

Necessary and sufficient conditions for the Marchenko-Pastur theorem

Pavel Yaskov*

Abstract

We obtain necessary and sufficient conditions for the Marchenko-Pastur theorem for matrices with IID isotropic rows. Our conditions are related to a weak concentration property for certain quadratic forms of the rows.

Keywords: the Marchenko-Pastur theorem; random matrices.

AMS MSC 2010: 60B20.

Submitted to ECP on December 6, 2015, final version accepted on September 20, 2016.

1 Introduction

The Marchenko-Pastur (MP) theorem [15] is a classical result in random matrix theory. It states that, with probability one, the empirical spectral distribution of

$$\widehat{\Sigma}_n = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_{pk} \mathbf{x}_{pk}^\top \quad (1.1)$$

converges weakly to the MP law with parameter $\rho > 0$ as $n \rightarrow \infty$ and $p = p(n) = \rho n + o(n)$ if $\{\mathbf{x}_{pk}\}_{k=1}^n$ are IID copies of an isotropic \mathbb{R}^p -valued random vector \mathbf{x}_p satisfying certain conditions.

In the simplest case, the entries of $\mathbf{x}_p = (X_{p1}, \dots, X_{pp})$ are assumed to be IID copies of a zero-mean random variable with unit variance (e.g., see Theorem 3.6 in [4]). More generally, the entries can be any independent zero-mean random variables that have unit variance and satisfy Lindeberg's condition

$$\lim_{p \rightarrow \infty} \frac{1}{p} \sum_{k=1}^p \mathbb{E} X_{pk}^2 I(|X_{pk}| > \varepsilon \sqrt{p}) = 0 \quad \text{for all } \varepsilon > 0 \quad (1.2)$$

(see [20]). The independence assumption can be relaxed in a number of ways. E.g., in [19], the MP theorem is proved for isotropic \mathbf{x}_p having a log-concave distribution.

All of the above assumptions imply that the quadratic forms $\mathbf{x}_p^\top A_p \mathbf{x}_p$ concentrate near their expectations up to an error term $o(p)$ with probability $1 - o(1)$, where A_p is any $p \times p$ complex matrix with the spectral norm $\|A_p\| \leq 1$. This concentration property is a widely used technical tool in random matrix theory. In fact, this condition alone is sufficient for the MP theorem (see [2], [5], [9], [11], [19], Theorem 19.1.8 in [21], and

*Steklov Mathematical Institute of the Russian Academy of Sciences, Moscow, Russia; National University of Science and Technology MISIS, Russia. E-mail: yaskov@mi.ras.ru

[27]). Recently, it has been proved in [8] that the extreme eigenvalues of $\widehat{\Sigma}_n$ converge in probability to the edges of the support of the limiting MP law when a version of the concentration property holds (see also [26]). There are many papers closely related to the MP theorem, where some other dependence assumptions are considered. E.g., see [1], [6], [7], [12], [16], [17], [18], [22], and [24].

As noted in [1], the concentration property is *not* a necessary condition for the MP theorem. In this paper, we show that this condition becomes necessary and sufficient if we consider only a restricted class of quadratic forms.

The paper is structured as follows. Section 2 contains our main results. Section 3 deals with the proofs. Some additional results are given in an Appendix.

2 Main results

We now introduce some notation that will be used throughout the paper.

For each $p \geq 1$, let \mathbf{x}_p be an isotropic random vector in \mathbb{R}^p , i.e. $\mathbb{E}\mathbf{x}_p\mathbf{x}_p^\top = I_p$ for the $p \times p$ identity matrix I_p . Assume further that all random elements are defined on the same probability space. Let also $\widehat{\Sigma}_n$ be given in (1.1), where $\{\mathbf{x}_{pk}\}_{k=1}^n$ are IID copies of \mathbf{x}_p . In what follows, $\widehat{\Sigma}_n$ and \mathbf{x}_p will be independent.

Define the MP law μ_ρ with parameter $\rho > 0$ by

$$d\mu_\rho = \max\{1 - 1/\rho, 0\} d\delta_0 + \frac{\sqrt{(b-x)(x-a)}}{2\pi x\rho} I(x \in [a, b]) dx,$$

where δ_c is a Dirac function with mass at c , $a = (1 - \sqrt{\rho})^2$, and $b = (1 + \sqrt{\rho})^2$. In this paper, all measures are defined on the Borel σ -algebra of \mathbb{R} . For a real symmetric $p \times p$ matrix A with eigenvalues $\lambda_1, \dots, \lambda_p$, its empirical spectral distribution is given by

$$\mu_A = \frac{1}{p} \sum_{k=1}^p \delta_{\lambda_k}$$

and $\|A\|$ denotes the spectral norm of A .

