



TITLE:

On spectral distribution of sample covariance matrices from large dimensional and large k-fold tensor products

AUTHOR(S):

Collins, Benoît; Yao, Jianfeng; Yuan, Wangjun

CITATION:

Collins, Benoît ...[et al]. On spectral distribution of sample covariance matrices from large dimensional and large k-fold tensor products. *Electronic Journal of Probability* 2022, 27: 1-18: 22-EJP825.

ISSUE DATE:

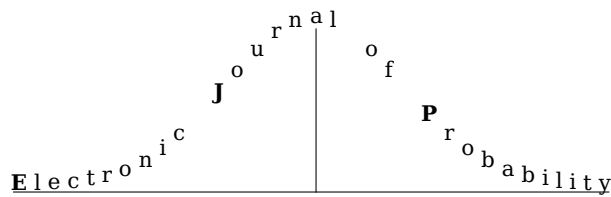
2022

URL:

<http://hdl.handle.net/2433/279255>

RIGHT:

Creative Commons Attribution License



Electron. J. Probab. **27** (2022), article no. 102, 1–18.
ISSN: 1083-6489 <https://doi.org/10.1214/22-EJP825>

On spectral distribution of sample covariance matrices from large dimensional and large k -fold tensor products*

Benoît Collins[†] Jianfeng Yao[‡] Wangjun Yuan[§]

Abstract

We study the eigenvalue distributions for sums of independent rank-one k -fold tensor products of large n -dimensional vectors. Previous results in the literature assume that $k = o(n)$ and show that the eigenvalue distributions converge to the celebrated Marčenko-Pastur law under appropriate moment conditions on the base vectors. In this paper, motivated by quantum information theory, we study the regime where k grows faster, namely $k = O(n)$. We show that the moment sequences of the eigenvalue distributions have a limit, which is different from the Marčenko-Pastur law, and the Marčenko-Pastur law limit holds if and only if $k = o(n)$ for this tensor model. The approach is based on the method of moments.

Keywords: large k -fold tensors; eigenvalue distribution; Marčenko-Pastur law; quantum information theory.

MSC2020 subject classifications: Primary 60B20; Secondary 15B52.

Submitted to EJP on December 22, 2021, final version accepted on July 5, 2022.

1 Introduction

For $n \in \mathbb{N}$, let $\{\xi_1, \dots, \xi_n\}$ be a family of i.i.d. centered complex random variables with unit variance. Denote $\mathbf{y} = \frac{1}{\sqrt{n}}(\xi_1, \dots, \xi_n) \in \mathbb{C}^n$. Let $\{\mathbf{y}_\alpha^{(l)} : 1 \leq \alpha \leq m, 1 \leq l \leq k\}$ be a family of independent copies of \mathbf{y} , and let $Y_\alpha = \mathbf{y}_\alpha^{(1)} \otimes \dots \otimes \mathbf{y}_\alpha^{(k)}$ for $1 \leq \alpha \leq m$. Let $\{\tau_1, \tau_2, \dots\}$ be a sequence of real numbers, and $Y = (Y_1, \dots, Y_m)$ be a $n^k \times m$ matrix. Consider the $n^k \times n^k$ Hermitian matrix

$$M_{n,k,m} = \sum_{\alpha=1}^m \tau_\alpha Y_\alpha Y_\alpha^*. \quad (1.1)$$

*BC was partially supported by JSPS Kakenhi 17H04823, 20K20882, 21H00987, and JPJSBP120203202.

[†]Department of Mathematics, Kyoto University. E-mail: collins@math.kyoto-u.ac.jp

[‡]School of Data Science, The Chinese University of Hong Kong (Shenzhen). E-mail: jeff Yao@cuhk.edu.cn

[§]Department of Mathematics and Statistics, University of Ottawa. E-mail: ywangjun@connect.hku.hk

Spectral distribution of sample covariance matrices from large tensor products

Therefore, each n^k -dimensional vector Y_α is a k -fold tensor product of n -dimensional i.i.d. vectors, and $M_{n,k,m}$ is a sum of m independent rank-1 Hermitian matrices of dimension n^k .

The simplest case of $k = 1$ was studied in the seminal paper [8] where the celebrated Marčenko-Pastur law was derived under appropriate moment conditions on the entries of the base vector \mathbf{y} and the limiting scheme where $n \rightarrow \infty$ and $m/n \rightarrow c > 0$. Many subsequent improvements on the Marčenko-Pastur law appeared in the literature, including [10], [3] and [9]. Notably, the latter paper extended the law to a broad family of \mathbf{y} -vectors, called *good vectors*, that includes the current setting with i.i.d. components. Later, for the setting $\tau_\alpha \equiv 1$, the necessary and sufficient conditions on Y_1 that the Marčenko-Pastur law serves as the limiting spectral distribution (LSD) of $M_{n,1,m}$ were carried out in [11].

