.1. The case (X_{it}) iid, $\mathbb{E}[X^4] < \infty$ and $\mathbb{E}[X^2] = 1$

In this case the behaviors 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 $\lambda_1(\mathbf{A}), \dots, \lambda_p(\mathbf{A})$ the *empirical spectral distribution* is defined by

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

Many functionals of the eigenvalues $\lambda_1(\mathbf{A}), \dots, \lambda_p(\mathbf{A})$ 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 [6,39] 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 $F_{\mathbf{A}}$:

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

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 $z \in \mathbb{C}^+$.

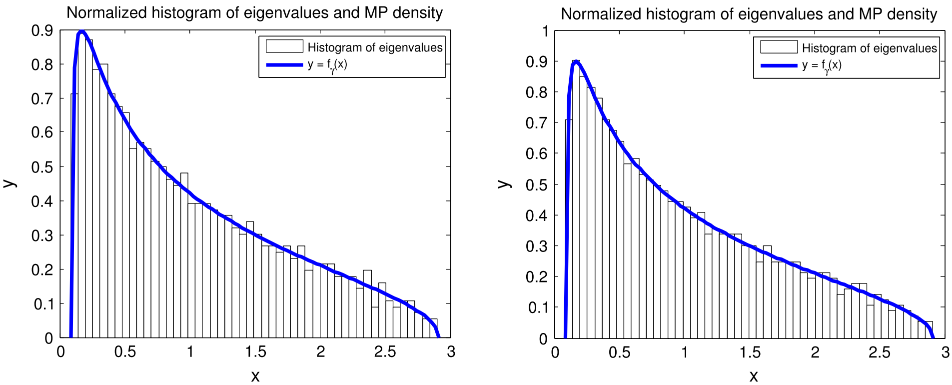
Under the growth condition (G_γ), the sequence of empirical spectral distributions of the normalized sample covariance matrix $n^{-1}\mathbf{XX}'$ converges weakly to the Marčenko–Pastur law with density

$$f_\gamma(x) = \begin{cases} \dfrac{1}{2\pi x \gamma} \sqrt{(b - x)(x - a)}, & \text{if } a \leq x \leq b, \\ 0, & \text{otherwise}, \end{cases} \tag{1.4}$$

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 [30]. 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 Fig. 1.

The Marčenko–Pastur law has kth moment

$$\beta_k = \beta_k(\gamma) = \int_a^b x^k f_\gamma(x) \, \mathrm{d}x = \sum_{r=1}^{k} \frac{1}{r} \binom{k}{r-1} \binom{k-1}{r-1} \gamma^{r-1}, \quad k \geq 1, \tag{1.5}$$



(a) Sample correlation: $\mu_{(p)} = 0.086$, $\mu_{(1)} = 2.898$.

(b) Sample covariance: $\lambda_{(p)}/(n\mathbb{E}[X^2]) = 0.085$, $\lambda_{(1)}/(n\mathbb{E}[X^2]) = 2.908$.

Fig. 1. Histogram and Marčenko–Pastur density for $X \sim t_6$, $n = 2000$, $p = 1000$. $\gamma = 0.5$, $(1-\sqrt{\gamma})^2 = 0.085$, $(1+\sqrt{\gamma})^2 = 2.914$.

and Stieltjes transform

$$s(z) = \int_{\mathbb{R}} \frac{1}{x-z} f_\gamma(x)\,\mathrm{d}x = \frac{1-\gamma-z+\sqrt{(1+\gamma-z)^2-4\gamma}}{2\gamma z}; \tag{1.6}$$

see [6, Chapter 3] or [5,39].

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

$$\limsup_{n\to\infty} n^{-1}\lambda_{(p)} \le (1-\sqrt{\gamma})^2 \le (1+\sqrt{\gamma})^2 \le \liminf_{n\to\infty} n^{-1}\lambda_{(1)} \quad \text{a.s.} \tag{1.7}$$

The finiteness of the fourth moment of X is necessary for the almost sure convergence of $\lambda_{(1)}/n$; see [7]. If $\mathbb{E}[X^4] < \infty$, one has (see [6])

$$n^{-1}\lambda_{(1)} \to (1+\sqrt{\gamma})^2 \quad \text{and} \quad n^{-1}\lambda_{(p)} \to (1-\sqrt{\gamma})^2 \quad \text{a.s.} \tag{1.8}$$

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 [36] 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 [28] complemented (1.8) by the corresponding central limit theorem in the special case of iid standard normal entries:

$$n^{2/3}\,\frac{(\sqrt{\gamma})^{1/3}}{\big(1+\sqrt{\gamma}\big)^{4/3}}\Big(\frac{\lambda_{(1)}}{n} - \Big(1+\sqrt{\tfrac{p}{n}}\Big)^2\Big) \xrightarrow{\mathrm{d}} \xi, \tag{1.9}$$

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

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

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 [37] 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 [21]". 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 [14]) yields

$$\begin{aligned} \max_{i=1,\ldots,p} |\mu_{(i)} - n^{-1}\lambda_{(i)}| &\le \|\mathbf{X}\mathbf{X}'\mathbf{F} - n^{-1}\mathbf{X}\mathbf{X}'\|_2 \\ &\le n^{-1}\|\mathbf{X}\mathbf{X}'\|_2 \|n\mathbf{F} - \mathbf{I}\|_2 \\ &= n^{-1}\lambda_{(1)} \max_{i=1,\ldots,p} \Big|\frac{n}{D_i} - 1\Big|, \end{aligned} \tag{1.10}$$

where for any matrix $\mathbf{A}$, $\|\mathbf{A}\|_2$ denotes its spectral norm, i.e., its largest singular value.

Lemma 2 in [8] implies that $\mathbb{E}[X^4] < \infty$ is equivalent to

$$\max_{i=1,\ldots,p} \Big|\frac{n}{D_i} - 1\Big| \stackrel{\text{a.s.}}{\to} 0,$$

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 [27,38] to derive

$$\mu_{(1)} \to (1+\sqrt{\gamma})^2 \quad \text{and} \quad \mu_{(p)} \to (1-\sqrt{\gamma})^2 \quad \text{a.s.} \tag{1.11}$$

If the assumption $\mathbb{E}[X^4] < \infty$ is weakened to $\lim_{n\to\infty} n\,\mathbb{P}(X^4 > n) = 0$, the paper [8] proves that $n^{-1}\lambda_{(1)} \stackrel{\mathbb{P}}{\to} (1+\sqrt{\gamma})^2$ and $\max_{i=1,\ldots,p}\big|n/D_i - 1\big| \stackrel{\mathbb{P}}{\to} 0$. 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 [13] 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 (1.9) with $\lambda_{(1)}/n$ replaced by $\mu_{(1)}$. A similar limit result holds for $\mu_{(p)}$.

1.2. The case (X_{it}) iid and $\mathbb{E}[X^4] = \infty$

Asymptotic theory for the eigenvalues of $\mathbf{X}\mathbf{X}'$ in the case of an entry distribution with infinite fourth moment was studied in [33,34,4] in the cases when $p/n \to \gamma \in (0,\infty)$, while the authors of [15,25] 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

$$\mathbb{P}(|X| > x) = x^{-\alpha} L(x), \tag{1.12}$$

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 $a_{np}^{-2}\lambda_{(p)} \stackrel{\mathbb{P}}{\to} 0$. 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 (1.8) that

$$\frac{\lambda_{(p)}}{\lambda_{(1)}} \xrightarrow{\text{a.s.}} \frac{(1-\sqrt{\gamma})^2}{(1+\sqrt{\gamma})^2} \quad \text{if } \alpha > 4,$$

$$\frac{a_{np}^2}{n}\frac{\lambda_{(p)}}{\lambda_{(1)}} \xrightarrow{d} \frac{(1-\sqrt{\gamma})^2}{\eta_{\alpha/2}} \quad \text{if } \alpha \in (2,4),$$

$$\frac{a_{np}^2}{n}\frac{\lambda_{(p)}}{\lambda_{(1)}} \xrightarrow{\text{a.s.}} 0 \quad \text{if } \alpha \in (0,2),$$

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 [9] assert that $(F_{n^{-1}\mathbf{XX}'})$ and $(F_\mathbf{R})$ converge weakly to the Marčenko–Pastur law. From [7] 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 (1.10) 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 [9] 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 (1.7), the weak limit of $(F_\mathbf{R})$ provides a first idea what the limits of the extreme eigenvalues might be.

1.3. (X_{it}) identically distributed, but dependent

For practical purposes it is important to work with arbitrary population covariance matrices and not just $n^{-1}\mathbb{E}[\mathbf{XX}'] = \mathbf{I}$. 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 [6,5,39]. Over the last couple of years significant progress on limiting spectral distributions for dependent time series was achieved; see for example [12,11,10]. 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 [22] the authors find that the bootstrap works for the top eigenvalues if they are sufficiently separated from the bulk. Among others, El Karoui [20] proposed to use the Marčenko–Pastur equation, which basically requires more insight into the empirical spectral distribution and its support. This was achieved in [18], where an algorithm for calculating the spectral distribution based on certain approximate integral equations for its Stieltjes transform was presented.

In view of [17,16,15] 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 (1.10) can be developed to show that methods for sample covariance matrices can be applied to sample correlation matrices; see for example [21]. Theorem 1 in [21] 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 [35,1,2,39].

1.4. About this paper

In Section 2 we introduce the basic assumptions of this paper and discuss their meaning. The main results are given in Section 3. We show that the limiting spectral distribution of the sample correlation matrices is the Marčenko–Pastur law (Theorem 3.1) and that the extreme eigenvalues converge a.s. to the endpoints of the limiting support (Theorem 3.3) 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 3 with a small simulation study which shows that the asymptotic results work nicely.

We continue with some technical results in Section 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 3.3 and 3.1 are given in Sections 5 and 6, respectively. Both proofs heavily depend on the techniques developed in Section 4. We conclude with an Appendix which contains some auxiliary analytical results.

Condition (C_q) is crucial for the proof of Theorem 3.1. In Section 2 we discuss this condition and find out that it is very close to condition (2.2) 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 3.1 may be proved only under (2.2).

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. In order to exclude the degenerate case we assume $\operatorname{var}(X) > 0$. Recall the notation

$$Y_{it} = \frac{X_{it}}{\sqrt{D_i}}, \quad i = 1, \ldots, p;\ t = 1, \ldots, n. \tag{2.1}$$

For ease of notation we will sometimes write $(Y_1, \ldots, Y_n) = (Y_{11}, \ldots, Y_{1n})$, $Y = Y_1$ and $D = D_1$.