We can now state our main result (proved in Section 3).

Theorem 2.1. *Let $p = p(n)$ satisfy $p/n \rightarrow \rho > 0$ as $n \rightarrow \infty$. If \mathbf{x}_p is isotropic for all $p = p(n)$, then the following conditions are equivalent:*

- (i) $\mu_{\widehat{\Sigma}_n}$ converges weakly to μ_ρ almost surely as $n \rightarrow \infty$,
- (ii) for all $\varepsilon > 0$,

$$\frac{1}{p} [\mathbf{x}_p^\top (\widehat{\Sigma}_n + \varepsilon I_p)^{-1} \mathbf{x}_p - \text{tr}(\widehat{\Sigma}_n + \varepsilon I_p)^{-1}] \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

Furthermore, (i) implies that $\mathbf{x}_p^\top \mathbf{x}_p/p \xrightarrow{\mathbb{P}} 1$.

Remark 2.2. For isotropic \mathbf{x}_p , the convergence in probability in (ii) can be replaced by the convergence in L_1 . By Jensen's inequality, the latter yields that

- (iii) for all $\varepsilon > 0$ and $A_n(\varepsilon) = \mathbb{E}(\widehat{\Sigma}_n + \varepsilon I_p)^{-1}$,

$$\frac{1}{p} [\mathbf{x}_p^\top A_n(\varepsilon) \mathbf{x}_p - \text{tr}(A_n(\varepsilon))] \xrightarrow{\mathbb{P}} 0, \quad n \rightarrow \infty.$$

Under certain assumptions, (iii) \Rightarrow (ii). E.g., one can assume that $p\mathbb{P}(|\mathbf{x}_p^\top y| > \varepsilon\sqrt{p}) \rightarrow 0$ uniformly in unit $y \in \mathbb{R}^p$ as $p \rightarrow \infty$ for all $\varepsilon > 0$. This will be shown elsewhere.

By Theorem 2.1, the following condition is sufficient for the MP theorem:

$$\begin{aligned} & [\mathbf{x}_p^\top A_p \mathbf{x}_p - \text{tr}(A_p)]/p \xrightarrow{\mathbb{P}} 0 \text{ as } p \rightarrow \infty \text{ for all sequences of real symmetric} \\ & \text{positive semi-definite } p \times p \text{ matrices } A_p \text{ with } \|A_p\| \leq 1. \end{aligned} \tag{2.1}$$

A short proof of this fact for not necessarily isotropic \mathbf{x}_p is given in [27].

Condition (2.1) holds in many cases of interest. In particular, (2.1) \Leftrightarrow (1.2) if each \mathbf{x}_p has zero-mean independent entries with unit variance (see Proposition 2.1 in [27]). More complicated models satisfying (2.1) are given in [2], [3], [5], [9], and [25].

In general, (2.1) does not follow from (ii) in Theorem 2.1. Recall the following example from [1]. Take $p = 2q$ for $q = q(n)$ and consider an isotropic random vector \mathbf{x}_p defined by

$$\mathbf{x}_p = \sqrt{2}(\mathbf{z}_q \xi, \mathbf{z}_q(1 - \xi)),$$

where \mathbf{z}_q is a standard normal vector in \mathbb{R}^q , ξ is a random variable independent of \mathbf{z}_q , and $\mathbb{P}(\xi = \alpha) = 1/2, \alpha \in \{0, 1\}$. Assume also that $n \rightarrow \infty$ and $p/n \rightarrow \rho > 0$.

As either $\xi = 0$ or $1 - \xi = 0$, the matrix $\widehat{\Sigma}_n$ will be block-diagonal with two $q \times q$ diagonal blocks $\widehat{\Sigma}_{n1}$ and $\widehat{\Sigma}_{n2}$. It is easy to verify that each $\mu_{\widehat{\Sigma}_{nk}}$ converges weakly to μ_ρ almost surely and, as a result, the same is true for $\mu_{\widehat{\Sigma}_n}$. Thus, (ii) in Theorem 2.1 holds. However, (2.1) does not hold for $A_p = \Pi_p$ being the orthogonal projection on the first q coordinates since

$$\frac{1}{p}[\mathbf{x}_p^\top \Pi_p \mathbf{x}_p - \text{tr}(\Pi_p)] = \frac{2\xi \mathbf{z}_q^\top \mathbf{z}_q - q}{p} \xrightarrow{\mathbb{P}} \frac{2\xi - 1}{2}, \quad q \rightarrow \infty.$$

We now give necessary and sufficient conditions in the classical setting.

