ALEA, Lat. Am. J. Probab. Math. Stat. **22**, 1301–1307 (2025)
DOI: 10.30757/ALEA.v22-49

A remark on the spectrum of sample covariance matrices from large random tensors

Pavel Yaskov

Steklov Mathematical Institute of RAS, Moscow, Russia
E-mail address: yaskov@mi-ras.ru

Abstract. We study limiting spectral distributions of sample covariance matrices associated with large isotropic random tensors generated by i.i.d. zero-mean random variables on the unit circle. We consider the symmetric and non-symmetric random tensor models. In the non-symmetric case, we give a short derivation of the result of Yuan (2024).

1. Introduction

For $n \in \mathbb{N}$ and $[n] = \{1, \dots, n\}$, consider a random vector $\mathbf{x} = (\xi_k)_{k\in[n]}$ in $\mathbb{C}^n$ and two models:

The non-symmetric random tensor (RT) *model.* Let $X = \mathbf{x}_1 \otimes \ldots \otimes \mathbf{x}_d$, where $d \in \mathbb{N}$, $\otimes$ is the Kronecker product, and $\mathbf{x}_1, \dots, \mathbf{x}_d$ are independent copies of $\mathbf{x}$. By definition, $X = (x_i)_{i=(i_1,\dots,i_d)\in[n]^d}$ takes values in $\mathbb{C}^p$ for $p = n^d$ and has the entries

$$x_i = \prod_{k=1}^{d} \xi_{i_k k},$$

when $\mathbf{x}_l = (\xi_{kl})_{k\in[n]}$, $l \in [d]$. We write $X \sim \mathrm{RT}(n, d, \xi)$ if the entries of $\mathbf{x}$ are i.i.d. copies of ξ.

The symmetric random tensor (SRT) *model.* Let $X = (x_i)_{i=\{i_1,\dots,i_d\}\subseteq[n],|i|=d}$ be a random vector in $\mathbb{C}^p$ for $p = \binom{n}{d}$ and $d \leqslant n$ with the entries

$$x_i = \prod_{k=1}^{d} \xi_{i_k},$$

here all i_k's are distinct. We write $X \sim \mathrm{SRT}(n, d, \xi)$ if the entries of $\mathbf{x}$ are i.i.d. copies of ξ.

Recall that Z is an *isotropic* random vector if it has zero mean and identity covariance matrix. The above models give isotropic X under the following assumptions: $\mathbf{x}$ is isotropic and X follows the RT model, $\mathbf{x}$ is isotropic with independent entries and X follows the SRT model.

Received by the editors November 10th, 2024; accepted May 23rd, 2025.
2010 *Mathematics Subject Classification.* 60B20, 15B52, 15A18.
Key words and phrases. Random tensors, Sample covariance matrices, Marchenko-Pastur law.
This work was supported by the Russian Science Foundation under grant no. 23-11-00375, https://rscf.ru/en/project/23-11-00375/.

Further, let $\{\tau_1, \tau_2, \ldots\}$ be a sequence of real numbers, and consider the matrix

$$M_{p,q} = \frac{1}{p} \sum_{\alpha=1}^{q} \tau_\alpha X_\alpha X_\alpha^*$$

with X_α being independent copies of an isotropic random vector X in $\mathbb{C}^p$. Pajor and Pastur (2009) considered so called *good vectors* $X = X(p)$ in $\mathbb{C}^p$ for $p \in \mathbb{N}$, i.e., such that

> there are $\varepsilon_p > 0$, $p \in \mathbb{N}$, such that $\varepsilon_p = o(1)$ as $p \to \infty$ and $\mathrm{Var}(X^* A X/p) \leqslant \varepsilon_p \|A\|^2$ for all p, $A \in \mathbb{C}^{p \times p}$, and $\|A\|$ being the spectral norm of A.