2.1. Domain of attraction type-condition for the Gaussian law

One of the basic assumptions in this paper is

$$\mathbb{E}\big[Y_1 Y_2\big] = o(n^{-2}) \quad \text{and} \quad \mathbb{E}\big[Y_1^4\big] = o(n^{-1}), \qquad n \to \infty. \tag{2.2}$$

In [24] it was proved that condition (2.2) 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 (2.2) is valid as a domination argument shows.

Jonsson [29] proved that $\mathbb{E}[(Y_1 + \cdots + Y_n)^2] \ge 1$, with equality if and only if X is symmetric. He also gave an explicit expression for the mixed moment

$$\mathbb{E}[Y_1Y_2] = \int_0^\infty (\mathbb{E}[X\mathrm{e}^{-\lambda X^2}])^2(\mathbb{E}[\mathrm{e}^{-\lambda X^2}])^{n-2}\,\mathrm{d}\lambda, \quad n \ge 2.$$

In view of the identity $\mathbb{E}[(Y_1 + \cdots + Y_n)^2] = 1 + n(n-1)\,\mathbb{E}[Y_1Y_2]$, one makes the interesting observation that $\mathbb{E}[Y_1Y_2]$ is always nonnegative.

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 $(k^3q)/n \to 0$, and the moment inequality

$$\mathbb{E}[Y_1^{2m_1}\cdots Y_r^{2m_r}] \le \frac{q_n}{n}\,\mathbb{E}[Y_1^{2m_1}\cdots Y_{r-1}^{2m_{(r-1)}}Y_r^{(2m_r-2)}] \qquad (C_q)$$

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 (2.2). 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. Condition (C_q) implies that

$$\lim_{n\to\infty} (\log n)^3\, n\,\mathbb{E}\big[Y_1^4\big] = 0,$$

which is satisfied for regularly varying distributions with index $\alpha > 2$.

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 [31] 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

$$\lim_{n\to\infty} n\,\mathbb{E}[Y_1^4] = 1 - \frac{\alpha}{2}, \qquad (2.3)$$

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 [24]:

$$\mathbb{E}[Y_1^{2m_1}\cdots Y_r^{2m_r}] = \frac{1}{(k-1)!}\int_0^\infty \lambda^{k-1}(\mathbb{E}[\mathrm{e}^{-\lambda X^2}])^{n-r}\prod_{j=1}^r \mathbb{E}[X^{2m_j}\mathrm{e}^{-\lambda X^2}]\,\mathrm{d}\lambda, \tag{2.4}$$

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 2.1 (*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 distribution via (2.4). 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

$$\mathbb{E}[X^{2m}\mathrm{e}^{-\lambda X^2}] = (2m-1)!!(1+2\lambda)^{-1/2-m}. \tag{2.5}$$

By (2.4) and (2.5), 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

$$\mathbb{E}[Y_1^{2m_1}\cdots Y_r^{2m_r}] = \frac{\Gamma(n/2)}{2^\ell\Gamma(n/2+\ell)}\prod_{j=1}^r(2m_j-1)!!, \tag{2.6}$$

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 2.2 (*Gamma Distribution*). Assume $X^2 \sim \mathrm{Gamma}(\alpha,\beta)$, $\alpha,\beta > 0$. In this case

$$\frac{\mathrm{d}^m}{\mathrm{d}\lambda^m}\mathbb{E}[\mathrm{e}^{-\lambda X^2}] = \frac{\mathrm{d}^m}{\mathrm{d}\lambda^m}\Big(1+\frac{\lambda}{\beta}\Big)^{-\alpha} = \frac{\Gamma(1-\alpha)}{\Gamma(1-\alpha-m)}\beta^{-m}\Big(1+\frac{\lambda}{\beta}\Big)^{-\alpha-m}.$$

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 $q_n = (k_n+\alpha)/\alpha$.

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 3.1 (*Limiting Spectral Distribution*). *Assume the condition* (G_γ).

(1) *If X is centered and (2.2) holds then $F_{\mathbf{R}}$ converges weakly to the Marčenko–Pastur law given in (1.4).*
(2) *If X is symmetric and (2.2) does not hold, i.e.,* $\liminf_{n\to\infty} n\,\mathbb{E}[Y^4] > 0$, *then*

$$\liminf_{n\to\infty} \mathbb{E}\Big[\int x^k F_{\mathbf{R}}(dx)\Big] > \beta_k(\gamma), \qquad k \geq 1,$$

where $\beta_k(\gamma)$ is the kth moment of the Marčenko–Pastur law given in (1.5).

The proof of parts (1) and (2) will be given in Sections 6.1 and 6.2, respectively.

Remark 3.2. Part (1) with condition (2.2) replaced by $\mathbb{E}[X^2] < \infty$ was proved in [27]. Later, in [9] 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 (2.2) holds under the latter condition. Part (2) shows that (2.2) is the minimal condition for part (1) if X is symmetric. By Lemma B.1 in [6], $\lim_{n\to\infty}\mathbb{E}\big[\int x^k F_{\mathbf{R}}(dx)\big] = \beta_k(\gamma), k \geq 1$, implies weak convergence of $F_{\mathbf{R}}$ to the Marčenko–Pastur distribution as the latter is uniquely determined by its moments $(\beta_k(\gamma))_{k\geq 1}$.

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

$$\mathbb{E}[\mu_{(1)}^{k_n}] \leq \mathbb{E}[\operatorname{tr}\mathbf{R}^{k_n}] = \sum_{i_1,\ldots,i_{k_n}=1}^{p}\ \sum_{t_1,\ldots,t_{k_n}=1}^{n} \mathbb{E}[Y_{i_1 t_{k_n}} Y_{i_1 t_1} Y_{i_2 t_1} Y_{i_2 t_2} \cdots Y_{i_{k_n} t_{k_n-1}} Y_{i_{k_n} t_{k_n}}]$$

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 5.

Theorem 3.3 (*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

$$\mu_{(1)} \to (1+\sqrt{\gamma})^2 \quad \text{a.s.}, \tag{3.1}$$

$$\mu_{(p)} \to (1-\sqrt{\gamma})^2 \quad \text{a.s.} \tag{3.2}$$

Remark 3.4. Part (1) was proved in [27,38]; see the discussion in Section 1. Theorem 3.3 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; see also the discussion in Section 2.2. 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.

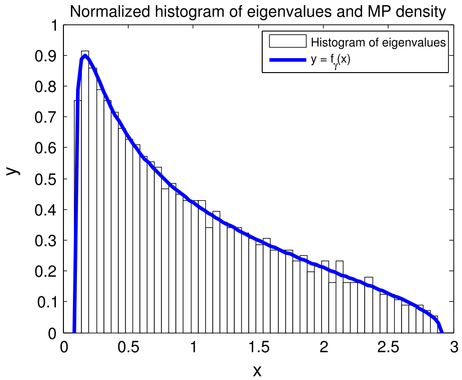
3.1. Simulation study

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 3.1. 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.

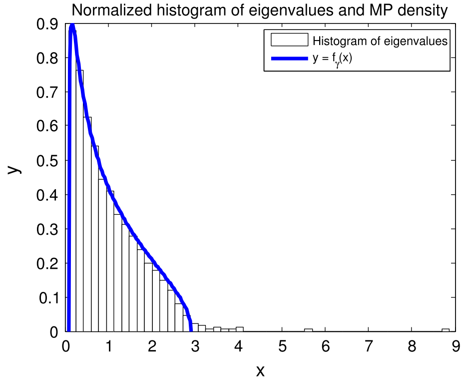
In Fig. 1 we simulated a 1000×2000 data matrix $\mathbf{X}$ with iid entries drawn from a t_6-distribution which we renormalized to meet the requirement $\mathbb{E}[X^2] = 1$. To illustrate the weak convergence of $(F_{\mathbf{R}})$ and $(F_{\mathbf{X}\mathbf{X}'/n})$ we plot the histograms of the eigenvalues $(\mu_{(i)})$ 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 Fig. 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 3.3 provides a formal justification.

In Figs. 3 and 4 we simulated from distributions with infinite fourth moment. We drew from a symmetrized Pareto distribution with parameter 3.99 to create the plots in Fig. 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 3.3, 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

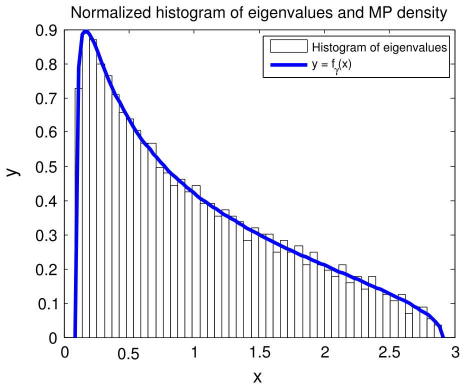


(a) Sample correlation: $\mu_{(p)} = 0.088$, $\mu_{(1)} = 2.880$.

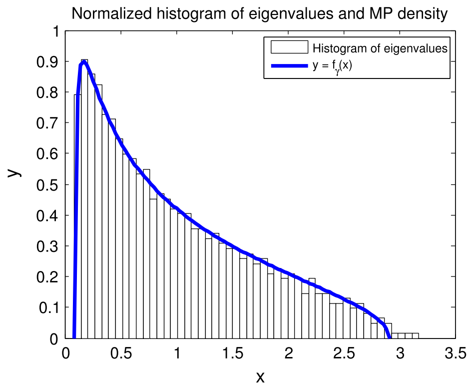


(b) Sample covariance: $\lambda_{(p)}/(n\mathbb{E}[X^2]) = 0.083$, $\lambda_{(1)}/(n\mathbb{E}[X^2]) = 8.870$.

Fig. 2. Histogram and Marčenko–Pastur density for $X \sim t_3$, $n = 2000$, $p = 1000$. Here $\gamma = p/n = 0.5$, $(1-\sqrt{\gamma})^2 = 0.085$, $(1+\sqrt{\gamma})^2 = 2.914$.



(a) Sample correlation: $\mu_{(p)} = 0.086$, $\mu_{(1)} = 2.902$.

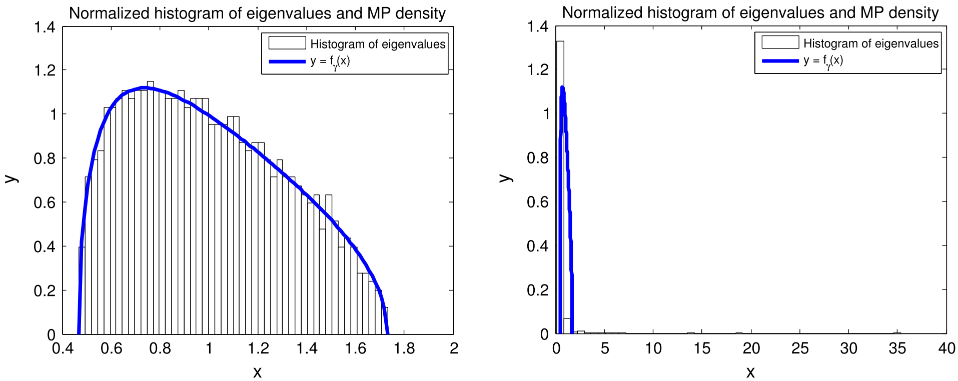


(b) Sample covariance: $\lambda_{(p)}/(n\mathbb{E}[X^2]) = 0.083$, $\lambda_{(1)}/(n\mathbb{E}[X^2]) = 3.176$.