Recently, [6] considered the k -fold tensor model $M_{n,k,m}$ and established a LSD for its n^k real-valued eigenvalues. For the special case $\tau_\alpha \equiv 1$, the LSD is exactly the Marčenko-Pastur law. A central limit theorem (CLT) is also established for a class of linear spectral statistics following the approach of [7]. The main setup [6] is that the tensor product number k must be small enough compared to the space dimension n . Precisely, $k/n \rightarrow 0$ is required for the validity of the LSD while $k \equiv 2$ is required when $n \rightarrow \infty$ for the validity of the CLT. It is natural to consider the setting where k is large. Note that larger values of k produce more dependence among the entries of the vector Y_1 : one expects that the LSD of $M_{n,k,m}$ does not obey the Marčenko-Pastur law when k is large enough. However, the martingale moment bound employed in [6] is less useful when k is large, and the method cannot extend directly to the case $k = O(n)$. The present paper deals with the case $k = O(n)$ for the model (1.1) and shows that the empirical spectral distribution (ESD) of $M_{n,k,m}$ with $\tau_\alpha \equiv 1$ converges to the Marčenko-Pastur law if and only if $k = o(n)$. Therefore, the matrix $M_{n,k,m}$ with $k = O(n)$ is a new example of *bad vectors*, for which the necessary and sufficient condition in [11] does not hold.

Another motivation for studying the model (1.1) is from quantum information theory. In [1], the model $M_{n,k,m}$ was introduced as a quantum analog of the classical probability problem of allocating r balls into s bins. The random vector Y_1 was interpreted as random product states in $(\mathbb{C}^n)^{\otimes k}$. When k is fixed and $m = cn^k$ for some $c > 0$, [1] established the convergence in expectation of the normalized trace of moments $n^{-k} \text{Tr} M_{n,k,m}^p$, which coincide with the corresponding moments of the Marčenko-Pastur law. In quantum physics and quantum information theory, it is natural to investigate the behavior of a large number of quantum states. The paper [12] characterizes the quantum entanglement of structured random states and studies the spectral density of the reduced state when the number of the quantum states k is large. Likewise, [4] studies the asymptotic behavior of the average entropy of entanglement for elements of an ensemble of random states associated with a graph when the dimension of the quantum subsystem is large.

This paper considers the scenario where k grows to infinity quickly. Namely, we assume that there exist constants $c, d \in (0, \infty)$, such that

$$\frac{k}{n} \rightarrow d, \quad \frac{m}{n^k} \rightarrow c. \quad (1.2)$$

Let us add one more word of motivation for the choice of regime $k = O(n)$. This is a natural threshold in our quest for limiting theorems, and from the point of view of probability theory, our result extends existing results. However, the need for large k is real in Quantum Information Theory. Actually, a typical real-world scenario would be n fixed (the dimension of the system) and k would go to infinity – the system can be used many times, i.e., one works on regularized quantities. In practice, we do not know how to fix n and send k to infinity, so we take this scenario as our next model.

Spectral distribution of sample covariance matrices from large tensor products

Our approach is based on the method of moments. Under appropriate moment conditions on the base variable ξ_1 and the sequence of coefficients $\{\tau_\alpha\}$, we derive the limits for the spectral moments of $M_{n,k,m}$ under the limiting scheme (1.2). A striking fact from our result is that contrary to [6], the limiting spectral moments found here involve the 4th moment of the base variable ξ_1 .

2 Almost sure convergence of spectral moments

We use $[s]$ to denote the set of integers from 1 to a positive integer s . For $\alpha = (\alpha_1, \dots, \alpha_p)$ and each value t in α , we count its frequency by

$$\text{deg}_t(\alpha) = |\{j \in [p] : \alpha_j = t\}|. \quad (2.1)$$

The main result of the paper is the following theorem.

Theorem 2.1. *Assume $d > 0$ and the following moment conditions hold.*

1. For all $p \in \mathbb{N}$, the p -th moment $m_p = E[|\xi_1|^p] < \infty$.
2. For all $q \in \mathbb{N}$,

$$\frac{1}{m} \sum_{j=1}^m \tau_j^q \rightarrow m_q^{(\tau)}, \quad m \rightarrow \infty. \quad (2.2)$$

We have

(i)

$$\lim_{n \rightarrow \infty} \frac{1}{n^k} \mathbb{E}[\text{Tr} M_{n,k,m}^p] = \sum_{s=1}^p c^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\prod_{t=1}^s m_{\text{deg}_t(\alpha)}^{(\tau)} \right) \exp \left(d(m_4 - 1) \sum_{t=1}^s \binom{\text{deg}_t(\alpha)}{2} \right). \quad (2.3)$$

Here $\mathcal{C}_{s,p}^{(1)}$ is a special class of graphs that will be defined in the course of the proof (see Lemma 3.1).