They proved the following fundamental result:

Theorem 1.1 (Pajor and Pastur (2009))**.** *Let $X = X(p) \in \mathbb{C}^p$, $p \in \mathbb{N}$, be good isotropic random vectors. If $q = q(p)$ satisfies $q/p \to c > 0$ and*

$$\frac{1}{q} \sum_{\alpha=1}^{q} \delta_{\tau_\alpha} \to \sigma \text{ weakly as } p \to \infty,$$

then, with probability one, the empirical spectral distribution (ESD) of $M_{p,q}$ defined by

$$\frac{1}{p} \sum_{\lambda \text{ in the spectrum of } M_{p,q}} \delta_\lambda$$

weakly converges to a non-random probability measure μ that is uniquely determined by equation

$$S(z) = -\left(z - c \int_{\mathbb{R}} \frac{\tau \sigma(d\tau)}{1 + \tau S(z)} \right)^{-1}$$

for its Stieltjes transform $S(z) = \int_{\mathbb{R}} \mu(d\lambda)/(\lambda - z)$, $z \in \mathbb{C}$, $\mathrm{Im}(z) > 0$. In particular, μ is the Marchenko-Pastur law with parameter $c > 0$, when $\tau_\alpha \equiv 1$ and $\sigma = \delta_1$.

The case $\tau_\alpha \equiv 1$ for good (not necessarily isotropic) random vectors $X = X(p) \in \mathbb{C}^p$ is studied by Bai and Zhou (2008). The literature on limiting spectral distributions (LSD) of sample covariance matrices obtained from a random vector with dependent entries is vast. However, to the best of our knowledge, the most general assumptions that cover most specific data models in the literature and guarantee the universality of LSDs are based on the idea of good vectors or its generalization (e.g., see Adamczak (2013), Dembczak-Kołodziejczyk and Lytova (2022), Girko and Gupta (1994), Pajor and Pastur (2009), Yaskov (2016), Yaskov (2018a), Yaskov (2018b), among others). There are very few examples of data models that give universal LSD with 'bad' vectors (see Adamczak (2011), this case is explained in Yaskov (2024)).

Lytova (2018) (Lemma 3.3) proved that

> if $\mathbf{x} = \mathbf{x}(n) \in \mathbb{C}^n$, $n \in \mathbb{N}$, are isotropic random vectors that are good w.r.t. the sequence $(\varepsilon_n)_{n \in \mathbb{N}}$, then $X = X(n, d)$ following the RT model will be good isotropic random vectors for $d = d(n)$ satisfying $d\varepsilon_n = o(1)$ as $n \to \infty$.

In particular, $X_n \sim \mathrm{RT}(n, d, \xi)$ are good vectors if $d = o(n)$ and $\mathbb{E}\xi = 0$, $\mathbb{E}|\xi| = 1$, $\mathbb{E}|\xi|^4 < \infty$. Thus, Theorem 1.1 gives the LSD of $M_{p,q}$ generated by this RT model under the condition $d = o(n)$, here $p = n^d$. Collins et al. (2022) studied the case $d/n \to \gamma > 0$, additionally assuming that all absolute moments of ξ are finite and imposing moment conditions on τ_α. They derived the limiting moment sequence of ESD of $M_{p,q}$ that differed from the case $d = o(n)$ when $\mathbb{P}(|\xi| = 1) < 1$. The case $|\xi| = 1$ a.s. was analyzed by Yuan (2024), who proved that the limiting moment sequence is the same regardless of $d = d(n)$ when $d \to \infty$. Yuan has used the method of moments, which required extensive computations. In this paper, we give a short derivation that $X_n \sim \mathrm{RT}(n, d, \xi)$ are good vectors when $\mathbb{E}\xi = 0$ and $|\xi| = 1$ a.s. and $d = d(n)$ is arbitrary. As a result, Theorem 1.1 could be applied to get LSD for the corresponding matrix model $M_{p,q}$.