Fig. 3. Histogram and Marčenko–Pastur density: $X \stackrel{d}{=} Z_1 - Z_2$ for $Z_i \sim$ Pareto(3.99), $n = 2000$, $p = 1000$. Here $\gamma = p/n = 0.5$, $(1-\sqrt{\gamma})^2 = 0.085$, $(1+\sqrt{\gamma})^2 = 2.914$.

Fig. 1(b) with a misspecified variance of the data could have resembled Fig. 3(b). In this respect sample correlations are more robust.

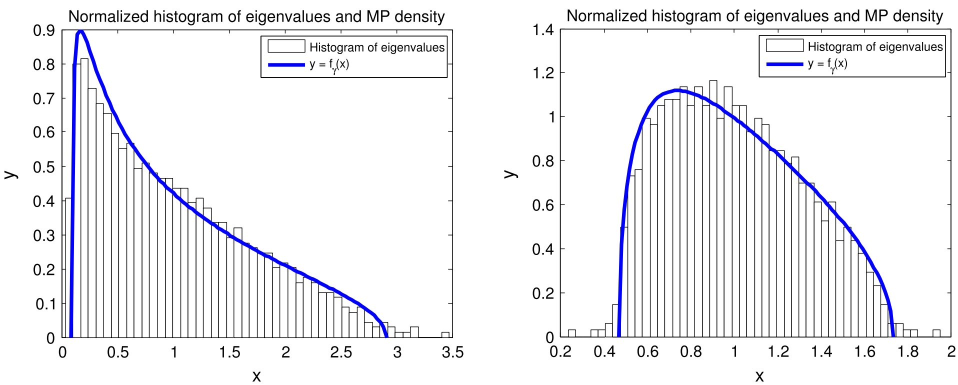
In Fig. 4, we choose X from the standardized $t_{2.1}$-distribution, moving closer to the infinite variance case. The histogram of $(\mu_{(i)})$ fits the Marčenko–Pastur density very well and the extreme eigenvalues are located in a close proximity of the endpoints of the Marčenko–Pastur support. The 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 [4,25,15], 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



(a) Sample correlation: $\mu_{(p)} = 0.469$, $\mu_{(1)} = 1.731$.

(b) Sample covariance: $\lambda_{(p)}/(n\mathbb{E}[X^2]) = 0.159$, $\lambda_{(1)}/(n\mathbb{E}[X^2]) = 35.319$.

Fig. 4. Histogram and Marčenko–Pastur density for $X \sim t_{2.1}$, $n = 10000$, $p = 1000$. Here $\gamma = p/n = 0.1$, $(1-\sqrt{\gamma})^2 = 0.467$, $(1+\sqrt{\gamma})^2 = 1.732$.



(a) $X \stackrel{d}{=} Z^2 - \mathbb{E}Z^2$ for $Z \sim t_{1.5}$, $n = 2000$, $p = 1000$, $\gamma = 0.5$.

(b) $X \sim t_{1.8}$, $n = 10000$, $p = 1000$, $\gamma = 0.1$.

Fig. 5. Histogram of $(\mu_{(i)})$ and Marčenko–Pastur density.

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 turn to the case where the assumptions of Theorem 3.1 are violated. We present two histograms of $(\mu_{(i)})$ with $\mathbb{E}[X^2] = \infty$ in Fig. 5. In (a), we choose the non-symmetric $X \stackrel{d}{=} Z^2 - \mathbb{E}Z^2$ 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 3.1(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.

3.2. A version of condition (C_q) for finite fourth moment and beyond

In applications it is quite challenging to check condition (C_q) for a given distribution. In principle, (C_q) can be verified by direct calculation using (2.4). However, for some classical distributions, such as the t-and symmetrized Pareto distributions, (2.4) requires the calculation of a high-dimensional integral which might not always be as easy to compute as in Examples 2.1 and 2.2. Due to the complexity of the calculations, this paper does not contain a nicely worked out example of a distribution with infinite fourth moment that satisfies (C_q).

In the existing literature the almost sure convergence of the extreme eigenvalues was established for finite fourth moment only. Section 3.1 indicates that the result holds for infinite fourth moment too. A necessary and sufficient condition, however, is unknown. Our condition (C_q) is a step towards the optimal condition which we believe to be $n\mathbb{E}[Y^4]\to 0$ for symmetric distributions. From our constructive method of proof one can see how $\mathbb{E}[Y^4]$ comes into play; see also Theorem 3.1 and especially the proof of part (2) for a better understanding how the behavior of $\mathbb{E}[Y^4]$ influences the limiting spectral distribution of $\mathbf{R}$.

In this subsection we show that a version of condition (C_q) holds for essentially all distributions with finite fourth moment. We also provide an explicit example of a distribution with infinite fourth moment such that $\mu_{(1)}\stackrel{\mathbb{P}}{\to}(1+\sqrt{\gamma})^2$ and $\mu_{(p)}\stackrel{\mathbb{P}}{\to}(1-\sqrt{\gamma})^2$.

For a slowly varying sequence $\delta_n\to 0$ define $\widehat{X}_{it}^{(n)}=X_{it}\mathbf{1}_{\{|X_{it}|\le\delta_n\sqrt{n}\}}$ and set

$$\widehat{\mathbf{X}}_n=\big(\widehat{X}_{it}^{(n)}\big)_{i=1,\dots,p;t=1,\dots,n}. \tag{3.3}$$

The sample correlation matrix of the truncated data $\widehat{\mathbf{R}}=\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}'$ with $\widehat{Y_{it}}=\widehat{X}_{it}^{(n)}/\sqrt{\sum_{t=1}^n(\widehat{X}_{jt}^{(n)})^2}$ and

$$\widehat{\mathbf{R}}=\big(\widehat{R}_{ij}\big)=\Big(\sum_{t=1}^n\frac{\widehat{X}_{it}^{(n)}\widehat{X}_{jt}^{(n)}}{\sqrt{\sum_{t=1}^n(\widehat{X}_{it}^{(n)})^2}\sqrt{\sum_{t=1}^n(\widehat{X}_{jt}^{(n)})^2}}\Big),\quad i,j=1,\dots,p, \tag{3.4}$$

has largest and smallest eigenvalues $\widehat{\mu}_{(1)}$ and $\widehat{\mu}_{(p)}$, respectively. If $\mathbb{E}[X^4]<\infty$, then by Lemma 2.2 in [40], $\mathbb{P}(\mathbf{X}_n\neq\widehat{\mathbf{X}}_n \text{ i.o.})=0$, which implies

$$\max_{1\le i\le p}\{|\mu_{(i)}-\widehat{\mu}_{(i)}|\}\stackrel{\text{a.s.}}{\to}0.$$

Next, we introduce a condition similar to (C_q). For ease of notation we will again write $(\widehat{Y}_1,\dots,\widehat{Y}_n)=(\widehat{Y}_{11},\dots,\widehat{Y}_{1n})$ and $\widehat{D}=\widehat{D}_1=\sum_{t=1}^n(\widehat{X}_{1t}^{(n)})^2$.

Condition $(\widehat{C}_q)$: There exists a slowly varying sequence $\delta_n\to 0$ such that

$$\lim_{n\to\infty}n^2\,\mathbb{P}(|X|>\delta_n\sqrt{n})=0 \tag{3.5}$$

and 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 $(k^3q)/n\to 0$ and the moment inequality

$$\mathbb{E}[\widehat{Y}_1^{2m_1}\cdots\widehat{Y}_r^{2m_r}]\le\frac{q_n}{n}\,\mathbb{E}[\widehat{Y}_1^{2m_1}\cdots\widehat{Y}_{r-1}^{2m_{(r-1)}}\widehat{Y}_r^{(2m_r-2)}] \tag{3.6}$$

holds for $1 \le r \le \ell - 1$ and any positive integers $m_1, \ldots, m_r$ satisfying $m_1 + \cdots + m_r = \ell$, where $\ell \le k$.

Remark 3.5. Loosely speaking, condition (C_q) fails if the probability of one Y_i^2 being much larger than the sum of the other Y_j^2 is "not small enough". In some cases it might be easier to check condition (3.6) instead of (C_q) since one is allowed to include an extra truncation step which alleviates the aforementioned issue at the cost of the additional restriction (3.5).

We have the following version of Theorem 3.3 part (2).

Theorem 3.6. *Assume* (G_γ).

(2') *If* X *is symmetric and satisfies condition* $(\widehat{C}_q)$, *then*

$$\widehat{\mu}_{(1)} \stackrel{\text{a.s.}}{\to} (1+\sqrt{\gamma})^2 \quad \text{and} \quad \widehat{\mu}_{(p)} \stackrel{\text{a.s.}}{\to} (1-\sqrt{\gamma})^2. \tag{3.7}$$

Proof. The truncated entries $\widehat{X}_{it}$ satisfy condition (C_q) and are symmetric. Following the proof of Theorem 3.3 under condition (C_q) we get (3.7). □

Eq. (3.5) in condition $(\widehat{C}_q)$ is designed for comparison of the eigenvalues of $\mathbf{R}$ and $\widehat{\mathbf{R}}$. Under (3.5), one has $\mathbb{P}(\mathbf{X}_n \neq \widehat{\mathbf{X}}_n) \le c\, n^2\, \mathbb{P}(|X| > \delta_n \sqrt{n}) \to 0$, which implies

$$\max_{1\le i\le p} \{|\mu_{(i)} - \widehat{\mu}_{(i)}|\} \stackrel{\mathbb{P}}{\to} 0. \tag{3.8}$$

Next we present a rather crude way to check $(\widehat{C}_q)$ for distributions with finite $(4+\varepsilon)$ moment.

Corollary 3.7. *Let* $\varepsilon > 0$. *If* $\mathbb{E}[X] = 0$, $|X| \ge c > 0$ *and* $\mathbb{E}[|X|^{4+\varepsilon}] < \infty$, *then condition* $(\widehat{C}_q)$ *is valid and* $\delta_n > 0$ *can be chosen as any slowly varying sequence tending to* 0.