(ii) For all fixed $p \in \mathbb{N}_+$, if $k \geq 2$, we have

$$\sum_{n=1}^{\infty} \text{Var} \left(\frac{1}{n^k} \text{Tr} M_{n,k,m}^p \right) < \infty.$$

In particular, (i) and (ii) imply that $n^{-k} \text{Tr} M_{n,k,m}^p$ converge almost surely to the limit given in the r.h.s. of (2.3).

The proof of the theorem is given in Section 3.

Remark 2.2. Here, we require the random variable ξ_1 to have finite moments of any order. For most matrix models, one can remove such moment conditions by a standard truncation argument. However, the centralization step in the truncation argument fails for our matrix model. More precisely, let $\hat{y}_\alpha^{(l)}$ be the truncated vector for all $1 \leq \alpha \leq m$ and $1 \leq l \leq k$, then we have the following identity

$$\begin{aligned} & \hat{y}_\alpha^{(1)} \otimes \dots \otimes \hat{y}_\alpha^{(k)} - (\hat{y}_\alpha^{(1)} - \mathbb{E}[\hat{y}_\alpha^{(1)}]) \otimes \dots \otimes (\hat{y}_\alpha^{(k)} - \mathbb{E}[\hat{y}_\alpha^{(k)}]) \\ &= \sum_{l=1}^k \hat{y}_\alpha^{(1)} \otimes \dots \otimes \hat{y}_\alpha^{(l-1)} \otimes \mathbb{E}[\hat{y}_\alpha^{(l)}] \otimes (\hat{y}_\alpha^{(l+1)} - \mathbb{E}[\hat{y}_\alpha^{(l+1)}]) \otimes \dots \otimes (\hat{y}_\alpha^{(k)} - \mathbb{E}[\hat{y}_\alpha^{(k)}]). \end{aligned} \quad (2.4)$$

Spectral distribution of sample covariance matrices from large tensor products

Note that for $1 \leq l \leq k$, the $n^k \times m$ matrix, whose α -th column is

$$\hat{\mathbf{y}}_\alpha^{(1)} \otimes \cdots \otimes \hat{\mathbf{y}}_\alpha^{(l-1)} \otimes \mathbb{E}[\hat{\mathbf{y}}_\alpha^{(l)}] \otimes (\hat{\mathbf{y}}_\alpha^{(l+1)} - \mathbb{E}[\hat{\mathbf{y}}_\alpha^{(l+1)}]) \otimes \cdots \otimes (\hat{\mathbf{y}}_\alpha^{(k)} - \mathbb{E}[\hat{\mathbf{y}}_\alpha^{(k)}]),$$

has rank at most n^{k-1} . Thus, the $n^k \times m$ matrix, whose α -th column is (2.4), has rank at most kn^{k-1} . Hence, by [2, Theorem A.44], the sup norm of the difference of the cumulative distribution functions in the centralization step does not exceed k/n , which is not negligible when $d > 0$.

Let $\theta = e^{d(m_4-1)}$. The limit moments in (2.3), say γ_p , are polynomial functions of θ . The first four moments are

$$\begin{aligned} \gamma_1 &= 1, \\ \gamma_2 &= c\theta + c^2, \\ \gamma_3 &= c\theta^3 + 3c^2\theta + c^3, \\ \gamma_4 &= c\theta^6 + 4c^2\theta^3 + 2c^2\theta^2 + 6c^3\theta + c^4. \end{aligned}$$

However, for higher exponent p , some computing code is needed to find an explicit expression for γ_p .

Remark 2.3. The limiting moment sequence (γ_p) grows to infinity extremely fast with p . To see this, let $\tau_\alpha \equiv 1$. Then γ_p is lower bounded by the first term ($s = 1$) in (2.3), namely $c\theta^{p(p-1)/2}$. Thus the Carleman's condition, that is, $\sum_{p=1}^\infty \gamma_{2p}^{-1/(2p)} = \infty$, is not satisfied (see also [5]). In particular, it is not clear whether the moment sequence (γ_p) uniquely determines a limiting distribution. However, by the convergence of the moments, we know that the sequence of eigenvalue distributions is tight (almost surely).

As a byproduct of our moment method, we give an alternative method for deriving a limiting spectral distribution in the case of $d = 0$, a result already given in [6] using the method of Stieltjes transform.

Proposition 2.4. Assume $d = 0$. We have

(i)

$$\lim_{n \rightarrow \infty} \frac{1}{n^k} \mathbb{E}[\text{Tr} M_{n,k,m}^p] = \sum_{s=1}^p c^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\prod_{t=1}^s m_{\deg_t(\alpha)}^{(\tau)} \right). \quad (2.5)$$

Here $\mathcal{C}_{s,p}^{(1)}$ is a special class of graphs defined later in Lemma 3.1.

