C. R. Acad. Sci. Paris, Ser. I 354 (2016) 319–322

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Probability theory

A short proof of the Marchenko–Pastur theorem

Une courte démonstration du théorème de Marchenko–Pastur

Pavel Yaskov [a,b,1]

[a] *Steklov Mathematical Institute of RAS, Russia*
[b] *National University of Science and Technology MISIS, Russia*

ARTICLE INFO

Article history:
Received 5 August 2015
Accepted after revision 14 December 2015
Available online 3 February 2016

Presented by Jean-François le Gall

ABSTRACT

We prove the Marchenko–Pastur theorem for random matrices with i.i.d. columns and a general dependence structure within the columns by a simple modification of the standard Cauchy–Stieltjes resolvent method.

RÉSUMÉ

Nous prouvons le théorème de Marchenko–Pastur pour les matrices aléatoires avec des colonnes i.i.d. et une structure de dépendance générale à l'intérieur des colonnes par une simple modification de la méthode standard résolvante de Cauchy–Stieltjes standard.

1. Introduction

Let $\mathbf{X}_{pn}$ be a $p \times n$ random matrix whose columns $\{\mathbf{x}_{pk}\}_{k=1}^n$ are i.i.d. copies of some random vector $\mathbf{x}_p$ in $\mathbb{R}^p$ for all $p, n \geqslant 1$. All random elements are defined on the same probability space. The object of our study is μ_{pn}, the empirical spectral distribution (ESD) of $n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top$. Here ESD of a $p \times p$ real symmetric matrix A is defined by

$$\mu = \frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i},$$

where δ_λ stands for the Dirac mass at $\lambda \in \mathbb{R}$ and $\lambda_1 \leqslant \ldots \leqslant \lambda_p$ are eigenvalues of A.

Recall that the Marchenko–Pastur law μ_c with parameter $c > 0$ is the probability distribution

$$(1 - 1/c)^+\delta_0 + \frac{\sqrt{(b-x)(x-a)}}{2\pi cx} I(x \in [a, b])\,\mathrm{d}x,$$

E-mail address: yaskov@mi.ras.ru.

[1] The research is supported by the Russian Science Foundation via grant 14-21-00162.

http://dx.doi.org/10.1016/j.crma.2015.12.008
1631-073X/

where $x^+ = \max\{x, 0\}$ for $x \in \mathbb{R}$, $a = (1-\sqrt{c})^2$, and $b = (1+\sqrt{c})^2$.

The Marchenko–Pastur theorem [8] states that, for any $p = p(n)$ with $p/n \to c > 0$ as $n \to \infty$,

$$\mathbb{P}(\mu_{pn} \Rightarrow \mu_c \text{ weakly}, \ n \to \infty) = 1 \tag{1}$$

if each $\mathbf{x}_p$ has centered orthonormal entries $\{X_{pk}\}_{k=1}^p$ satisfying certain conditions. The standard conditions include the independence of $\{X_{pk}\}_{k=1}^p$ and the Lindeberg condition

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

These conditions first appeared in [10]. Succeeding work related to the Marchenko–Pastur theorem were done in many papers (see [5,7,13,14], among others). In particular, Bai and Zhou [3], Pastur and Pajor [9], and Pastur and Shcherbina [11] (see Theorem 19.1.8) proved the Marchenko–Pastur theorem, assuming that $\mathrm{Var}(\mathbf{x}_p^\top A_p \mathbf{x}_p/p) \to 0$, $p \to \infty$, for all sequences of $p \times p$ complex matrices A_p with uniformly bounded spectral norms $\|A_p\|$ (see also [2]). If entries of $\mathbf{x}_p$ are independent, this assumption is much stronger than (2).

In this note we give a short proof of the Marchenko–Pastur theorem under weaker conditions that cover all mentioned results.

2. Main results

Consider the following assumption.

(A) $(\mathbf{x}_p^\top A_p \mathbf{x}_p - \mathrm{tr}(A_p))/p \xrightarrow{\mathbb{P}} 0$ as $p \to \infty$ for all sequences of $p \times p$ complex matrices A_p with uniformly bounded spectral norms $\|A_p\|$.