Proof. If $\mathbb{E}[X] = 0$ and $\mathbb{E}[|X|^{4+\varepsilon}] < \infty$, Eq. (3.5) and the fact that (δ_n) can be chosen as any slowly varying sequence follow from the Borel–Cantelli lemma and the proof of Lemma 2.2 in [40]. Since $|X| \ge c > 0$ and $\widehat{X}^2 \le \delta_n^2 n$ we get

$$\begin{aligned} \mathbb{E}[\widehat{Y}_1^{2m_1} \cdots \widehat{Y}_r^{2m_r}] &\le \delta_n^2 n\, \mathbb{E}[\widehat{Y}_1^{2m_1} \cdots \widehat{Y}_{r-1}^{2m_{r-1}} \widehat{Y}_r^{2m_r-2} D^{-1}] \\ &\le \delta_n^2\, \mathbb{E}[\widehat{Y}_1^{2m_1} \cdots \widehat{Y}_{r-1}^{2m_{r-1}} \widehat{Y}_r^{2m_r-2}]/c \\ &= \frac{q_n}{n}\, \mathbb{E}[\widehat{Y}_1^{2m_1} \cdots \widehat{Y}_{r-1}^{2m_{r-1}} \widehat{Y}_r^{2m_r-2}], \end{aligned} \tag{3.9}$$

for $q_n = n\delta_n^2/c$. Choose $\delta_n = (\log n)^{-2}$ and $k_n = (\log n)^{1+\varepsilon/3}$ with $\varepsilon \in (0,1)$. We calculate

$$\frac{k_n^3 q_n}{n} = c^{-1} (\log n)^{3+\varepsilon-4} \to 0,$$

which finishes the proof. □

If $\mathbb{E}[X^4] = \infty$, a slowly varying sequence (δ_n) that satisfies (3.5) does not always exist. For instance if $|X|$ is regularly varying with index $\alpha > 0$, we have

$$n^2 \mathbb{P}(|X| > \delta_n \sqrt{n}) = n^{2-\alpha/2} \delta_n^{-\alpha} \ell(\delta_n \sqrt{n})$$

for some slowly varying function ℓ. For a slowly varying sequence $\delta_n \to 0$ we therefore have

$$\lim_{n\to\infty} n^2\,\mathbb{P}(|X| > \delta_n \sqrt{n}) = \begin{cases} 0, & \text{if } \alpha > 4, \\ \infty, & \text{if } \alpha < 4. \end{cases} \tag{3.10}$$

As a consequence we need $\alpha = 4$ and a particular relationship between ℓ and δ_n for both $\mathbb{E}[X^4] = \infty$ and (3.5) to hold. An example of such a distribution is presented next.

Example 3.8 (*Weak Limit of Extreme Eigenvalues Under Infinite Fourth Moment*)**.** Let $u > 4$ and $s \in (4/u, 1)$. Consider a symmetric random variable X satisfying $|X| \ge c > 0$ and

$$\mathbb{P}(|X| > x) = x^{-4}(\log x)^{-s}, \qquad x \ge x_0.$$

Then $\mathbb{E}[X^4] = \infty$. If $p/n \to \gamma \in (0, 1]$, we have for the largest eigenvalue $\mu_{(1)}$ and the smallest eigenvalue $\mu_{(p)}$ of the sample correlation matrix $\mathbf{R}$:

$$\mu_{(1)} \xrightarrow{\mathbb{P}} (1+\sqrt{\gamma})^2 \quad \text{and} \quad \mu_{(p)} \xrightarrow{\mathbb{P}} (1-\sqrt{\gamma})^2. \tag{3.11}$$

Using the crude bounds in (3.9), the almost sure convergence of the extreme eigenvalues is replaced by weak convergence.

Proof of 3.11. For the slowly varying sequence $\delta_n = (\log n)^{-u}$ and $\widehat{X}_{it}^{(n)} = X_{it}\mathbf{1}_{\{|X_{it}|\le \delta_n\sqrt{n}\}}$, we have due to $4u - s < 0$,

$$\begin{aligned} \mathbb{P}(\mathbf{X}_n \ne \widehat{\mathbf{X}}_n) &\le cn^2\,\mathbb{P}(|X| > \delta_n\sqrt{n}) \\ &\sim c\,2^s(\log n)^{4u-s} \to 0. \end{aligned}$$

By the proof of Corollary 3.7, condition $(\widehat{C}_q)$ holds. By Theorem 3.6, we have (3.7). Together with (3.8) this finishes the proof. $\square$

3.3. *A remark on the centered sample correlation matrix*

We presented results for the matrices $\mathbf{R}$ and $\mathbf{X}\mathbf{X}'$, 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

$$\widetilde{F}_{ii} = \frac{1}{(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}')_{ii}}, \qquad i = 1, \dots, p.$$

In contrast to (1.10) an application of Weyl's inequality [6] yields

$$n^{-1}|\lambda_{(1)}(\mathbf{X}\mathbf{X}') - \lambda_{(1)}(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}')| \le n^{-1}\|\mathbf{X}\mathbf{X}' - \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}'\|_2, \tag{3.12}$$

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 [21], we let $\mathbf{H} = \mathbf{I}_n - n^{-1}\mathbf{1}\mathbf{1}'$, where $\mathbf{1} = (1, \dots, 1)'$. 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 $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}'$ is 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 [27]. 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 [27] 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. 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 4.7.

Throughout (X_{it}) are iid symmetric, which implies that the Y_{it} are symmetric as well. We will study the moments

$$\sum_{t_1,\dots,t_k=1}^{n} \mathbb{E}[Y_{i_1 t_k} Y_{i_1 t_1} Y_{i_2 t_1} Y_{i_2 t_2} Y_{i_3 t_2} Y_{i_3 t_3} \cdots Y_{i_k t_{k-1}} Y_{i_k t_k}]$$

for $k \ge 1$ and various choices of *paths* $I = (i_1, i_2, \dots, i_k) \in \{1, \dots, p\}^k$. In this case, $\text{length}(I) = k$ is the *length of the path*. We say that a *path* $(i_1, i_2, \dots, i_k)$ is an *r-path* if it contains exactly r distinct components. A path is *canonical* if $i_1 = 1$ and $i_l \le \max\{i_1, \dots, i_{l-1}\}+1$, $l \ge 2$. 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 more details and examples of these path notions consult Section 3.1.2 in [6] and the references therein. For $k \ge 1$, 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.

We will show that $F(I)$ can be often expressed as $F(S(I))$ times a certain power of n, where $S(I)$ is a shorter path than I. We start with two examples.

Example 4.1. Let $I=(1,1,2,2)$ and recall that (Y_{it}) are symmetric. We have

$$\begin{aligned}
F(1,1,2,2) &= \sum_{t_1,\dots,t_4=1}^{n} \mathbb{E}[Y_{1t_4}Y_{1t_1}^2 Y_{1t_3}]\,\mathbb{E}[Y_{2t_2}Y_{2t_3}^2 Y_{2t_4}] \\
&= \sum_{t_2,t_4=1}^{n} \mathbb{E}\Big[Y_{1t_4}\underbrace{\sum_{t_1=1}^{n} Y_{1t_1}^2}_{=1} Y_{1t_3}\Big]\,\mathbb{E}\Big[Y_{2t_2}\underbrace{\sum_{t_3=1}^{n} Y_{2t_3}^2}_{=1} Y_{2t_4}\Big] \\
&= \sum_{t_2=1}^{n} \underbrace{\mathbb{E}[Y_{1t_2}^2]}_{=n^{-1}}\underbrace{\mathbb{E}[Y_{2t_2}^2]}_{=n^{-1}} = n^{-1}.
\end{aligned}$$

Since $F(\emptyset)=n$ we get $F(1,1,2,2)=F(\emptyset)n^{-2}$, where $S(I)=\emptyset$ is interpreted as a path of length zero.

Example 4.2. For $I=(1,2,1,2,3,3)$ we get

$$\begin{aligned}
F(1,2,1,2,3,3) &= \sum_{t_1,t_2,t_3,t_4,t_6=1}^{n} \mathbb{E}[Y_{1t_6}Y_{1t_1}Y_{1t_2}Y_{1t_3}]\,\mathbb{E}[Y_{2t_1}Y_{2t_2}Y_{2t_3}Y_{2t_4}] \\
&\quad \times \underbrace{\mathbb{E}\Big[Y_{3t_4}\sum_{t_5=1}^{n} Y_{3t_5}^2 Y_{3t_6}\Big]}_{=n^{-1}\mathbf{1}_{\{t_4=t_6\}}} \\
&= n^{-1}\sum_{t_1,\dots,t_4=1}^{n} \mathbb{E}[Y_{1t_4}Y_{1t_1}Y_{1t_2}Y_{1t_3}]\,\mathbb{E}[Y_{2t_1}Y_{2t_2}Y_{2t_3}Y_{2t_4}] \\
&= n^{-1}F(1,2,1,2).
\end{aligned}$$

Therefore we have $F(1,2,1,2,3,3)=F(S(I))n^{-1}$ for the shorter path $S(I)=(1,2,1,2)$.

The ideas of Examples 4.1 and 4.2 will be formalized in the path-shortening algorithm and Lemma 4.4. 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 $R=0$, 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 $1\le j\le l$, 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 4.3. 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, $\text{length}(S(I)) \le \text{length}(I)$. 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 $R(I) \le r$.