Corollary 2.3. *Let $p = p(n)$ satisfy $p/n \rightarrow \rho > 0$ as $n \rightarrow \infty$. If $\mathbf{x}_p = (X_{p1}, \dots, X_{pp})$ has zero-mean independent entries with unit variance for all $p = p(n)$, then $\mu_{\widehat{\Sigma}_n}$ converges weakly to μ_ρ almost surely as $n \rightarrow \infty$ iff (1.2) holds for given $p = p(n)$.*

This result proved in Section 3 is not new. As far as we know, it was initially obtained by Girko via a different method (see Theorem 4.1 in Chapter 3 in [10]).

3 Proofs

Proof of Theorem 2.1. Let further $n \rightarrow \infty$ and $p = p(n) = \rho n + o(n)$. Recall some useful facts and definitions. For a finite measure μ with support in \mathbb{R}_+ , its Stieltjes transform on \mathbb{R}_+ is given by

$$S(\varepsilon, \mu) = \int_0^\infty \frac{\mu(d\lambda)}{\lambda + \varepsilon}, \quad \varepsilon > 0.$$

The next lemma proved in the Appendix is a version of the Stieltjes continuity theorem.

Lemma 3.1. *Let μ, μ_1, μ_2, \dots be random probability measures with support in \mathbb{R}_+ . Then μ_n converges weakly to μ a.s. iff $\mathbb{P}(S(\varepsilon, \mu_n) \rightarrow S(\varepsilon, \mu)) = 1$ for all $\varepsilon > 0$.*

Denote $S_n(\varepsilon) = S(\varepsilon, \mu_{\widehat{\Sigma}_n})$. Then $S_n(\varepsilon) = p^{-1} \text{tr}(\widehat{\Sigma}_n + \varepsilon I_p)^{-1}$ by the definition of $\mu_{\widehat{\Sigma}_n}$. By the standard martingale argument,¹

$$S_n(\varepsilon) - \mathbb{E}S_n(\varepsilon) \rightarrow 0 \quad \text{a.s.} \tag{3.1}$$

for any $\varepsilon > 0$. The latter and Lemma 3.1 imply that (i) holds iff

$$\mathbb{E}S_n(\varepsilon) \rightarrow S(\varepsilon, \mu_\rho) \quad \text{for all } \varepsilon > 0. \tag{3.2}$$

The next lemma that assumes neither (i) nor (ii) will play a key role in our analysis.

Lemma 3.2. *Under the conditions of Theorem 2.1,*

$$1 = \mathbb{E} \frac{\mathbf{x}_p^\top (\widehat{\Sigma}_n + \varepsilon I_p)^{-1} \mathbf{x}_p / p}{1 + \rho \mathbf{x}_p^\top (\widehat{\Sigma}_n + \varepsilon I_p)^{-1} \mathbf{x}_p / p} + \varepsilon \mathbb{E}S_n(\varepsilon) + o(1)$$

for any $\varepsilon > 0$ as $n \rightarrow \infty$.

¹See Step 1 in the proof of Theorem 1.1 in [5] or Lemma 4.1 in [1] and the trace bound above (4.2).

The proof of Lemma 3.2 is deferred to the Appendix.

Let us now show that (i) \Leftrightarrow (3.3) \Leftrightarrow (ii), where

$$\mathbb{E} \frac{\mathbf{x}_p^\top (\widehat{\Sigma}_n + \varepsilon I_p)^{-1} \mathbf{x}_p / p}{1 + \rho \mathbf{x}_p^\top (\widehat{\Sigma}_n + \varepsilon I_p)^{-1} \mathbf{x}_p / p} = \mathbb{E} \frac{S_n(\varepsilon)}{1 + \rho S_n(\varepsilon)} + o(1) \quad \text{for all } \varepsilon > 0. \quad (3.3)$$

First, we prove that (3.2) \Leftrightarrow (3.3). This will imply that (i) \Leftrightarrow (3.3) as (i) \Leftrightarrow (3.2).

Assume that (3.3) holds. By (3.1) and the dominated convergence theorem,

$$\mathbb{E} \frac{S_n(\varepsilon)}{1 + \rho S_n(\varepsilon)} = \frac{\mathbb{E} S_n(\varepsilon)}{1 + \rho \mathbb{E} S_n(\varepsilon)} + o(1). \quad (3.4)$$

Therefore, Lemma 3.2 yields

$$1 = \frac{\mathbb{E} S_n(\varepsilon)}{1 + \rho \mathbb{E} S_n(\varepsilon)} + \varepsilon \mathbb{E} S_n(\varepsilon) + o(1) \quad (3.5)$$

and we see that $\mathbb{E} S_n(\varepsilon)$ converges to the unique positive solution of the equation

$$1 = \frac{S}{1 + \rho S} + \varepsilon S. \quad (3.6)$$

Lemma 3.3. For all $\varepsilon > 0$, $S = S(\varepsilon, \mu_\rho)$ is a unique positive root of (3.6).