The SRT model was studied by Bryson et al. (2021) and Yaskov (2023). In particular, it was shown by Yaskov (2023) that

> if $d = d(n) \leqslant n$ and $\mathbf{x} = \mathbf{x}(n) \in \mathbb{R}^n$, $n \in \mathbb{N}$, have independent entries distributed as ξ (in $\mathbb{R}$) with $\mathbb{E}\xi = 0$, $\mathbb{E}\xi^2 = 1 > \mathbb{P}(|\xi| = 1)$, $\mathbb{E}\xi^4 < \infty$, then the following conditions are equivalent:
> (i) $X_n \sim \mathrm{SRT}(n, d, \xi)$ are good isotropic random vectors,
> (ii) $d = o(\sqrt{n})$ as $n \to \infty$.

Furthermore, it was proved that the condition $d = o(\sqrt{n})$ is also a necessary and sufficient condition for the Marchenko-Pastur law for this SRT model (when $\mathbb{P}(|\xi| = 1) < 1$). In this paper, we study the case $|\xi| = 1$ a.s. and show that the condition $\min\{d, n-d\} = o(n)$ implies that $X_n \sim \mathrm{SRT}(n, d, \xi)$ are good vectors and, as a result, Theorem 1.1 can be applied.

The results and their proofs are presented in the next section.

2. Main results

Let us state our main result.

Proposition 2.1. *Let ξ be a complex random variable with $\mathbb{E}\xi = 0$ and $|\xi| = 1$ a.s. If $d, n \in \mathbb{N}$, then, for all $A \in \mathbb{C}^{p\times p}$,*
(i) $\mathrm{Var}(X^*AX) \leqslant 2p\,\mathrm{tr}(AA^*)/n$ *when* $X \sim \mathrm{RT}(n, d, \xi)$ *and* $p = n^d$,
(ii) $\mathrm{Var}(X^*AX) \leqslant 4p\,\mathrm{tr}(AA^*)\min\{d, n-d\}/n$ *when* $X \sim \mathrm{SRT}(n, d, \xi)$ *and* $p = \binom{n}{d}$.

The proof of Proposition 2.1 is deferred to the end of this section. As follows from the proof, (i) and (ii) hold for $X = \mathbf{x}_1 \otimes \ldots \otimes \mathbf{x}_d$ and X following the SRT model with $\mathbf{x} = \mathbf{x}_1$, respectively, where $\mathbf{x}_j$ are columns of a random matrix $\mathbf{X} = (\xi_{kl})_{k\in[n], l\in[d]}$ in $\mathbb{C}^{n\times d}$ such that $\mathbb{E}[\xi_{kl}|\xi_{pq}, (p,q) \neq (i,j)] = 0$ and $|\xi_{kl}| = 1$ a.s. for all k, l.

By Proposition 2.1 and the bound $\mathrm{tr}(AA^*) \leqslant \|A\|^2 p$, $X_n \sim \mathrm{RT}(n, d, \xi)$ are good vectors for arbitrary $d = d(n)$ and zero-mean ξ with $|\xi| = 1$ a.s. Also, $X_n \sim \mathrm{SRT}(n, d, \xi)$ are good vectors when $d = d(n)$ is such that $\min\{d, n-d\} = o(n)$ as $n \to \infty$. So, Theorem 1.1 could be applied.

The following proposition shows the sharpness of the condition $\min\{d, n-d\} = o(n)$ in the case of the SRT model.

Proposition 2.2. *Let ξ be a complex random variable with $\mathbb{E}\xi = 0$ and $|\xi| = 1$ a.s. If $d = d(n) \leqslant n$ and $\min\{d, n-d\} \neq o(n)$ as $n \to \infty$, then $X_n \sim \mathrm{SRT}(n, d, \xi)$, $n \in \mathbb{N}$, are not good vectors.*

Let us prove Propositions 2.1 and 2.2.