Lemma 4.4. *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

$$\begin{aligned} F(I) &= \sum_{t_1,\dots,t_{j-1},t_{j+1},\dots,t_k=1}^{n} \mathbb{E}[Y_{i_1 t_k} Y_{i_1 t_1} \cdots Y_{i_{j-1} t_{j-2}} Y_{i_{j-1} t_{j-1}} Y_{i_j t_{j-1}} \sum_{t_j=1}^{n} Y_{i_j t_j}^2 Y_{i_j t_{j+1}} \\ &\qquad Y_{i_{j+2} t_{j+1}} Y_{i_{j+2} t_{j+2}} \cdots Y_{i_k t_{k-1}} Y_{i_k t_k}] \\ &= \sum_{t_1,\dots,t_{j-1},t_{j+1},\dots,t_k=1}^{n} \mathbb{E}[Y_{i_1 t_k} Y_{i_1 t_1} \cdots Y_{i_{j-1} t_{j-2}} Y_{i_{j-1} t_{j-1}} Y_{i_j t_{j-1}} Y_{i_j t_{j+1}} \\ &\qquad Y_{i_{j+2} t_{j+1}} Y_{i_{j+2} t_{j+2}} \cdots Y_{i_k t_{k-1}} Y_{i_k t_k}] = F(S_1(I)), \end{aligned} \tag{4.1}$$

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

$$\begin{aligned} F(I) &= \sum_{t_1,\dots,t_{j-1},t_{j+1},\dots,t_k=1}^{n} \sum_{t_j=1}^{n} \mathbb{E}[Y_{i_j t_{j-1}} Y_{i_j t_j}] \\ &\qquad \times \mathbb{E}[Y_{i_1 t_k} Y_{i_1 t_1} \cdots Y_{i_{j-1} t_{j-2}} Y_{i_{j-1} t_{j-1}} Y_{i_{j+1} t_j} Y_{i_{j+1} t_{j+1}} \cdots Y_{i_k t_{k-1}} Y_{i_k t_k}] \\ &= \sum_{t_1,\dots,t_{j-1},t_{j+1},\dots,t_k=1}^{n} n^{-1} \mathbb{E}[Y_{i_1 t_k} Y_{i_1 t_1} \cdots Y_{i_{j-1} t_{j-2}} Y_{i_{j-1} t_{j-1}} Y_{i_{j+1} t_{j-1}} Y_{i_{j+1} t_{j+1}} \end{aligned}$$

$$\cdots Y_{i_k t_{k-1}} Y_{i_k t_k}]$$
$$= F(S_2(I)) n^{-1}, \tag{4.2}$$

where first and third equality come from writing out the definition of F. For the second equality we used that $t_{j-1} = t_j$ is necessary for $\mathbb{E}[Y_{i_j t_{j-1}} Y_{i_j t_j}]$ to be non-zero and $\mathbb{E}[Y_{i_j t_j}^2] = n^{-1}$. 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. $\square$

Define for $k \geq 1$, a function g by $g(\emptyset) = 1$ and

$$g(I) = \max_{T \in \{1,\ldots,k\}^k} \{|T| : f(I,T) > 0\}, \qquad I \in \{1,\ldots,k\}^k. \tag{4.3}$$

From now on, we assume the I-paths to be canonical.

Lemma 4.5. *Let I be a canonical r-path of length k. For any $T \in \{1,\ldots,k\}^k$ such that $f(I,T) > 0$ we have $|T| \leq k - r + 1$.*

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 $|\{t_{i_1}, t_{i_2}, \ldots, t_{i_{2a}}\}| \leq a$ 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 $|T| \leq k = k - r + 1$. 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 4.4. Clearly, $1 \leq g(I) \leq k$. From the definition of runs in the PSA we have $\mathrm{runs}(I) \leq k - r$. From (4.1) and (4.2) one infers

$$\mathrm{runs}(I) + 1 \leq g(I) \leq k - R(I). \tag{4.4}$$

First, we analyze paths I with $S(I) = \emptyset$, or equivalently $\mathrm{length}(S(I)) = 0$. This implies $\mathrm{runs}(I) = k - r$; otherwise the path-shortening function stops earlier and $S(I) \neq \emptyset$. Therefore, we get from the identity

$$g(I) = g(S(I)) + \mathrm{runs}(I) \tag{4.5}$$

that $g(I) = k - r + 1$, which finishes the proof in the case $S(I) = \emptyset$. Note that (4.5) holds for all I. This follows from the proof of Lemma 4.4.

Next, assume $S(I) \neq \emptyset$. In this case, we immediately see

$$\mathrm{length}(S(I)) = k - R(I) - \mathrm{runs}(I). \tag{4.6}$$

Since each element in $S(I)$ has to appear at least twice and $r \geq 2$ we have $\mathrm{length}(S(I)) \geq 4$. Moreover, $S(I)$ has $r - R(I) \geq 2$ distinct components. As a consequence, it must hold

$$\mathrm{runs}(I) \leq k - r - (r - R(I)) = k - 2r + R(I). \tag{4.7}$$

In view of (4.5), 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 $S(I) = (s_1, \dots, s_{\mathrm{length}(S(I))})$. Since $S(I)$ is canonical, we have

$$\{s_1, \dots, s_{\mathrm{length}(S(I))}\} = \{1, \dots, r - R(I)\}.$$

For $i = 1, \dots, r - R(I)$ define $N_i := |\{1 \le j \le \mathrm{length}(S(I)) : s_j = i\}|$. 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

$$T_i := \{t_j \,:\, j \in L_i\} \quad \text{and} \quad \widetilde{T}_i := (t_j \,:\, j \in L_i), \quad i = 1, \dots, r - R(I).$$

For example, if $I = (1, 2, 1, 2, 3, 3)$, then $k = 6, r = 3$, $N_1 = N_2 = 2$ and we have $(S(I), R(I), \mathrm{runs}(I)) = ((1, 2, 1, 2), 1, 1)$ 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 $\mathrm{length}(\widetilde{T}_i)/2$ distinct t-indices in $\widetilde{T}_i$. Since each t_j appears exactly twice in $(\widetilde{T}_1, \dots, \widetilde{T}_{r-R(I)})$,

$$g(S(I)) \le 0.5 \ \mathrm{length}(S(I)). \tag{4.8}$$

Now we are ready to finish the proof of the lemma. By (4.5), (4.8), (4.6), (4.7), in this order, one obtains

$$\begin{aligned} g(I) = g(S(I)) + \mathrm{runs}(I) &\le \frac{\mathrm{length}(S(I))}{2} + \mathrm{runs}(I) \\ &= \frac{k - R(I) - \mathrm{runs}(I)}{2} + \mathrm{runs}(I) \\ &\le \frac{k - R(I) + k - 2r + R(I)}{2} = k - r. \quad \square \end{aligned}$$

Remark 4.6. 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 $S(I) \ne \emptyset$, the bound $g(I) \le k - r$ is sharp. Consider for instance $I = (1, 2, 1, 2)$, where

$$f(I, T) = \mathbb{E}[Y_{1t_1} Y_{1t_2} Y_{1t_3} Y_{1t_4}]\, \mathbb{E}[Y_{2t_1} Y_{2t_2} Y_{2t_3} Y_{2t_4}] = (\mathbb{E}[Y_{t_1} Y_{t_2} Y_{t_3} Y_{t_4}])^2.$$

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 (4.3) 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

$$d(I) = k - r + 1 - g(I). \tag{4.9}$$

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 $r \le k$,

$$|\mathcal{I}_{r,k}(0)| = \frac{1}{r}\binom{k}{r-1}\binom{k-1}{r-1}. \tag{4.10}$$

Proposition 4.7. *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*

$$F(I) \le 2\, n^{1-r-d(I)} (2k)^{d(I)} q^{d(I)}. \tag{4.11}$$

Proof. $S(I) = \emptyset$ is equivalent to $R(I) = r$. By Lemma 4.4,

$$F(I) = F(S(I))n^{-R(I)} = F(\emptyset)n^{-r} = n^{1-r}.$$

Therefore, we only have to prove (4.11) for paths I with $d(I) \ge 1$. Without loss of generality we assume $S(I)$ is a canonical $(r - R(I))$-path. We use the notation for paths with $S(I) \ne \emptyset$ developed in the proof of Lemma 4.5. We know that

$$S(I) = (\pi_1, \dots, \pi_{\mathrm{length}(S(I))}),$$

where $\pi_1, \dots, \pi_{\mathrm{length}(S(I))}$ is a permutation of the path

$$I_0 = (\underbrace{1, \dots, 1}_{N_1}, \underbrace{2, \dots, 2}_{N_2}, \dots, \underbrace{r - R(I), \dots, r - R(I)}_{N_{r-R(I)}}).$$

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 4.5,

$$g(I_0) = (k - R(I) - \mathrm{runs}(I)) - (r - R(I)) + 1 = k - r - \mathrm{runs}(I) + 1$$

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

$$\prod_{i=1}^{r-R(I)} \mathbb{E}\Big[Y_{i1}^{2m_{i,1}} \cdots Y_{is_i}^{2m_{i,s_i}}\Big] = \prod_{i=1}^{r-R(I)} \mathbb{E}\Big[Y_1^{2m_{i,1}} \cdots Y_{s_i}^{2m_{i,s_i}}\Big], \tag{4.12}$$

where for all $i = 1, \dots, r-R(I)$, $1 \le s_i \le N_i$, $m_{i,j} \ge 1$ for all $j \ge 1$ and $m_{i,1} + \cdots + m_{i,s_i} = N_i$. We write

$$\mathbf{s} = (s_1, \dots, s_{r-R(I)}) \quad \text{and} \quad \mathbf{m}_i = (m_{i,1}, \dots, m_{i,s_i}), \quad i = 1, \dots, r - R(I). \tag{4.13}$$

Observe that in

$$F(I_0) = \sum_{t_1,\ldots,t_{N_1+\cdots+N_{r-R(I)}}=1}^{n} \mathbb{E}\Big[Y_{t_{N_1+\cdots+N_{r-R(I)}}} Y_{t_1}^2 \cdots Y_{t_{N_1-1}}^2 Y_{t_{N_1}}\Big] \cdots$$
$$\cdots \mathbb{E}\Big[Y_{t_{N_1+\cdots+N_{r-R(I)-1}}} Y^2_{t_{N_1+\cdots+N_{r-R(I)-1}+1}}$$
$$\cdots Y^2_{t_{N_1+\cdots+N_{r-R(I)}-1}} Y_{,t_{N_1+\cdots+N_{r-R(I)}}}\Big]$$

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 4.6. 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 $|\{t_1, t_2, t_3, t_4\}| \leq 2$ 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 4.6.

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 (4.8) 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 $L_1 = L_2 = \{1, \ldots, \text{length}(S(I))\}$.