Theorem 2.1. *If (A) holds, then (1) holds.*

If entries of $\mathbf{x}_p$ are orthonormal, then $\mathbb{E}(\mathbf{x}_p^\top A_p \mathbf{x}_p) = \mathrm{tr}(A_p)$, and the assumption considered in [3,9,11] (see § Introduction) is stronger than (A). In addition, we have the following proposition.

Proposition 2.1. *Let $\{X_{pk}\}_{k=1}^p$ be independent random variables with $\mathbb{E}X_{pk} = 0$, $\mathbb{E}X_{pk}^2 = 1$ for all $p \geqslant k \geqslant 1$. Then (2) holds if and only if (A) holds for $\mathbf{x}_p = (X_{p1}, \ldots, X_{pp})$, $p \geqslant 1$.*

Assumption (A) also covers the case where entries of $\mathbf{x}_p$ are orthonormal infinite linear combinations (in L_2) of some i.i.d. random variables $\{\varepsilon_k\}_{k=1}^\infty$ with $\mathbb{E}\varepsilon_k = 0$ and $\mathbb{E}\varepsilon_k^2 = 1$ (see Corollary 4.9 in arXiv:1410.5190).

Remark. We get an equivalent reformulation of (A) if we consider real symmetric positive semi-definite matrices A_p instead of matrices with complex entries.

3. Proofs

Proof of Theorem 2.1. We will use the Cauchy–Stieltjes transform method. By the Stieltjes continuity theorem (e.g., see Exercise 2.4.10(i) in [12]), we only need to show that $s_n(z) \to s(z)$ a.s. for all $z \in \mathbb{C}$ with $\mathrm{Im}(z) > 0$, where $s_n = s_n(z)$ and $s = s(z)$ are the Stieltjes transforms of μ_{pn} and μ_c defined by

$$s_n(z) = \int_{\mathbb{R}} \frac{\mu_{pn}(\mathrm{d}\lambda)}{\lambda - z} \quad \text{and} \quad s(z) = \int_{\mathbb{R}} \frac{\mu_c(\mathrm{d}\lambda)}{\lambda - z}.$$

By the definition of μ_{pn}, $s_n(z) = \mathrm{tr}(n^{-1}\mathbf{X}_{pn}\mathbf{X}_{pn}^\top - zI_p)^{-1}/p$ for the $p \times p$ identity matrix I_p.

Fix any $z \in \mathbb{C}$ with $v = \mathrm{Im}(z) > 0$. By the standard martingale argument (e.g., see Step 1 in the proof of Theorem 1.1 in [3] or Lemma 4.1 in [1]), we derive that $s_n(z) - \mathbb{E}s_n(z) \to 0$ a.s. We finish the proof by checking that $\mathbb{E}s_n(z) \to s(z)$. We need a technical lemma.

Lemma 3.1. *Let C be a $p \times p$ real symmetric positive semi-definite matrix and $x \in \mathbb{R}^p$. If $z \in \mathbb{C}$ is such that $v = \mathrm{Im}(z) > 0$, then (1) $\|(C - zI_p)^{-1}\| \leqslant 1/v$, (2) $|\mathrm{tr}(C + xx^\top - zI_p)^{-1} - \mathrm{tr}(C - zI_p)^{-1}| \leqslant 1/v$, (3) $|x^\top(C + xx^\top - zI_p)^{-1}x| \leqslant 1 + |z|/v$, (4) $\mathrm{Im}(z + z\mathrm{tr}(C - zI_p)^{-1}) \geqslant v$ and $\mathrm{Im}(\mathrm{tr}(C - zI_p)^{-1}) > 0$, (5) $\mathrm{Im}(z + zx^\top(C - zI_p)^{-1}x) \geqslant v$.*

All bounds in Lemma 3.1 are well known. Part (1) can be proved by diagonalizing C. Part (2) is given in Lemma 2.6 in [4]. Part (3) follows from the Sherman–Morrison formula and Part (5), since

$$x^\top(C+xx^\top-zI_p)^{-1}x=x^\top(C-zI_p)^{-1}x-\frac{(x^\top(C-zI_p)^{-1}x)^2}{1+x^\top(C-zI_p)^{-1}x}=1-\frac{z}{z+zx^\top(C-zI_p)^{-1}x}.$$

Parts (4)–(5) can be checked by showing that $\mathrm{Im}(\mathrm{tr}((1/z)C-I_p)^{-1})\geqslant 0$ and $\mathrm{Im}(x^\top((1/z)C-I_p)^{-1}x)\geqslant 0$.