Proof of Proposition 2.1. (i) Let $X \sim \mathrm{RT}(n, d, \xi)$. Write $X = (x_i)$ and $A = (a_{ij})$, where $i = (i_1, \ldots, i_d), j = (j_1, \ldots, j_d)$ run all elements of $[n]^d$. Set $|i \cap j| = \sum_{s=1}^d \mathbf{1}(i_s = j_s)$ and let $\Delta_t = \sum a_{ij} x_i^* x_j \mathbf{1}(|i \cap j| = t)$ for $t \in [d] \cup \{0\}$, where the sum is over all i, j. Then $X^*AX = \sum_{t=0}^d \Delta_t$, $\mathbb{E}\Delta_t = 0$ for $t < d$, $\Delta_d = \sum a_{ii}$, and

$$\mathrm{Var}(X^*AX) = \mathbb{E}\Big|\sum_{t=0}^{d-1} \Delta_t\Big|^2 = \sum_{s,t=0}^{d-1} \mathbb{E}\Delta_s^*\Delta_t. \tag{2.1}$$

We have

$$\mathbb{E}\Delta_s^*\Delta_t = \sum_{i,j,k,l:|i\cap j|=s,|k\cap l|=t} a_{ij}^* a_{kl} \mathbb{E} x_i x_j^* x_k^* x_l, \tag{2.2}$$

here and in what follows $i = (i_s)_{s=1}^d, j = (j_s)_{s=1}^d, k = (k_s)_{s=1}^d, l = (l_s)_{s=1}^d \in [n]^d$.

Let us show that $\mathbb{E}\Delta_s^*\Delta_t = 0$ when $s \neq t$. To see that the terms $|\mathbb{E} x_i x_j^* x_k^* x_l|$ are zero, we may assume w.l.o.g. that $s > t$ and $i_1 = j_1$, …, $i_s = j_s$ (all other cases could be considered similarly). As $|k \cap l| = t < s$, there is a pair (k_p, l_p) among $(k_1, l_1), \ldots, (k_s, l_s)$ such that $k_p \neq l_p$ and either

$k_p \neq i_p$ or $l_p \neq i_p$, say, $k_p \neq i_p$. Therefore, $\mathbb{E}x_i x_j^* x_k^* x_l = \mathbb{E}[(\xi_{k_p p})^* \times \ldots] = 0$, where everything in $\ldots$ is independent of $\xi_{k_p p}$.

Let us estimate $\mathbb{E}|\Delta_t|^2$ for fixed $t < d$. By the Cauchy-Schwarz inequality,

$$\mathbb{E}|\Delta_t|^2 \leqslant \frac{1}{2} \sum_{i,j,k,l:|i\cap j|=|k\cap l|=t} (|a_{ij}|^2 + |a_{kl}|^2)|\mathbb{E}x_i x_j^* x_k^* x_l| \leqslant \sum_{i,j:|i\cap j|=t} |a_{ij}|^2 \sum_{k,l:|k\cap l|=t} |\mathbb{E}x_i x_j^* x_k^* x_l|. \quad (2.3)$$

As $|\mathbb{E}x_i x_j^* x_k^* x_l| \leqslant \mathbb{E}|x_i x_j^* x_k^* x_l| \leqslant 1$, the very last sum does not exceed the number of its non-zero terms. Let us compute this number.

Recalling that $|i \cap j| = t$, we may assume w.l.o.g. that $i_s = j_s$ for $s \in [t]$ and $i_s \neq j_s$ for $s \in [d] \setminus [t]$ (all other cases could be considered similarly). When $\mathbb{E}x_i x_j^* x_k^* x_l \neq 0$, we necessarily have that $k_s = l_s$ for $s \in [t]$ and $(k_s, l_s) \in \{(i_s, j_s), (j_s, i_s)\}$ for $s \in [d] \setminus [t]$. So

(1) the number of possible choices of $((k_s)_{s=1}^t, (l_s)_{s=1}^t)$ is n^t;

(2) for fixed i, j, the number of possible choices of $((k_s)_{s=t+1}^d, (l_s)_{s=t+1}^d)$ is 2^{d-t}.