Lemma 3.3 is proved in the Appendix. Combining this lemma with (3.4) and (3.5), we get (3.3) \Rightarrow (3.2). Conversely, assume that (3.2) holds. By Lemma 3.3, (3.2) \Rightarrow (3.5). Using Lemma 3.2 and (3.4), we see that (3.5) \Rightarrow (3.3).

We have proved that (3.2) \Leftrightarrow (3.3) and, as a result, (i) \Leftrightarrow (3.3). Now, we need to verify that (3.3) \Leftrightarrow (ii). If (ii) holds, then (3.3) holds by the following fact: if $\xi_n \xrightarrow{\mathbb{P}} 0$ and there is $C > 0$ such that $\mathbb{P}(|\xi_n| \leq C) = 1$ for every $n \geq 1$, then $\mathbb{E} \xi_n \rightarrow 0$.

Suppose (3.3) holds. Note that, by $\mathbb{E} \mathbf{x}_p \mathbf{x}_p^\top = I_p$ and the independence of \mathbf{x}_p and $\widehat{\Sigma}_n$,

$$\mathbb{E}[\mathbf{x}_p^\top (\widehat{\Sigma}_n + \varepsilon I_p)^{-1} \mathbf{x}_p | \widehat{\Sigma}_n] = \text{tr}(\widehat{\Sigma}_n + \varepsilon I_p)^{-1} = p S_n(\varepsilon).$$

Then (ii) follows from (3.3) and the next lemma, where we put $Z_n = \rho \mathbf{x}_p^\top (\widehat{\Sigma}_n + \varepsilon I_p)^{-1} \mathbf{x}_p / p$ and $Y_n = \widehat{\Sigma}_n$ (for a proof, see the Appendix).

Lemma 3.4. Let $\{Z_n\}_{n=1}^\infty$ be non-negative random variables such that $\mathbb{E} Z_n$ is bounded over n . If $Y_n, n \geq 1$, are random elements satisfying

$$\mathbb{E} \frac{Z_n}{1 + Z_n} - \mathbb{E} \frac{\mathbb{E}(Z_n | Y_n)}{1 + \mathbb{E}(Z_n | Y_n)} \rightarrow 0, \quad n \rightarrow \infty,$$

then $Z_n - \mathbb{E}(Z_n | Y_n) \xrightarrow{\mathbb{P}} 0$.

We have proved that (i) \Leftrightarrow (3.3) \Leftrightarrow (ii). Let us show that (ii) implies that $\mathbf{x}_p^\top \mathbf{x}_p / p \xrightarrow{\mathbb{P}} 1$. Suppose (i)–(ii) hold. Then

$$\frac{1}{p} [\mathbf{x}_p^\top (\varepsilon \widehat{\Sigma}_n + I_p)^{-1} \mathbf{x}_p - \text{tr}(\varepsilon \widehat{\Sigma}_n + I_p)^{-1}] \xrightarrow{\mathbb{P}} 0$$

for any given $\varepsilon > 0$. Hence, we can find $\{\varepsilon_n\}_{n=1}^\infty$ that slowly tend to 0 and are such that

$$\Delta_n = \frac{1}{p} [\mathbf{x}_p^\top (\varepsilon_n \widehat{\Sigma}_n + I_p)^{-1} \mathbf{x}_p - \text{tr}(\varepsilon_n \widehat{\Sigma}_n + I_p)^{-1}] \xrightarrow{\mathbb{P}} 0.$$

By (i), $\mu_{\widehat{\Sigma}_n}$ converges weakly to μ_ρ a.s.. The support of μ_ρ is bounded. Hence, writing $\varepsilon_n \widehat{\Sigma}_n = \sum_{k=1}^p \lambda_k e_k e_k^\top$ for some $\lambda_k = \lambda_k(n) \geq 0$ and orthonormal vectors $e_k = e_k(n) \in \mathbb{R}^p$, $k = 1, \dots, p$, we conclude that

$$\frac{1}{p} \sum_{k=1}^p I(\lambda_k > \delta_n) \xrightarrow{\mathbb{P}} 0$$

when $\delta_n = K\varepsilon_n \rightarrow 0$ and $K > 0$ is large enough. In addition, we have

$$\Delta_n - \frac{\mathbf{x}_p^\top \mathbf{x}_p - p}{p} = U_n + V_n,$$

where

$$U_n = \frac{1}{p} \sum_{k: \lambda_k \leq \delta_n} (|\mathbf{x}_p^\top e_k|^2 - 1) \left(\frac{1}{\lambda_k + 1} - 1 \right),$$

$$V_n = \frac{1}{p} \sum_{k: \lambda_k > \delta_n} (|\mathbf{x}_p^\top e_k|^2 - 1) \left(\frac{1}{\lambda_k + 1} - 1 \right).$$