Take $\mathbf{x}_p=\mathbf{x}_{p,n+1}$ to be independent of $\mathbf{X}_{pn}$ and distributed as $\mathbf{X}_{pn}$'s columns $\{\mathbf{x}_{pk}\}_{k=1}^n$. Define also

$$A_n=\mathbf{X}_{pn}\mathbf{X}_{pn}^\top=\sum_{k=1}^n\mathbf{x}_{pk}\mathbf{x}_{pk}^\top\quad\text{and}\quad B_n=A_n+\mathbf{x}_p\mathbf{x}_p^\top=\sum_{k=1}^{n+1}\mathbf{x}_{pk}\mathbf{x}_{pk}^\top.$$

The matrix B_n-znI_p is non-degenerate and

$$p=\mathrm{tr}\big((B_n-znI_p)(B_n-znI_p)^{-1}\big)=\sum_{k=1}^{n+1}\mathbf{x}_{pk}^\top(B_n-znI_p)^{-1}\mathbf{x}_{pk}-zn\,\mathrm{tr}(B_n-znI_p)^{-1}.$$

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

$$p=(n+1)\mathbb{E}\mathbf{x}_p^\top(B_n-znI_p)^{-1}\mathbf{x}_p-zn\,\mathbb{E}\mathrm{tr}(B_n-znI_p)^{-1}.\tag{3}$$

Define $S_n(z)=\mathrm{tr}(A_n-znI_p)^{-1}$ and note that $S_n(z)=(p/n)s_n(z)$. By Lemma 3.1(2)–(3),

$$\mathbb{E}\mathrm{tr}(B_n-znI_p)^{-1}=\mathbb{E}S_n(z)+O(1/n)\quad\text{and}\quad\mathbb{E}\mathbf{x}_p^\top(B_n-znI_p)^{-1}\mathbf{x}_p=O(1).$$

Moreover, we will show below that

$$\mathbb{E}\mathbf{x}_p^\top(B_n-znI_p)^{-1}\mathbf{x}_p=\frac{\mathbb{E}S_n(z)}{1+\mathbb{E}S_n(z)}+o(1).\tag{4}$$

Suppose for a moment that (4) holds (and $p/n=c+o(1)$). Then (3) reduces to

$$\frac{\mathbb{E}S_n(z)}{1+\mathbb{E}S_n(z)}-z\mathbb{E}S_n(z)=c+o(1).$$

By (1) and (4) in Lemma 3.1, $\mathcal{S}=(\mathbb{E}S_n(z))_{n=1}^\infty$ is a bounded sequence with $\mathrm{Im}(\mathbb{E}S_n(z))>0$, $n\geqslant 1$. Hence, any limiting point of $\mathcal{S}$ has a non-negative imaginary part. In addition, it can be directly checked that the limiting quadratic equation $S/(1+S)-zS=c$ or $zS^2+(z-1+c)S+c=0$ has a unique solution $S=S(z)$ with $\mathrm{Im}(S(z))\geqslant 0$ when $\mathrm{Im}(z)>0$. As a result, any limiting point of $\mathcal{S}$ is equal to $S(z)$. Thus, $\mathbb{E}S_n(z)=(p/n)\mathbb{E}s_n(z)\to S(z)$.

One can also show that $S(z)=cs(z)$ is the above unique solution, where $s(z)$ is the Stieltjes transform of the Marchenko–Pastur law (see Remark 1.1 in [3]). Combining all above relations, we conclude that $s_n(z)\to s(z)$ a.s.