Therefore, for all i, j with $|i \cap j| = t$,

$$\sum_{k,l:|k\cap l|=t} |\mathbb{E}x_i x_j^* x_k^* x_l| \leqslant n^t 2^{d-t}$$

and

$$\mathrm{Var}(X^*AX) = \sum_{t=0}^{d-1} \mathbb{E}|\Delta_t|^2 \leqslant n^d \sum_{t=0}^{d-1} \Big(\frac{2}{n}\Big)^{d-t} \sum_{i,j:|i\cap j|=t} |a_{ij}|^2 \leqslant \frac{2n^d}{n} \sum_{t=0}^{d-1} \sum_{i,j:|i\cap j|=t} |a_{ij}|^2 \leqslant 2n^{d-1}\mathrm{tr}(AA^*).$$

(ii) Let $X \sim \mathrm{SRT}(n, d, \xi)$, i.e., $X = (x_i)_{i\subseteq[n],|i|=d}$, where $x_i = \prod_{s\in i} \xi_s$ for $i \subseteq [n]$, $|i| = d$ and ξ_s, $s \in [n]$, being independent copies of ξ. Let $A = (a_{ij})$, where i, j run all d-element subsets of $[n]$. We may assume w.l.o.g. that $d \leqslant n - d$. Indeed, when $n - d < d$, we have $x_i = x^*_{[n]\setminus i} x_{[n]}$ for all i (since $|\xi_s| = 1$ a.s.) and could always write

$$X^*AX = \sum_{i,j\subseteq[n]:|i|=|j|=d} a_{ij} x_i^* x_j = \sum_{i,j\subseteq[n]:|i|=|j|=d} a_{ij} x_{[n]\setminus i} x^*_{[n]\setminus j} = Z^*BZ$$

for $Z \sim \mathrm{SRT}(n, n-d, \xi^*)$ and $B = (b_{ij}) \in \mathbb{C}^{p\times p}$ with $p = \binom{n}{d}$, $b_{ij} = a_{[n]\setminus i,[n]\setminus j}$, and $\mathrm{tr}(B^*B) = \mathrm{tr}(A^*A)$.

Define $i \cap j$ in the standard manner and let $\Delta_t = \sum a_{ij} x_i^* x_j \mathbf{1}(|i \cap j| = t)$, where i, j run all d-element subsets of $[n]$. In particular, Δ_t is zero for all t such that $n < 2(d-t) + t$ (if $|i \cap j| = t$, then $n \geqslant |i\Delta j| + |i \cap j| = 2(d-t) + t$). As in the proof of (i), we have $X^*AX = \sum_{t=0}^d \Delta_t$, $\mathbb{E}\Delta_t = 0$ for $t < d$, $\Delta_d = \sum a_{ii}$ is constant, and

$$\mathrm{Var}(X^*AX) = \mathbb{E}\Big|\sum_{t=0}^{d-1} \Delta_t\Big|^2 = \sum_{s,t=0}^{d-1} \mathbb{E}\Delta_s^*\Delta_t, \quad (2.4)$$

where

$$\mathbb{E}\Delta_s^*\Delta_t = \sum_{i,j,k,l:|i\cap j|=s,|k\cap l|=t} a_{ij}^* a_{kl} \mathbb{E}x_i x_j^* x_k^* x_l \quad (2.5)$$

and here and in what follows, the sum is taken over $i, j, k, l \subseteq [n]$ with $|i| = |j| = |k| = |l| = d$.