For our purpose we will use a much larger bound, namely

$$(2d(I) + 1)!! = (2d(I) + 1)(2d(I) - 1) \cdots 3 \leq (2d(I) + 1)^{d(I)} \leq (2k)^{d(I)}. \tag{4.14}$$

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 (4.3) is attained. By the above considerations, $|Q_{S(I)}| \leq (2k)^{d(I)}$. 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

$$F(S(I)) \leq \sum_{Q \in Q_{S(I)}} F_Q(S(I)), \tag{4.15}$$

where $F_Q(S(I))$ is defined as follows. Write $Q = (q_1, \ldots, q_{\text{length}(S(I))})$. By construction, $\{q_1, \ldots, q_{\text{length}(S(I))}\} = \{1, \ldots, g(S(I))\}$. Set $K_j = \{1 \leq i \leq \text{length}(S(I)) : q_i = j\}$. Then

$$F_Q(S(I)) = \sum_{\substack{t_1,\ldots,t_{\text{length}(S(I))}=1 \\ t_l = t_m \ \forall l,m \in K_j, 1 \leq j \leq g(S(I))}}^{n} f(S(I), (t_1, \ldots, t_{\text{length}(S(I))})). \tag{4.16}$$

We will show later that

$$F_Q(S(I)) \leq 2\, q^{d(I)} n^{-d(I)} F(I_0), \qquad Q \in Q_{S(I)}. \tag{4.17}$$

Then it follows from (4.15) and (4.17) that

$$F(S(I)) \le \sum_{Q \in \mathcal{Q}_{S(I)}} F_Q(S(I))$$
$$\le (2k)^{d(I)} 2\, q^{d(I)} n^{-d(I)} F(I_0) = 2\,(2k)^{d(I)} q^{d(I)} n^{-d(I)} n^{1-r+R(I)}.$$

Finally, an application of Lemma 4.4 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 (4.17) by matching each of the $n^{g(I_0)-d(I)}$ positive summands in (4.16) with $n^{d(I)}$ of the $n^{g(I_0)}$ positive summands in $F(I_0)$, where we recall that

$$F(I_0) = \sum_{\substack{t_1,\ldots,t_{\mathrm{length}(S(I))}=1 \\ t_{N_1}=t_{N_2}=\cdots=t_{N_1+\cdots+N_{r-R(I)}}}}^{n} f(I_0, (t_1, \ldots, t_{\mathrm{length}(S(I))})). \tag{4.18}$$

By *matching* we mean the following. Assume we want to prove

$$\sum_{i=1}^{n} A_i \le \sum_{j=1}^{m} B_j \tag{4.19}$$

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 (4.19) 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

$$\prod_{i=1}^{r-R(I)} \mathbb{E}\big[Y_1^{2m_{i,1}} \cdots Y_{s_i}^{2m_{i,s_i}}\big]. \tag{4.20}$$

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 (4.13) 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 Eq. (4.12).

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 (4.18) and (4.16), 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

$$2\, F(I_0) = 2 \sum_{y=r-R(I)}^{\mathrm{length}(S(I))} \sum_{(\mathbf{s},\mathbf{m}) \in \mathcal{T}_0(y)} f_{\mathbf{s},\mathbf{m}}, \tag{4.21}$$

$$n^{d(I)}\, F_Q(S(I)) = n^{d(I)} \sum_{y=r-R(I)}^{\mathrm{length}(S(I))} \sum_{(\mathbf{s},\mathbf{m}) \in \mathcal{T}_Q(y)} f_{\mathbf{s},\mathbf{m}}. \tag{4.22}$$

We show (4.17) by a *matching argument*. We start by matching summands in class length$(S(I))$. From (4.20) we see that elements of class 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

$$|\mathcal{T}_0(\text{length}(S(I)))| = n(n-1)\cdots(n-g(I_0)+1),$$
$$|\mathcal{T}_Q(\text{length}(S(I)))| = n(n-1)\cdots(n-g(I_0)+d(I)+1).$$

Therefore, we have

$$n^{d(I)} \sum_{(\mathbf{s},\mathbf{m})\in\mathcal{T}_Q(\text{length}(S(I)))} f_{\mathbf{s},\mathbf{m}} \le 2 \sum_{(\mathbf{s},\mathbf{m})\in\mathcal{T}_0(\text{length}(S(I)))} f_{\mathbf{s},\mathbf{m}}.$$

This shows that for each summand in class length$(S(I))$ on the left-hand side of (4.17) we can find at least one summand of the same type on the right-hand side of (4.17).

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 length$(S(I))-1$, length$(S(I))-2,\ldots$, max(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 (4.17) 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 (4.17). Then match the class length$(S(I))-1$ summands on the left-hand side of (4.17) 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 (4.17).

Note that the factor 2 in (4.17) 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 (4.17), 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 (4.21) and (4.22) 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 (4.17). The general matching strategy applies and the proof of (4.17) is complete. □

5. Proof of Theorem 3.3

The following proposition contains our main technical novelty. Its proof is given after the proof of Theorem 3.3.

Proposition 5.1. *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* $\mathbf{R}$*:*

$$\limsup_{n\to\infty} \mu_{(1)} \le (1 + \sqrt{\gamma})^2 \quad \text{a.s.} \tag{5.1}$$

$$\liminf_{n\to\infty} \mu_{(p)} \ge (1 - \sqrt{\gamma})^2 \quad \text{a.s.} \tag{5.2}$$

Proof of Theorem 3.3. (1) If $\mathbb{E}[X^4] < \infty$, (3.1) and (3.2) hold for any mean zero distribution as seen in (1.11).
(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 3.1(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 5.1 this completes the proof. □

Proof of Eq. (5.1) in Proposition 5.1. We follow [23]. By Borel–Cantelli, (5.1) holds if

$$\sum_{n=1}^{\infty} \mathbb{E}\Big[\Big(\frac{\mu_{(1)}}{z}\Big)^k\Big] < \infty, \tag{5.3}$$

where $z > (1+\sqrt{\gamma})^2$ and $k = k_n \to \infty$. We choose k such that $k/\log n \to \infty$ and $(k^3 q)/n \to 0$, which exists by condition (C_q). Our goal is to show (5.3). 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 (4.9) for the definition of $d(I)$. Hence we have

$$\{1,\ldots,p\}^k = \bigcup_{r=1}^{k}\bigcup_{u=0}^{k-r} \mathcal{J}_{r,k}(u). \tag{5.4}$$

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

$$|\mathcal{J}_{r,k}(u)| = p(p-1)\cdots(p-r+1)|\mathcal{I}_{r,k}(u)|, \tag{5.5}$$

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 $u \geq 1$.

In view of (5.4) and (5.5) we obtain

$$\begin{aligned}
\mathbb{E}[\mathrm{tr}(\mathbf{R})^k] &= \sum_{r=1}^{k}\sum_{u=0}^{k-r}\sum_{I\in\mathcal{J}_{r,k}(u)} F(I) \\
&= \sum_{r=1}^{k} p(p-1)\cdots(p-r+1)\sum_{u=0}^{k-r}\sum_{I\in\mathcal{I}_{r,k}(u)} F(I) \qquad (5.6)\\
&\leq \sum_{r=1}^{k} p^r \sum_{I\in\mathcal{I}_{r,k}(0)} F(I) + \sum_{r=1}^{k-1} p^r \sum_{u=1}^{k-r}|\mathcal{I}_{r,k}(u)| \max_{I\in\mathcal{I}_{r,k}(u)} F(I) =: S_1 + S_2.
\end{aligned}$$

By Proposition 4.7, (4.10) and since $|\mathcal{I}_{r,k}(0)| \leq \binom{k-1}{r-1}^2$, we have

$$S_1 \leq \sum_{r=1}^{k} p^r \binom{k-1}{r-1}^2 n^{1-r} = p\sum_{r=1}^{k}\binom{k-1}{r-1}^2\left(\frac{p}{n}\right)^{r-1}. \tag{5.7}$$

Next we bound S_2. Consider $1 \leq u \leq k-r$. We will see how elements of $\mathcal{I}_{r,k}(u)$ can be constructed by modifying elements of $\mathcal{I}_{r,k}(0)$. Let $I \in \mathcal{I}_{r,k}(u)(N_1,\ldots,N_r)$ be the subset of $\mathcal{I}_{r,k}(u)$ for whose elements the integer i appears exactly N_i times as a component. Here $N_i, i = 1,\ldots,r$ are positive integers satisfying $N_1 + \cdots + N_r = k$. Obviously it is possible to obtain I by permuting the components of any $I_0 \in \mathcal{I}_{r,k}(0)(N_1,\ldots,N_r)$. Consider the following permutation of I_0: two components of I_0 exchange places, all others 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

$$(k-1)+(k-2)+\cdots+1 = \sum_{j=1}^{k-1} j = \frac{(k-1)k}{2} \leq \frac{k^2}{2}$$

ways how two components can switch positions.

Let $u = 1$. For any $I \in \mathcal{I}_{r,k}(u)(N_1,\ldots,N_r)$ there exists an $I_0 \in \mathcal{I}_{r,k}(0)(N_1,\ldots,N_r)$ 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 4.5. This implies

$$|\mathcal{I}_{r,k}(u)| \leq |\mathcal{I}_{r,k}(0)|\frac{k^2}{2}.$$

Similarly, for $1 \le u \le k-r$ and $I \in \mathcal{I}_{r,k}(u)(N_1, \ldots, N_r)$ there exists an $I_0 \in \mathcal{I}_{r,k}(0)(N_1, \ldots, N_r)$ and switching permutations $SP_1, \ldots, SP_u$ such that $I = SP_1 \circ \cdots \circ SP_u(I_0)$, which shows

$$|\mathcal{I}_{r,k}(u)| \le |\mathcal{I}_{r,k}(0)| \Big(\frac{k^2}{2}\Big)^u. \tag{5.8}$$

Now we are ready to bound S_2. From Proposition 4.7 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

$$\begin{aligned} \mathbb{E}[\operatorname{tr}(\mathbf{R})^k] &\le S_1 + S_2 \le p \sum_{r=1}^{k} \binom{k-1}{r-1}^2 \Big(\frac{p}{n}\Big)^{r-1} \Big(1 + 2\Big[\Big(1 - \frac{k^3 q}{n}\Big)^{-1} - 1\Big] \mathbf{1}_{\{r<k\}}\Big) \\ &\le p \sum_{r=1}^{k} \binom{2k-2}{2r-2} \Big(\frac{p}{n}\Big)^{r-1} \Big(1 + 2\Big[\Big(1 - \frac{k^3 q}{n}\Big)^{-1} - 1\Big] \mathbf{1}_{\{r<k\}}\Big) \\ &\le p \sum_{r=0}^{2k-2} \binom{2k-2}{r} \Big(\sqrt{\frac{p}{n}}\Big)^r \Big(2\Big(1 - \frac{k^3 q}{n}\Big)^{-1} - 1\Big)^{2k-2-r} \\ &= \Big[p^{1/(k-1)} \Big(2\Big(1 - \frac{k^3 q}{n}\Big)^{-1} - 1 + \sqrt{\frac{p}{n}}\Big)^2\Big]^{k-1} \le \eta^k, \end{aligned}$$

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 (5.3) which concludes the proof.

5.1. *Proof of Eq. (5.2) in Proposition 5.1*

We start with the following result.