We finish the proof by showing that $U_n \xrightarrow{\mathbb{P}} 0$ and $V_n \xrightarrow{\mathbb{P}} 0$. By the independence of $\widehat{\Sigma}_n$ and \mathbf{x}_p , we have $\mathbb{E}(|\mathbf{x}_p^\top e_k|^2 | \widehat{\Sigma}_n) = e_k^\top e_k = 1$. Furthermore,

$$\mathbb{E}|U_n| = \mathbb{E}[\mathbb{E}(|U_n| | \widehat{\Sigma}_n)] \leq \frac{2}{p} \mathbb{E} \sum_{k: \lambda_k \leq \delta_n} \frac{\lambda_k}{\lambda_k + 1} \leq 2\delta_n = o(1),$$

$$\mathbb{E}|V_n| = \mathbb{E}[\mathbb{E}(|V_n| | \widehat{\Sigma}_n)] \leq \frac{2}{p} \mathbb{E} \sum_{k=1}^p I(\lambda_k > \delta_n) = o(1).$$

Finally, we conclude that $(\mathbf{x}_p^\top \mathbf{x}_p - p)/p = \Delta_n - (U_n + V_n) \xrightarrow{\mathbb{P}} 0$. □

Proof of Corollary 2.3. If Lindeberg's condition (1.2) holds, then $\mu_{\widehat{\Sigma}_n}$ converges weakly to μ_ρ almost surely by Theorem 3.10 in [4]. Conversely, suppose the latter holds. Recall the Gnedenko-Kolmogorov conditions for relative stability (see (A) and (B) in [13]):

if $\{Z_{pk}\}_{p \geq k \geq 1}$ are non-negative independent random variables with $\mathbb{E}Z_{pk} \rightarrow 0$ uniformly in k as $p \rightarrow \infty$ and $\sum_{k=1}^p \mathbb{E}Z_{pk} = 1$ for all $p \geq 1$, then

$$\sum_{k=1}^p Z_{pk} \xrightarrow{\mathbb{P}} 1 \quad \text{iff} \quad \sum_{k=1}^p \mathbb{E}Z_{pk} I(Z_{pk} > \varepsilon) \rightarrow 0 \text{ for all } \varepsilon > 0.$$

As $\mathbb{E}[\mathbf{x}_p^\top \mathbf{x}_p] = p$ and $\mathbf{x}_p^\top \mathbf{x}_p/p \xrightarrow{\mathbb{P}} 1$ by Theorem 2.1, the above conditions yield (1.2). □

4 Appendix

Proof of Lemma 3.1. If μ_n converges weakly to μ a.s., then

$$S(\varepsilon, \mu_n) = \int_{\mathbb{R}_+} \frac{\mu_n(d\lambda)}{\lambda + \varepsilon} = \int_{\mathbb{R}} f d\mu_n \rightarrow \int_{\mathbb{R}} f d\mu = \int_{\mathbb{R}_+} \frac{\mu(d\lambda)}{\lambda + \varepsilon} = S(\varepsilon, \mu) \quad \text{a.s.}$$

for all $\varepsilon > 0$ as $\mu_n(\mathbb{R}_+) = \mu(\mathbb{R}_+) = 1$ a.s. and $f = f(\lambda)$ is a bounded continuous function on \mathbb{R} , where $f(\lambda) = (\lambda + \varepsilon)^{-1}$, $\lambda \geq 0$, and $f(\lambda) = \varepsilon^{-1}$, $\lambda < 0$.

Suppose now $\mathbb{P}(S(\varepsilon, \mu_n) \rightarrow S(\varepsilon, \mu)) = 1$ for all $\varepsilon > 0$. Then

$$\mathbb{P}(S(\varepsilon, \mu_n) \rightarrow S(\varepsilon, \mu) \text{ for all } \varepsilon \in \mathbb{Q} \cap (0, \infty)) = 1.$$

Taking into account that $|S(\varepsilon, \nu) - S(\varepsilon_0, \nu)| \leq |\varepsilon - \varepsilon_0| \nu(\mathbb{R}_+) / (\varepsilon \varepsilon_0)$, $\varepsilon, \varepsilon_0 > 0$, we get

$$\mathbb{P}(S(\varepsilon, \mu_n) \rightarrow S(\varepsilon, \mu) \text{ for all } \varepsilon > 0) = 1.$$