Let us show that $\mathbb{E}\Delta_s^*\Delta_t = 0$ when $s \neq t$. To see that the terms $|\mathbb{E}x_i x_j^* x_k^* x_l|$ are zero, we may assume without loss of generality that $s > t$. In this case, there is $\alpha \in (k\Delta l) \setminus (i\Delta j)$ (as $|k\Delta l| = 2(d-t) > 2(d-s) = |i\Delta j|$) and

$$\mathbb{E}x_i x_j^* x_k^* x_l = \mathbb{E}x_{i\setminus j} x^*_{j\setminus i} x^*_{k\setminus l} x_{l\setminus k} = \begin{cases} \mathbb{E}[\xi_\alpha \cdot x_{i\setminus j} x^*_{j\setminus i} x^*_{k\setminus l} x_{l\setminus(k\cup\{\alpha\})}], & \text{if } \alpha \in (l \setminus k) \setminus (i\Delta j), \\ \mathbb{E}[\xi_\alpha^* \cdot x_{i\setminus j} x^*_{j\setminus i} x^*_{k\setminus(l\cup\{\alpha\})} x_{l\setminus k}], & \text{if } \alpha \in (k \setminus l) \setminus (i\Delta j), \end{cases}$$

here ξ_α (or its complex conjugate) is independent of the second factor under the above expectations, as a result, $\mathbb{E}x_i x_j^* x_k^* x_l = 0$. This proves that $\mathbb{E}\Delta_s^*\Delta_t = 0$.

Let us estimate $\mathbb{E}|\Delta_t|^2$ for fixed $t < d$. As in the proof of (i), by the Cauchy-Schwarz inequality,

$$\mathbb{E}|\Delta_t|^2 \leqslant \frac{1}{2} \sum_{i,j,k,l:|i\cap j|=|k\cap l|=t} (|a_{ij}|^2 + |a_{kl}|^2)|\mathbb{E}x_i x_j^* x_k^* x_l| \leqslant \sum_{i,j:|i\cap j|=t} |a_{ij}|^2 \sum_{k,l:|k\cap l|=t} |\mathbb{E}x_i x_j^* x_k^* x_l|. \quad (2.6)$$

As $|\mathbb{E}x_i x_j^* x_k^* x_l| \leqslant \mathbb{E}|x_i x_j^* x_k^* x_l| \leqslant 1$, the very last sum in (2.6) does not exceed the number of its non-zero terms. Let us bound this number.

When $|i \cap j| = |k \cap l| = t(< d)$ and $|\xi| = 1$, we see that $x_i x_j^*$ is the product of ξ_α, $\alpha \in i\Delta j$, (or its complex conjugate ξ_α^*), similarly, $x_k^* x_l$ is the product of ξ_α, $\alpha \in k\Delta l$, (or its complex conjugate ξ_α^*). As a result, when $\mathbb{E}x_i x_j^* x_k^* x_l \neq 0$, we are to have $i\Delta j = k\Delta l$ and $k \cap l \in [n] \setminus (i\Delta j)$. Therefore, there are no more than $\binom{n-2(d-t)}{t}$ possible choices of $k \cap l$ (we have used that $|i\Delta j| = 2(d-t)$); there are no more than $\binom{2(d-t)}{d-t}$ possible choices of $(k \setminus l, l \setminus k)$.
Thus, for all i, j with $|i \cap j| = t$,

$$\sum_{k,l:|k\cap l|=t} |\mathbb{E}x_i x_j^* x_k^* x_l| \leqslant A_t = \binom{2(d-t)}{d-t}\binom{n-2(d-t)}{t}\mathbf{1}(n \geqslant 2(d-t)+t),$$

hereinafter the sum over the empty set iz zero. Using inequality $\binom{a}{b}\binom{c}{d} \leqslant \binom{a+c}{b+d}$ for $a \geqslant b$ and $c \geqslant d$, we see that

$$A_t \leqslant \binom{n}{d} \leqslant \binom{n}{d}\frac{4d}{n}$$

whenever $4d > n$. Let us consider the case $4d \leqslant n$. As is well known, $\binom{2a}{a} \leqslant 2^{2a}$. Also, it follows from $t < d \leqslant n$ that