To finish the proof, we only need to check (4). By the Sherman–Morrison formula,

$$\mathbf{x}_p^\top(B_n-znI_p)^{-1}\mathbf{x}_p=\mathbf{x}_p^\top(A_n+\mathbf{x}_p\mathbf{x}_p^\top-znI_p)^{-1}\mathbf{x}_p=\frac{\mathbf{x}_p^\top(A_n-znI_p)^{-1}\mathbf{x}_p}{1+\mathbf{x}_p^\top(A_n-znI_p)^{-1}\mathbf{x}_p}.$$

Using Lemma 3.1(1), (A), and the independence of $\mathbf{x}_p$ and A_n, we get $\mathbf{x}_p^\top(A_n-znI_p)^{-1}\mathbf{x}_p-S_n(z)\xrightarrow{p}0$. We also have

$$S_n(z)-\mathbb{E}S_n(z)=(p/n)(s_n(z)-\mathbb{E}s_n(z))\xrightarrow{p}0.$$

Hence, Lemma 3.1(4)–(5) and the inequality $|1+w|\geqslant\mathrm{Im}(z+zw)/|z|$, $w\in\mathbb{C}$, yield

$$\left|\frac{\mathbf{x}_p^\top(A_n-znI_p)^{-1}\mathbf{x}_p}{1+\mathbf{x}_p^\top(A_n-znI_p)^{-1}\mathbf{x}_p}-\frac{\mathbb{E}S_n(z)}{1+\mathbb{E}S_n(z)}\right|\leqslant\frac{|z|^2}{v^2}|\mathbf{x}_p^\top(A_n-znI_p)^{-1}\mathbf{x}_p-\mathbb{E}S_n(z)|\xrightarrow{p}0.$$

Finally, (4) follows from Lebesgue's dominated convergence theorem and Lemma 3.1(3). □

Proof of Proposition 2.1. For each $p\geqslant 1$, let $A_p=\big(a_{kj}^{(p)}\big)_{k,j=1}^p$ be a complex $p\times p$ matrix with $\|A_p\|\leqslant 1$. If D_p is a diagonal matrix with diagonal entries $(a_{kk}^{(p)})_{k=1}^p$, then

$$\mathbb{E}\Big|\mathbf{x}_p^\top(A_p-D_p)\mathbf{x}_p\Big|^2\leqslant 2\mathbb{E}\Big|\sum_{1\leqslant k<j\leqslant p}a_{kj}^{(p)}X_{pk}X_{pj}\Big|^2+2\mathbb{E}\Big|\sum_{1\leqslant j<k\leqslant p}a_{kj}^{(p)}X_{pk}X_{pj}\Big|^2=2\sum_{j\neq k}|a_{jk}^{(p)}|^2\leqslant 4\mathrm{tr}(A_pA_p^*),$$

where A_p^* is the complex conjugate of A_p. By the definition of the spectral norm, $\mathrm{tr}(A_pA_p^*)\leqslant\|A_p\|^2p$. Thus,

$$\frac{\mathbf{x}_p^\top(A_p-D_p)\mathbf{x}_p}{p}\xrightarrow{p}0.$$

To finish the proof, we need to show that (2) holds if and only if

$$\frac{1}{p}\sum_{k=1}^{p} a_k^{(p)}(X_{pk}^2-1) \xrightarrow{\mathbb{P}} 0 \quad \text{for any triangular array } \{a_k^{(p)}, 1\leqslant k\leqslant p, p\geqslant 1\} \text{ with } |a_k^{(p)}|\leqslant 1. \tag{5}$$

Let (5) hold. Then $Z_p = p^{-1}\sum_{k=1}^{p} X_{pk}^2 \xrightarrow{\mathbb{P}} 1$. Note also that $\mathbb{E}Z_p = 1$ and $Z_p \geqslant 0$ a.s. Extracting almost surely converging subsequences from $\{Z_p\}_{p=1}^{\infty}$ and applying Sheffé's lemma, one can prove that $\mathbb{E}|Z_p - 1| \to 0$. Using inequalities $p^{-1}\sum_{k=1}^{p}\mathbb{E}|X_{pk}^2-1|\leqslant 2$, $p\geqslant 1$, we derive from [6] that

$$\frac{1}{p}\sum_{k=1}^{p}\mathbb{E}|X_{pk}^2-1|I(|X_{pk}^2-1|>\varepsilon p)\to 0 \quad \text{for all } \varepsilon>0.$$

Obviously, this is equivalent to (2).