By Theorem 2.2 and Remark 2.3 in [23], the latter implies that $\bar{\mu}_n \rightarrow \bar{\mu}$ vaguely on the compact set $[0, \infty]$ a.s., where, for a finite measure ν on \mathbb{R}_+ , the measure $\bar{\nu}$ on $[0, \infty]$ is defined by $\bar{\nu}(\{\infty\}) = 0$ and

$$\bar{\nu}(B) = \int_B \frac{\nu(d\lambda)}{\lambda + 1} \quad \text{for all Borel sets } B \subseteq \mathbb{R}_+.$$

Necessary and sufficient conditions for the Marchenko-Pastur theorem

The function $f_z(\lambda) = (\lambda + 1)/(\lambda - z)$ with $f(\infty) = 1$ is continuous on $[0, \infty]$ for all $z \in \mathbb{C}$ with $\text{Im}(z) > 0$. Hence, the above vague convergence implies that

$$s(z, \mu_n) = \int_{\mathbb{R}_+} \frac{\mu_n(d\lambda)}{\lambda - z} = \int_{[0, \infty]} f_z d\bar{\mu}_n \rightarrow \int_{[0, \infty]} f_z d\bar{\mu} = \int_{\mathbb{R}_+} \frac{\mu(d\lambda)}{\lambda - z} = s(z, \mu)$$

a.s. for any given z . By the standard Stieltjes continuity theorem (e.g., see Theorem B.9 on page 515 in [4]), $\mu_n \rightarrow \mu$ vaguely a.s.. For probability measures, vague convergence is equivalent to weak convergence. This finishes the proof of the lemma. \square

Proof of Lemma 3.2. Proceeding as in [27], we now do some algebraic computations. Let $\mathbf{x}_{p, n+1} = \mathbf{x}_p$,

$$A_n = n\widehat{\Sigma}_n = \sum_{k=1}^n \mathbf{x}_{pk} \mathbf{x}_{pk}^\top, \quad \text{and} \quad B_n = A_n + \mathbf{x}_p \mathbf{x}_p^\top = \sum_{k=1}^{n+1} \mathbf{x}_{pk} \mathbf{x}_{pk}^\top.$$

For any given $\varepsilon > 0$, the matrix $B_n + \varepsilon n I_p$ is invertible and

$$p = \text{tr}((B_n + \varepsilon n I_p)(B_n + \varepsilon n I_p)^{-1}) = \sum_{k=1}^{n+1} \mathbf{x}_{pk}^\top (B_n + \varepsilon n I_p)^{-1} \mathbf{x}_{pk} + \varepsilon n \text{tr}(B_n + \varepsilon n I_p)^{-1}.$$

Taking expectations and using the exchangeability of $\{\mathbf{x}_{pk}\}_{k=1}^{n+1}$,

$$p = (n+1) \mathbb{E} \mathbf{x}_p^\top (B_n + \varepsilon n I_p)^{-1} \mathbf{x}_p + \varepsilon n \mathbb{E} \text{tr}(B_n + \varepsilon n I_p)^{-1}. \quad (4.1)$$

Recall the Sherman-Morrison formula:

$$(C + x x^\top)^{-1} = C^{-1} - \frac{C^{-1} x x^\top C^{-1}}{1 + x^\top C^{-1} x} \quad \text{if } x \in \mathbb{R}^p, C \in \mathbb{R}^{p \times p} \text{ is positive definite, and } C = C^\top.$$

In particular, by a direct calculation,

$$\begin{aligned} \text{tr} C^{-1} - \text{tr}(C + x x^\top)^{-1} &= \frac{x^\top C^{-2} x}{1 + x^\top C^{-1} x} \leq \|C^{-1}\| \frac{x^\top C^{-1} x}{1 + x^\top C^{-1} x} \leq \|C^{-1}\|, \\ x^\top (C + x x^\top)^{-1} x &= \frac{x^\top C^{-1} x}{1 + x^\top C^{-1} x} \leq 1. \end{aligned} \quad (4.2)$$

Since $\|(A_n + \varepsilon n I_p)^{-1}\| \leq (\varepsilon n)^{-1}$, the latter implies that

$$\mathbb{E} \text{tr}(B_n + \varepsilon n I_p)^{-1} = \mathbb{E} \text{tr}(A_n + \varepsilon n I_p)^{-1} + o(1) \quad \text{and} \quad \mathbb{E} \mathbf{x}_p^\top (B_n + \varepsilon n I_p)^{-1} \mathbf{x}_p = O(1).$$

Thus, by (4.1),

$$p/n = \mathbb{E} \mathbf{x}_p^\top (B_n + \varepsilon n I_p)^{-1} \mathbf{x}_p + \varepsilon \mathbb{E} \text{tr}(A_n + \varepsilon n I_p)^{-1} + o(1). \quad (4.3)$$