Proposition 5.2. *Assume (G_γ). If the iid entries (X_{it}) are symmetric and satisfy condition (C_q) then*

$$\limsup_{n\to\infty} \|\mathbf{R} - (1+\gamma)\mathbf{I}\|_2 \le 2\sqrt{\gamma} \quad \text{a.s.} \tag{5.9}$$

Proof. The general idea is the same as in the proof of Eq. (5.1) in Proposition 5.1: 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,

$$\mathbb{E}[\operatorname{tr}(\mathbf{R} - (1+\gamma)\mathbf{I})^{2k}] = p\,(1+\gamma)^{2k}\Big[1 + p^{-1}\sum_{i=1}^{2k}\binom{2k}{i}\Big(\frac{-1}{1+\gamma}\Big)^i \mathbb{E}[\operatorname{tr}\mathbf{R}^i]\Big]. \tag{5.10}$$

From (5.6) combined with (4.10) we know that for n sufficiently large

$$\begin{aligned}\mathbb{E}[\operatorname{tr}\mathbf{R}^i] &\geq p\sum_{r=1}^{i}\frac{(p-1)(p-2)\cdots(p-r+1)}{n^{r-1}}\frac{1}{r}\binom{i}{r-1}\binom{i-1}{r-1}\\ &= p\,\beta_i(\gamma)\,(1-\delta_n),\end{aligned} \tag{5.11}$$

where $\delta_n = O(1/n)$. Additionally, we established

$$\mathbb{E}[\operatorname{tr}\mathbf{R}^i] \leq p\,\beta_i(\gamma)\Big(1 + \frac{2k^3 q}{n}\Big)(1+\delta_n). \tag{5.12}$$

Hence, by (5.10), Lemma A.2, and noting that f_k is continuous on $\mathbb{R}$, and $p/n \to \gamma \in (0,1]$, we have for n sufficiently large,

$$\begin{aligned}\mathbb{E}[\operatorname{tr}(\mathbf{R} - (1+\gamma)\mathbf{I})^{2k}] &= p(1+\gamma)^{2k}\Big[1 + \sum_{i=1}^{2k}\binom{2k}{i}\Big(\frac{-1}{1+\gamma}\Big)^i \beta_i(\gamma)\Big]\big(1 + O(2k^3 q_n/n)\big)\\ &= p(1+\gamma)^{2k} f_k(\gamma)\big(1 + O(2k^3 q_n/n)\big)\\ &\leq p\,(1+\gamma)(4\gamma)^k\big(1 + O(2k^3 q_n/n)\big) < z^{2k},\end{aligned}$$

for any $z > 2\sqrt{\gamma}$. The last inequality follows from

$$\lim_{n\to\infty} p^{1/(2k)}\big(1 + 2k^3 q_n/n\big)^{1/(2k)}(1+\gamma)^{1/(2k)} = 1.$$

Using the same Borel–Cantelli argument as in the proof of (5.1), one obtains the desired relation

$$\limsup_{n\to\infty}\|\mathbf{R} - (1+\gamma)\mathbf{I}\|_2 \leq 2\sqrt{\gamma} \quad \text{a.s.} \quad \square$$

With Proposition 5.2 we can finish the proof of (5.2). We have

$$\|\mathbf{R} - (1+\gamma)\mathbf{I}\|_2 = \max\{\mu_{(1)} - (1+\gamma), -\mu_{(p)} + (1+\gamma)\}.$$

From (5.9) we conclude

$$\begin{aligned}\limsup_{n\to\infty}\mu_{(1)} &\leq 2\sqrt{\gamma} + 1 + \gamma = (1+\sqrt{\gamma})^2 \quad \text{a.s.,}\\ \liminf_{n\to\infty}\mu_{(p)} &\geq -2\sqrt{\gamma} + 1 + \gamma = (1-\sqrt{\gamma})^2 \quad \text{a.s.}\end{aligned}$$

6. Proof of Theorem 3.1

6.1. *Proof of Theorem 3.1(1)*

We appeal to the proof of Theorem 2.3 in [9]. The following lemma is a version of Corollary 1.1 in [9].

Lemma 6.1. *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 $\mathbb{E}[Y^4] = o(n^{-1})$,

$$n \operatorname{var}(Y_1 Y_2) \to 0, \tag{6.1}$$

$$V_n = n^2 \sum_{\mathcal{S}} \big(\operatorname{cov}(Y_{i_1} Y_{j_1}, Y_{i_2} Y_{j_2})\big)^2 \to 0, \tag{6.2}$$

then Condition 1 of Theorem 1.1 in [9] 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 $c > 0$,

$$\begin{aligned}
&\mathbb{E}[|\mathbf{Y}_0 \mathbf{B} \mathbf{Y}_0' - \operatorname{tr}(\mathbf{B}\mathbb{E}[\mathbf{Y}_0 \mathbf{Y}_0'])|^2] \\
&\quad = \mathbb{E}\Big[\Big| \sum_{i_1, j_1 = 1}^{n} B_{i_1 j_1} (Y_{i_1} Y_{j_1} - \mathbb{E}[Y_{i_1} Y_{j_1}]) \Big|^2\Big] \\
&\quad = \sum_{i_1, j_1 = 1}^{n} \sum_{i_2, j_2 = 1}^{n} B_{i_1 j_1} B_{i_2 j_2} \operatorname{cov}(Y_{i_1} Y_{j_1}, Y_{i_2} Y_{j_2}) \\
&\quad \le c \big[n \operatorname{var}(Y_1^2) + n \operatorname{var}(Y_{11} Y_{12})\big] + \sum_{\mathcal{S}} B_{i_1 j_1} B_{i_2 j_2} \operatorname{cov}(Y_{i_1} Y_{j_1}, Y_{i_2} Y_{j_2}).
\end{aligned}$$

By assumption, $n \operatorname{var}(Y^2) = n\, (\mathbb{E}[Y^4] - n^{-2}) \to 0$. The second summand converges to zero by (6.1). It is shown in [9] 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 (6.2). $\square$

Remark 6.2. Lemma 6.1 corrects the proof of Theorem 2.3 and Corollary 1.1 in [9]. 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

$$V_n' \geq np^2\big(\mathrm{var}(Y^2)\big)^2 = n^{-1}\, p^2\,(n\,\mathbb{E}[Y^4])^2 - 2\,\frac{p^2}{n^2}\,(n\,\mathbb{E}[Y^4]) + \frac{p^2}{n^3},$$

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 3.1(1). If the distribution of X is in the domain of attraction of the normal law the claim follows from Theorem 2.3 in [9], using our Lemma 6.1.

Now assume the alternative condition (2.2). We will apply Theorem 2.2 in [9] and our Lemma 6.1. 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

$$\begin{aligned} s_{\mathbf{R}}(z) &= \int \frac{1}{x-z}\,\mathrm{d}F_{\mathbf{R}}(x) \\ &= \int \frac{1}{x-z}\,\mathrm{d}\Big(\frac{n}{p} F_{\mathbf{Y}'\mathbf{Y}} - \Big(\frac{n}{p} - 1\Big)\mathbf{1}_{[0,\infty)}\Big)(x) \\ &= \frac{n}{p} s_{\mathbf{Y}'\mathbf{Y}}(z) - \Big(\frac{n}{p} - 1\Big)\frac{1}{-z}, \qquad z \in \mathbb{C}^+, \end{aligned} \tag{6.3}$$

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 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 [24], we have $T_{ij} = O(n^{-1})$ for $i \neq j$ and hence $\|\mathbf{T}\|_2$ is bounded. The empirical spectral distribution

$$F_{\mathbf{T}}(x) = \frac{1}{n}\sum_{j=1}^n \mathbf{1}_{\{\lambda_j(\mathbf{T}) \leq x\}} = \frac{\mathbf{1}_{\{T_{11}+(n-1)T_{12} \leq x\}}}{n} + \frac{n-1}{n}\mathbf{1}_{\{T_{11}-T_{12} \leq x\}}$$

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 6.1. We have

$$\begin{aligned} n\,\mathbb{E}[\mathrm{var}(Y_1Y_2)] &= n\,\big(\mathbb{E}[(Y_1Y_2)^2] - (\mathbb{E}[Y_1Y_2])^2\big) \\ &\leq n\Big(\frac{1}{n(n-1)} - o(n^{-2})\Big) \to 0, \qquad n \to \infty. \end{aligned}$$

This implies (6.1).

Now we turn to V_n in (6.2). 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

$c > 0$,

$$\begin{aligned} V_n &\le c\Big(n^4\left(\operatorname{cov}(Y_1^2, Y_1Y_2)\right)^2 + n^5\left(\operatorname{cov}(Y_1^2, Y_2Y_3)\right)^2 + n^5\left(\operatorname{cov}(Y_1Y_2, Y_2Y_3)\right)^2 \\ &\quad + n^6\left(\operatorname{cov}(Y_1Y_2, Y_3Y_4)\right)^2\Big) \\ &\le c\Big(n^4\left(\mathbb{E}[Y_1^3Y_2] - (1/n)\mathbb{E}[Y_1Y_2]\right)^2 + n^5\left(\mathbb{E}[Y_1^2Y_2Y_3] - (1/n)\mathbb{E}[Y_1Y_2]\right)^2 \\ &\quad + n^5\left(\mathbb{E}[Y_1Y_2^2Y_3] - (\mathbb{E}[Y_1Y_2])^2\right)^2 + n^6\left(\mathbb{E}[Y_1Y_2Y_3Y_4] - (\mathbb{E}[Y_1Y_2])^2\right)^2\Big). \end{aligned}$$

The right-hand side converges to zero in view of assumption (2.2) and because (see [24])

$$\mathbb{E}[Y_1^3Y_2] \le \mathbb{E}[Y_1Y_2], \quad \mathbb{E}[Y_1^2Y_2Y_3] = O(n^{-3}) \quad \text{and} \quad \mathbb{E}[Y_1Y_2Y_3Y_4] = O(n^{-4}).$$

Applications of Theorem 2.2 in [9] and our Lemma 6.1 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 (6.3) 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 (1.4); see (1.6). The proof is complete.

6.2. *Proof of Theorem 3.1(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

$$\widetilde{\beta}_k = \mathbb{E}\Big[\int x^k\,\mathrm{d}F_{\mathbf{R}}(x)\Big] = p^{-1}\mathbb{E}[\operatorname{tr}\mathbf{R}^k] = p^{-1}\sum_{i_1,\ldots,i_k=1}^{p} F(i_1,\ldots,i_k). \tag{6.4}$$

From (5.6) we know that

$$p^{-1}\mathbb{E}[\operatorname{tr}(\mathbf{R})^k] \ge \sum_{r=1}^{k}(p-1)(p-2)\cdots(p-r+1)\Big(\sum_{I\in\mathcal{I}_{r,k}(0)} + \sum_{I\in\mathcal{I}_{r,k}(1)}\Big)F(I) =: S_3 + S_4.$$

By Proposition 4.7 and (4.10), we have

$$\lim_{n\to\infty} S_3 = \sum_{r=1}^{k} \frac{1}{r}\binom{k}{r-1}\binom{k-1}{r-1}\gamma^{r-1} = \beta_k(\gamma), \tag{6.5}$$

which we recognize from (1.5) as the kth 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 4.3 for the definitions of $R(\cdot)$ and $S(\cdot)$. Moreover, by symmetry of Y_{it} we have

$$\begin{aligned} F(1,2,1,2) &= \sum_{t_1,\ldots,t_4=1}^{n} \mathbb{E}[Y_{1t_1}Y_{1t_2}Y_{1t_3}Y_{1t_4}Y_{2t_1}Y_{2t_2}Y_{2t_3}Y_{2t_4}] = \sum_{t_1,\ldots,t_4=1}^{n} (\mathbb{E}[Y_{t_1}Y_{t_2}Y_{t_3}Y_{t_4}])^2 \\ &= \sum_{t_1=1}^{n}(\mathbb{E}[Y_{t_1}^4])^2 + 3\sum_{t_1\neq t_2=1}^{n}(\mathbb{E}[Y_{t_1}^2Y_{t_2}^2])^2 \ge \frac{1}{n}(n\mathbb{E}[Y^4])^2. \end{aligned}$$

By Lemma 4.4 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 (6.5) proves $\liminf_{n\to\infty}\widetilde{\beta}_k > \beta_k(\gamma)$, as desired.