Let (2) hold. By the Marcinkiewicz–Zygmund inequality, there exists a universal constant $C>0$ such that

$$\mathbb{E}\Big|\frac{1}{p}\sum_{k=1}^{p} a_k^{(p)}(X_{pk}^2-1)\Big| \leqslant \frac{C}{p}\mathbb{E}\Big(\sum_{k=1}^{p}(X_{pk}^2-1)^2\Big)^{1/2}, \text{ where each } a_k^{(p)} \text{ is as in (5).}$$

Using (2), Jensen's inequality, and $\sqrt{x+y}\leqslant\sqrt{x}+\sqrt{y}$, $x,y\geqslant 0$, we get

$$\begin{aligned}\mathbb{E}\Big(\sum_{k=1}^{p}(X_{pk}^2-1)^2\Big)^{1/2} &\leqslant \Big(\sum_{k=1}^{p}\mathbb{E}(X_{pk}^2-1)^2 I(|X_{pk}^2-1|\leqslant\varepsilon p)\Big)^{1/2} + \sum_{k=1}^{p}\mathbb{E}|X_{pk}^2-1|I(|X_{pk}^2-1|>\varepsilon p)\\ &\leqslant p\sqrt{2\varepsilon}+o(p)\end{aligned}$$

for all $\varepsilon>0$, where we also applied the bound $\mathbb{E}(X_{pk}^2-1)^2 I(|X_{pk}^2-1|\leqslant\varepsilon p)\leqslant\varepsilon p\mathbb{E}|X_{pk}^2-1|\leqslant 2\varepsilon p$. Therefore,

$$\overline{\lim_{p\to\infty}}\,\mathbb{E}\Big|\frac{1}{p}\sum_{k=1}^{p} a_k^{(p)}(X_{pk}^2-1)\Big|\leqslant\sqrt{2\varepsilon}.$$

Tending ε to zero, we get (5). □

Acknowledgements

The author thanks the referee for her/his valuable suggestions.

References

[1] R. Adamczak, On the Marchenko–Pastur and circular laws for some classes of random matrices with dependent entries, Electron. J. Probab. 16 (2011) 1065–1095.
[2] R. Adamczak, Some remarks on the Dozier–Silverstein theorem for random matrices with dependent entries, Random Matrices: Theory Appl. 2 (2013) 1250017, 46pp.
[3] Z. Bai, W. Zhou, Large sample covariance matrices without independence structures in columns, Stat. Sin. 18 (2008) 425–442.
[4] Z. Bai, J.W. Silverstein, On the empirical distribution of eigenvalues of a class of large dimensional random matrices, J. Multivar. Anal. 54 (1995) 175–192.
[5] U. Grenander, J.W. Silverstein, Spectral analysis of networks with random topologies, SIAM J. Appl. Math. 32 (2) (1977) 499–519.
[6] P. Hall, On the L_p convergence of sums of independent random variables, Math. Proc. Camb. Philos. Soc. 82 (1977) 439–446.
[7] D. Jonsson, Some limit theorems for the eigenvalues of a sample covariance matrix, J. Multivar. Anal. 12 (1982) 1–38.
[8] V.A. Marcenko, L.A. Pastur, Distribution of eigenvalues in certain sets of random matrices, Mat. Sb. (N. S.) 72 (1967) 507–536.
[9] A. Pajor, L. Pastur, On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution, Stud. Math. 195 (2009) 11–29.
[10] L. Pastur, On the spectrum of random matrices, Teor. Mat. Fiz. 10 (1972) 102–112.
[11] L. Pastur, M. Shcherbina, Eigenvalue Distribution of Large Random Matrices, Mathematical Surveys and Monographs, vol. 171, American Mathematical Society, Providence, RI, USA, 2011.
[12] T. Tao, Topics in Random Matrix Theory, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, USA, 2012.
[13] K.W. Wachter, The strong limits of random matrix spectra for sample matrices of independent elements, Ann. Probab. 6 (1) (1978) 1–18.
[14] Y.Q. Yin, P.R. Krishnaiah, Limit theorem for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic, Teor. Veroâtn. Primen. 30 (4) (1985) 810–816.