Recall that $A_n = n\widehat{\Sigma}_n$. Then $\text{tr}(A_n + \varepsilon n I_p)^{-1} = (p/n) S_n(\varepsilon)$ and, by (4.2),

$$\begin{aligned} \mathbb{E} \mathbf{x}_p^\top (B_n + \varepsilon n I_p)^{-1} \mathbf{x}_p &= \mathbb{E} \frac{\mathbf{x}_p^\top (A_n + \varepsilon n I_p)^{-1} \mathbf{x}_p}{1 + \mathbf{x}_p^\top (A_n + \varepsilon n I_p)^{-1} \mathbf{x}_p} \\ &= \mathbb{E} \frac{\mathbf{x}_p^\top (\widehat{\Sigma}_n + \varepsilon I_p)^{-1} \mathbf{x}_p / n}{1 + \mathbf{x}_p^\top (\widehat{\Sigma}_n + \varepsilon I_p)^{-1} \mathbf{x}_p / n} \\ &= \frac{p}{n} \mathbb{E} \frac{\mathbf{x}_p^\top (\widehat{\Sigma}_n + \varepsilon I_p)^{-1} \mathbf{x}_p / p}{1 + \rho \mathbf{x}_p^\top (\widehat{\Sigma}_n + \varepsilon I_p)^{-1} \mathbf{x}_p / p} + o(1). \end{aligned}$$

Combining the above relations, we obtain the desired result. \square

Proof of Lemma 3.3. Denote further $\mathbb{C}_+ = \{z \in \mathbb{C} : \text{Im}(z) > 0\}$ and let

$$s(z, \mu_\rho) = \int_0^\infty \frac{\mu_\rho(d\lambda)}{\lambda - z}$$

be the Stieltjes transform (on \mathbb{C}_+) of μ_ρ . It is well-known (e.g., see Remark 1.1 in [5]) that $s = s(z, \mu_\rho)$, $z \in \mathbb{C}_+$, is a unique solution in \mathbb{C}_+ of the equation

$$\rho z s^2 + (\rho + z - 1)s + 1 = 0 \quad \text{or, equivalently,} \quad 1 = \frac{s}{1 + \rho s} - z s. \quad (4.4)$$

Letting $z \rightarrow -\varepsilon < 0$ and using the dominated convergence theorem, we conclude that $s(z, \mu_\rho) \rightarrow S(\varepsilon, \mu_\rho)$ and (4.4) becomes (3.6). \square

Proof of Lemma 3.4. We have

$$\begin{aligned} \mathbb{E} \frac{Z_n}{1 + Z_n} - \mathbb{E} \frac{\mathbb{E}(Z_n|Y_n)}{1 + \mathbb{E}(Z_n|Y_n)} &= \mathbb{E} \frac{Z_n - \mathbb{E}(Z_n|Y_n)}{(1 + Z_n)(1 + \mathbb{E}(Z_n|Y_n))} \\ &= \mathbb{E} \frac{Z_n - \mathbb{E}(Z_n|Y_n)}{(1 + \mathbb{E}(Z_n|Y_n))^2} - \mathbb{E} \frac{(Z_n - \mathbb{E}(Z_n|Y_n))^2}{(1 + Z_n)(1 + \mathbb{E}(Z_n|Y_n))^2} \\ &= -\mathbb{E} \frac{(Z_n - \mathbb{E}(Z_n|Y_n))^2}{(1 + Z_n)(1 + \mathbb{E}(Z_n|Y_n))^2}. \end{aligned}$$

As a result, we see that

$$\frac{(Z_n - \mathbb{E}(Z_n|Y_n))^2}{(1 + Z_n)(1 + \mathbb{E}(Z_n|Y_n))^2} \xrightarrow{\mathbb{P}} 0.$$

Since $\mathbb{E}Z_n$ is bounded and $Z_n \geq 0$ a.s., we conclude that Z_n and $\mathbb{E}(Z_n|Y_n)$ are bounded asymptotically in probability and $Z_n - \mathbb{E}(Z_n|Y_n) \xrightarrow{\mathbb{P}} 0$. \square