Acknowledgments

A large part of this paper was written during a research visit of JH at the Statistics Department of Columbia University in the City of New York. JH thanks the Department for its great hospitality. JH and TM thank Richard A. Davis, Olivier Wintenberger and Gennady Samorodnitsky for valuable discussions on the subject.

Appendix

In this section we provide some auxiliary tools for the proofs of the main results.

Lemma A.1. *Let $k \in \mathbb{N}$ and $1 \le j \le k$. Then*

$$-\sum_{i=2j-1}^{2k}(-1)^i\binom{2k}{i}\binom{i-1}{2j-2} = 1. \tag{A.1}$$

Proof. For $1 \le j \le k$ we rewrite (A.1) as

$$-\sum_{i=2j-1}^{2k} (-1)^i \binom{2k}{i} \frac{(i-1)!}{(i+1-2j)!} = (2j-2)!. \tag{A.2}$$

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

$$\frac{(i-1)!}{(i+1-2j)!} = (i-1)(i-2)\cdots(i-2j+2),$$

Eq. (A.2) 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, $w^{(2j-2-\ell)}(1) = (2j-2-\ell)!$. 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

$$f_k(x) = 1 + \sum_{i=1}^{2k} \binom{2k}{i} \Big(\frac{-1}{1+x}\Big)^i \sum_{r=1}^{i} \frac{1}{r} \binom{i}{r-1} \binom{i-1}{r-1} x^{r-1}. \tag{A.3}$$

The following is our key lemma.

Lemma A.2. *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 A.1 and its equivalent formulation (A.2).

For $j \in \mathbb{N}$ define

$$
g_j(x) = \frac{(1+x)^{2j-1}}{x^j} f_j(x). \tag{A.4}
$$

We have $g_1(x) = 1$ and $g_2(x) = 2 + x$. A straightforward induction proves the recursion

$$
g_j(x) = \frac{(1+x)^2 g_{j-1}(x) - g_{j-1}(0)}{x}, \quad j \ge 2. \tag{A.5}
$$

From this recursive construction one deduces that $g_j(x)$ is a polynomial of degree $j-1$ with positive coefficients.

Next we show $g_k(x) \le 4^k$. Clearly we have $g_1(x) \le 4$ and $g_2(x) \le 4^2$. Therefore assume

$$
\|g_{k-1}\|_{[0,1]} := \sup_{y\in[0,1]} |g_{k-1}(y)| \le 4^{k-1}.
$$

Then for $x \in [0, 1]$,

$$
\begin{aligned}
g_k(x) &= \frac{x(2+x)g_{k-1}(x) + g_{k-1}(x) - g_{k-1}(0)}{x} \\
&\le (2+x)g_{k-1}(x) + g_{k-1}(1) \le (2+x+1)\|g_{k-1}\|_{[0,1]} \\
&\le (3+x)4^{k-1} \le 4^k.
\end{aligned}
$$

In view of (A.4), this finishes the proof. □

References

[1] R. Adamczak, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann, Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles, J. Amer. Math. Soc. 23 (2) (2010) 535–561.

[2] R. Adamczak, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann, 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] T.W. Anderson, Asymptotic theory for principal component analysis, Ann. Math. Statist. 34 (1963) 122–148.

[4] A. Auffinger, G. Ben Arous, S. Péché, Poisson convergence for the largest eigenvalues of heavy tailed random matrices, Ann. Inst. Henri Poincaré Probab. Stat. 45 (3) (2009) 589–610.

[5] Z. Bai, Z. Fang, Y.-C. Liang, Spectral Theory of Large Dimensional Random Matrices and its Applications to Wireless Communications and Finance Statistics, World Scientific Publishing Co. Pte. Ltd., University of Science and Technology of China Press, Hackensack, NJ, Hefei, 2014, p. xii+220 Random matrix theory and its applications.

[6] Z. Bai, J.W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, second ed., in: Springer Series in Statistics, Springer, New York, 2010, p. xvi+551.

[7] Z.D. Bai, J.W. Silverstein, Y.Q. Yin, A note on the largest eigenvalue of a large-dimensional sample covariance matrix, J. Multivariate Anal. 26 (2) (1988) 166–168.
[8] Z.D. Bai, Y.Q. Yin, Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix, Ann. Probab. 21 (3) (1993) 1275–1294.
[9] Z. Bai, W. Zhou, Large sample covariance matrices without independence structures in columns, Statist. Sinica 18 (2) (2008) 425–442.
[10] M. Banna, Limiting spectral distribution of Gram matrices associated with functionals of β-mixing processes, J. Math. Anal. Appl. 433 (1) (2016) 416–433.
[11] M. Banna, F. Merlevède, Limiting spectral distribution of large sample covariance matrices associated with a class of stationary processes, J. Theoret. Probab. 28 (2) (2015) 745–783.
[12] M. Banna, F. Merlevède, M. Peligrad, On the limiting spectral distribution for a large class of symmetric random matrices with correlated entries, Stochastic Process. Appl. 125 (7) (2015) 2700–2726.
[13] Z. Bao, G. Pan, W. Zhou, Tracy-Widom law for the extreme eigenvalues of sample correlation matrices, Electron. J. Probab. 17 (88) (2012) 32.
[14] R. Bhatia, Matrix Analysis, in: Graduate Texts in Mathematics, vol. 169, Springer-Verlag, New York, 1997, p. xii+347.
[15] R.A. Davis, J. Heiny, T. Mikosch, X. Xie, Extreme value analysis for the sample autocovariance matrices of heavy-tailed multivariate time series, Extremes 19 (3) (2016) 517–547.
[16] R.A. Davis, T. Mikosch, O. Pfaffel, Asymptotic theory for the sample covariance matrix of a heavy-tailed multivariate time series, Stochastic Process. Appl. 126 (3) (2016) 767–799.
[17] R.A. Davis, O. Pfaffel, R. Stelzer, Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails, Stochastic Process. Appl. 124 (1) (2014) 18–50.
[18] E. Dobriban, Efficient computation of limit spectra of sample covariance matrices, Random Matrices Theory Appl. 4 (4) (2015) 1550019 36.
[19] D. Donoho, High-dimensional data analysis: the curses and blessings of dimensionality, Technical Report, Stanford University, 2000.
[20] N. El Karoui, Spectrum estimation for large dimensional covariance matrices using random matrix theory, Ann. Statist. 36 (6) (2008) 2757–2790.
[21] N. El Karoui, Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond, Ann. Appl. Probab. 19 (6) (2009) 2362–2405.
[22] N. El Karoui, E. Purdom, The bootstrap, covariance matrices and PCA in moderate and high-dimensions, Technical Report, University of California, 2016.
[23] S. Geman, A limit theorem for the norm of random matrices, Ann. Probab. 8 (2) (1980) 252–261.
[24] E. Giné, F. Götze, D.M. Mason, When is the student t-statistic asymptotically standard normal? Ann. Probab. 25 (3) (1997) 1514–1531.
[25] J. Heiny, T. Mikosch, Eigenvalues and eigenvectors of heavy-tailed sample covariance matrices with general growth rates: the iid case, Stochastic Process. Appl. (2016) 29.
[26] A. Janssen, T. Mikosch, R. Mohsen, X. Xiaolei, The eigenvalues of the sample covariance matrix of a multivariate heavy-tailed stochastic volatility model, 2016. Available at https://arxiv.org/bbs/1605.02563.
[27] T. Jiang, The limiting distributions of eigenvalues of sample correlation matrices, Sankhyā 66 (1) (2004) 35–48.
[28] I.M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. 29 (2) (2001) 295–327.
[29] F. Jonsson, On the quadratic moment of self-normalized sums, Statist. Probab. Lett. 80 (17-18) (2010) 1289–1296.
[30] V.A. Marčenko, L.A. Pastur, Distribution of eigenvalues in certain sets of random matrices, Mat. Sb. (NS) 72 (114) (1967) 507–536.
[31] D.M. Mason, J. Zinn, Acknowledgment of priority: “When does a randomly weighted self-normalized sum converge in distribution?”, Electron. Commun. Probab. 10 (2005) 297 [Electron. Comm. Probab. **10** (2005), 70–81 (electronic); MR2133894], (electronic).
[32] R.J. Muirhead, Aspects of Multivariate Statistical Theory, in: Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1982, p. xix+673.
[33] A. Soshnikov, Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails, Electron. Commun. Probab. 9 (2004) 82–91 (electronic).
[34] A. Soshnikov, Poisson statistics for the largest eigenvalues in random matrix ensembles, in: Mathematical Physics of Quantum Mechanics, in: Lecture Notes in Phys., vol. 690, Springer, Berlin, 2006, pp. 351–364.
[35] N. Srivastava, R. Vershynin, Covariance estimation for distributions with $2+\varepsilon$ moments, Ann. Probab. 41 (5) (2013) 3081–3111.

[36] K. Tikhomirov, The limit of the smallest singular value of random matrices with i.i.d. entries, Adv. Math. 284 (2015) 1–20.

[37] C.A. Tracy, H. Widom, Distribution functions for largest eigenvalues and their applications, in: Proceedings of the International Congress of Mathematicians, Vol. I, (Beijing, 2002), Higher Ed. Press, Beijing, 2002, pp. 587–596.

[38] H. Xiao, W. Zhou, Almost sure limit of the smallest eigenvalue of some sample correlation matrices, J. Theoret. Probab. 23 (1) (2010) 1–20.

[39] J. Yao, S. Zheng, Z. Bai, Large Sample Covariance Matrices and High-Dimensional Data Analysis, in: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, New York, 2015, p. xiv+308.

[40] Y.Q. Yin, Z.D. Bai, P.R. Krishnaiah, On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix, Probab. Theory Related Fields 78 (4) (1988) 509–521.