$$\frac{\binom{n-2d+2t}{t}}{\binom{n}{d}} = \frac{\frac{(n-2d+2t)!}{(n-2d+t)!}}{\frac{(n-d+t)!}{(n-d)!}}\frac{\frac{d!}{t!}}{\frac{n!}{(n-d+t)!}} = \prod_{r=1}^{t}\frac{n-d-(d-t)+r}{n-d+r}\prod_{m=0}^{d-t-1}\frac{d-m}{n-m} \leqslant$$

$$\leqslant \prod_{m=0}^{d-t-1}\frac{d-m}{n-m} = \Big(\frac{d}{n}\Big)^{d-t}\prod_{m=0}^{d-t-1}\frac{1-m/d}{1-m/n} \leqslant \Big(\frac{d}{n}\Big)^{d-t},$$

where $\prod_{r=1}^{t}$ is equal to one if $t = 0$. Hence, for all $t \in [d-1] \cup \{0\}$ and $4d \leqslant n$,

$$A_t \leqslant \binom{n}{d}\Big(\frac{4d}{n}\Big)^{d-t} \leqslant \binom{n}{d}\frac{4d}{n}.$$

Using (2.4) and the above estimates, we finally conclude that

$$\mathrm{Var}(X^*AX) \leqslant \binom{n}{d}\frac{4d}{n}\sum_{t=0}^{d-1}\sum_{i,j:|i\cap j|=t}|a_{ij}|^2 = \binom{n}{d}\frac{4d}{n}\mathrm{tr}(AA^*).$$

The proof of the proposition is finished. □

Proof of Proposition 2.2. Let $\mathbf{x} = (\xi_\alpha)_{\alpha\in[n]}$ generates $X_n \sim \mathrm{SRT}(n, d, \xi)$ for each n, where $\mathbf{x} = \mathbf{x}(n)$ and $\xi_\alpha = \xi_\alpha(n)$, $\alpha \in [n]$, are independent copies of ξ. Take zero-diagonal $A_n = \big(a_{ij}\big)_{i,j\subseteq[n],|i|=|j|=d}$ with $a_{ij} = a_{ij}(n) = \mathbf{1}((i,j) \in \mathcal{S}_n)$ and

$$\mathcal{S}_n = \{(i,j) : i, j \subseteq [n], |i| = |j| = d, |i \cap j| = d-1, i\Delta j = \{1,2\}\}.$$

By definition, A_n is symmetric and for each i with $|\{1,2\} \cap i| = 1$, there is only one j such that $|\{1,2\}\cap j| = 1$ and $a_{ij} = 1$ (such j will be not equal to i as A is zero-diagonal and $a_{ii} = 0 \neq 1 = a_{ij}$). Therefore, the submatrix

$$\bar{A}_n = (a_{ij})_{i,j\subseteq[n],|\{1,2\}\cap i|=|\{1,2\}\cap j|=1,|i|=|j|=d}$$

is the permutation matrix and $\|\bar{A}_n\| = 1$. As $a_{ij} = 0$ when a_{ij} is not an element of $\bar{A}_n$, we have $\|A_n\| = \|\bar{A}_n\| = 1$. Also, if $(i,j) \in \mathcal{S}_n$ and $\iota = i \cap j$, then $|\iota| = d-1$ and either $1 \in i$, $2 \in j$ or $2 \in i$, $1 \in j$. Therefore,

$$\begin{aligned}\operatorname{Var}(X_n^* A_n X_n) &= \sum_{i,j,k,l:(i,j),(k,l)\in\mathcal{S}_n} \mathbb{E} x_i x_j^* x_k^* x_l \\ &= \sum_{\iota,\kappa\subseteq[n]\setminus\{1,2\},|\iota|=|\kappa|=d-1} (\mathbb{E}\xi_1\xi_2^*\xi_1^*\xi_2 + \mathbb{E}\xi_1\xi_2^*\xi_2^*\xi_1 + \mathbb{E}\xi_2\xi_1^*\xi_2^*\xi_1 + \mathbb{E}\xi_2\xi_1^*\xi_1^*\xi_2) \\ &= \sum_{\iota,\kappa\subseteq[n]\setminus\{1,2\},|\iota|=|\kappa|=d-1} (2 + 2|\mathbb{E}\xi^2|) \geqslant 2\Big(\sum_{\iota\subseteq[n]\setminus\{1,2\},|\iota|=d-1} 1\Big)^2 = 2\binom{n-2}{d-1}^2\end{aligned}$$