References

- [1] Adamczak, R.: On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries. *Electronic Journal of Probability*, **16**, (2011), 1065–1095. MR-2820070
- [2] Adamczak, R.: Some remarks on the Dozier-Silverstein theorem for random matrices with dependent entries, *Random Matrices: Theory Appl.*, **2**, (2013), 46 pp. MR-3077829
- [3] Anatolyev, S., and Yaskov, P.: Asymptotics of diagonal elements of projection matrices under many instruments/regressors. *Econometric Theory*, DOI:10.1017/S0266466616000165, (2016).
- [4] Bai, Z., and Silverstein, J.: *Spectral analysis of large dimensional random matrices*. Second edition. New York: Springer, 2010. MR-2567175
- [5] Bai, Z., and Zhou, W.: Large sample covariance matrices without independence structures in columns. *Stat. Sinica*, **18**, (2008), 425–442. MR-2411613
- [6] Banna, M., and Merlevède, F.: Limiting spectral distribution of large sample covariance matrices associated with a class of stationary processes. *J. of Theor. Probab.*, **28**(2), (2015), 745–783. MR-3370674
- [7] Banna, M., Merlevède, F., and Peligrad, M.: On the limiting spectral distribution for a large class of symmetric random matrices with correlated entries. *Stoch. Proc. Appl.*, **125**, (2015), 2700–2726. MR-3332852
- [8] Chafaï, D., and Tikhomirov, K.: On the convergence of the extremal eigenvalues of empirical covariance matrices with dependence. arXiv:1509.02231
- [9] El Karoui, N.: Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.*, **19**, (2009), 2362–2405. MR-2588248

- [10] Girko, V.L.: *Statistical analysis of observations of increasing dimension*. Vol. 28. Springer Science & Business Media, 1995. MR-1473719
- [11] Girko, V., and Gupta, A.K.: Asymptotic behavior of spectral function of empirical covariance matrices. *Random Oper. and Stoch. Eqs.*, **2**(1), (1994), 44–60. MR-1276248
- [12] Götze, F., Naumov, A.A., and Tikhomirov, A.N.: Limit theorems for two classes of random matrices with dependent entries. *Theory Probab. Appl.*, **59**(1), (2015), 23–39. MR-3416062
- [13] Hall, P.: On the L_p convergence of sums of independent random variables. *Math. Proc. Cambridge Philos. Soc.*, **82**, (1977), 439–446. MR-0448489
- [14] Hui, J., and Pan, G.M.: Limiting spectral distribution for large sample covariance matrices with m -dependent elements. *Commun. Stat. – Theory Methods*, **39**, (2010), 935–941. MR-2745352
- [15] Marčenko, V.A., and Pastur, L.A.: Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, **72**, (1967), 507–536. MR-0208649
- [16] Merlève, F., and Peligrad, M.: On the empirical spectral distribution for matrices with long memory and independent rows. *Stoch. Proc. Appl.*, **126**(9), (2016), 2734–2760. MR-3522299
- [17] Merlève, F., Peligrad, C., and Peligrad, M.: On the universality of spectral limit for random matrices with martingale differences entries. *Random Matrices: Theory Appl.*, **4**, 1550003, (2015), 33 pp. MR-3334667
- [18] O’Rourke, S.: A note on the Marchenko-Pastur law for a class of random matrices with dependent entries. *Elect. Comm. Probab.*, **17**, Article 28, 1–13. MR-2955493
- [19] Pajor, A., and Pastur L.: On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution. *Studia Math.*, **195**, (2009), 11–29. MR-2539559
- [20] Pastur, L.: On the spectrum of random matrices. *Teor. Mat. Fiz.*, **10**, (1972), 102–112. MR-0475502
- [21] Pastur, L., and Shcherbina, M.: *Eigenvalue distribution of large random matrices*. Mathematical Surveys and Monographs, **171**. American Mathematical Society, Providence, RI, 2011. MR-2808038
- [22] Pfaffel, O., and Schlemm, E.: Eigenvalue distribution of large sample covariance matrices of linear processes. *Probab. Math. Statist.*, **31**, (2011), 313–329. MR-2853681
- [23] Schilling, R.L., Song, R., and Vondraček, Z.: *Bernstein Functions: Theory and Applications*. De Gruyter Studies in Mathematics, **37**. Walter de Gruyter, Berlin, 2010. MR-2598208
- [24] Yao, J.: A note on a Marčenko-Pastur type theorem for time series. *Statist. Probab. Lett.*, **82**, (2012), 22–28. MR-2863018
- [25] Yaskov, P.: Variance inequalities for quadratic forms with applications. *Math. Methods Statist.*, **24**(4), (2015), 309–319. MR-3437388
- [26] Yaskov, P.: Controlling the least eigenvalue of a random Gram matrix. *Linear Algebra Appl.*, **504**, (2016), 108–123. MR-3502532
- [27] Yaskov, P.: A short proof of the Marchenko-Pastur theorem. *C. R. Math. Acad. Sci. Paris*, **354**, (2016), 319–322. MR-3463031

Acknowledgments. We thank the referee for very useful suggestions that help to make the proofs much more transparent.