and

$$\binom{n-2}{d-1}^2 = \binom{n}{d}^2 \left(\frac{d(n-d)}{n(n-1)}\right)^2.$$

This proves that $\operatorname{Var}(X_n^* A_n X_n / \binom{n}{d}) \neq o(1)$ when $\min\{d, n-d\} \neq o(n)$.

□

References

Adamczak, R. On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries. *Electron. J. Probab.*, **16**, no. 37, 1068–1095 (2011). DOI: 10.1214/EJP.v16-899.

Adamczak, R. Some remarks on the Dozier-Silverstein theorem for random matrices with dependent entries. *Random Matrices Theory Appl.*, **2** (2), 1250017, 46 (2013). DOI: 10.1142/S2010326312500177.

Bai, Z. and Zhou, W. Large sample covariance matrices without independence structures in columns. *Statist. Sinica*, **18** (2), 425–442 (2008). MR2411613.

Bryson, J., Vershynin, R., and Zhao, H. Marchenko-Pastur law with relaxed independence conditions. *Random Matrices Theory Appl.*, **10** (4), Paper No. 2150040, 28 (2021). DOI: 10.1142/S2010326321500404.

Collins, B., Yao, J., and Yuan, W. On spectral distribution of sample covariance matrices from large dimensional and large k-fold tensor products. *Electron. J. Probab.*, **27**, Paper No. 102, 18 (2022). DOI: 10.1214/22-ejp825.

Dembczak-Kołodziejczyk, A. and Lytova, A. On the empirical spectral distribution for certain models related to sample covariance matrices with different correlations. *Random Matrices Theory Appl.*, **11** (3), Paper No. 2250030, 23 (2022). DOI: 10.1142/S2010326322500307.

Girko, V. L. and Gupta, A. K. Asymptotic behavior of spectral function of empirical covariance matrices. *Random Oper. Stochastic Equations*, **2** (1), 43–60 (1994). DOI: 10.1515/rose.1994.2.1.43.

Lytova, A. Central limit theorem for linear eigenvalue statistics for a tensor product version of sample covariance matrices. *J. Theoret. Probab.*, **31** (2), 1024–1057 (2018). DOI: 10.1007/s10959-017-0741-9.

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** (1), 11–29 (2009). DOI: 10.4064/sm195-1-2.

Yaskov, P. Necessary and sufficient conditions for the Marchenko-Pastur theorem. *Electron. Commun. Probab.*, **21**, Paper No. 73, 8 (2016). DOI: 10.1214/16-ECP4748.

Yaskov, P. LLN for quadratic forms of long memory time series and its applications in random matrix theory. *J. Theoret. Probab.*, **31** (4), 2032–2055 (2018a). DOI: 10.1007/s10959-017-0767-z.

Yaskov, P. Marchenko-Pastur law for a random tensor model. *Electron. Commun. Probab.*, **28**, Paper No. 23, 17 (2023). DOI: 10.1214/23-ECP527.

Yaskov, P. A. On the spectrum of sample covariance matrices for time series. *Theory Probab. Appl.*, **62** (3), 432–443 (2018b). DOI: 10.1137/S0040585X97T988721.

Yaskov, P. A. On sufficient conditions in the Marchenko-Pastur theorem. *Theory Probab. Appl.*, **68** (4), 657–673 (2024). MR4912046.

Yuan, W. On spectrum of sample covariance matrices from large tensor vectors. *ALEA Lat. Am. J. Probab. Math. Stat.*, **21** (2), 1527–1545 (2024). DOI: 10.30757/alea.v21-57.