(ii) For all fixed $p \in \mathbb{N}_+$, if $k \geq 2$, we have

$$\sum_{n=1}^\infty \text{Var} \left(\frac{1}{n^k} \text{Tr} M_{n,k,m}^p \right) < \infty.$$

In particular, (i) and (ii) imply that $n^{-k} \text{Tr} M_{n,k,m}^p$ converge almost surely to the limit given in the r.h.s. of (2.5).

The proof of the proposition is given in Section 4.

Finally, comparing the two cases $d > 0$ and $d = 0$, it is worth noticing that the fourth moment m_4 contributes to the limiting spectral moments only in the case of $d > 0$ (Theorem 2.1).

3 Proof of Theorem 2.1

We first introduce some preliminaries on graph combinatorics. We call $\alpha = (\alpha_1, \dots, \alpha_p) \in [m]^p$ a sequence of length p with vertices α_j for $1 \leq j \leq p$. We denote by $|\alpha|$ the

Spectral distribution of sample covariance matrices from large tensor products

number of distinct elements in α . If $s = |\alpha|$, then we call α an s -sequence. Let $\mathcal{J}_{s,p}(m)$ be the set of all s -sequences $\alpha \in [m]^p$. Then

$$[m]^p = \bigcup_{s=1}^p \mathcal{J}_{s,p}(m). \tag{3.1}$$

Two sequences are equivalent if one becomes the other by a suitable permutation on $[m]$. The sequence α is canonical if $\alpha_1 = 1$ and $\alpha_u \leq \max\{\alpha_1, \dots, \alpha_{u-1}\} + 1$ for $u \geq 2$. We denote by $\mathcal{C}_{s,p}$ the set of all canonical s -sequences of length p . From the definition above, one can see that the set of distinct vertices of a canonical s -sequence is $[s]$. Denote by $\mathcal{I}_{s,m}$ the set of injective maps from $[s]$ to $[m]$. For a canonical s -sequence α and a map $\varphi \in \mathcal{I}_{s,m}$, we call $\varphi(\alpha)$ the s -sequence $(\varphi(\alpha_1), \dots, \varphi(\alpha_p))$. For each canonical s -sequence, its image under the maps in $\mathcal{I}_{s,m}$ gives all its equivalent sequences, and hence its equivalent class of sequences in $[m]^p$ has exactly $m(m-1) \cdots (m-s+1)$ distinct elements.

Let $\alpha = (\alpha_1, \dots, \alpha_p) \in [m]^p$ be a fixed canonical s -sequence. For $i_1 = (i_1^{(1)}, \dots, i_1^{(p)}) \in [n]^p$, draw two parallel lines referred as the α -line and the i -line, respectively. Plot $i_1^{(1)}, \dots, i_1^{(p)}$ on the i -line and $\alpha_1, \dots, \alpha_p$ on the α -line. Draw p down edges from α_u to $i_1^{(u)}$ and p up edges from $i_1^{(u)}$ to α_{u+1} for $1 \leq u \leq p$. Here, we use the convention that $\alpha_1 = \alpha_{p+1}$. We call a down edge from α_u to $i_1^{(u)}$ a down innovation if $i_1^{(u)}$ is different from $i_1^{(1)}, \dots, i_1^{(u-1)}$. We also call an up edge from $i_1^{(u)}$ to α_{u+1} an up innovation if α_{u+1} is different from $\alpha_1, \dots, \alpha_u$. We denote the graph by $g(i_1, \alpha)$ and call such graph a $\Delta(p; \alpha)$ -graph. Two graphs $g(i_1, \alpha)$ and $g(i'_1, \alpha)$ are called equivalent if the two sequences i_1 and i'_1 are equivalent, and we write $g(i_1, \alpha) \sim g(i'_1, \alpha)$ for this equivalence. For each equivalent class, we choose the canonical graph such that $i_1 = (i_1^{(1)}, \dots, i_1^{(p)}) \in [n]^p$ is a canonical r -sequence for some $r \in \mathbb{N}_+$. A canonical $\Delta(p; \alpha)$ -graph is denoted by $\Delta(p, r, s; \alpha)$ if it has r noncoincident i -vertices and s noncoincident α -vertices. We classify $\Delta(p, r, s; \alpha)$ -graphs into the following five categories.

- $\Delta_1(p, s; \alpha)$. $\Delta(p; \alpha)$ -graphs in which each down edge must coincide with one and only one up edge. If we glue the coincident edges, the resulting graph is a tree of p edges and $p + 1$ vertices, which implies $r + s = p + 1$.
- $\Delta_2(p, r, s; \alpha)$. $\Delta(p; \alpha)$ -graphs that contain at least one single edge.
- $\Delta_3(p, s; \alpha)$. $\Delta(p; \alpha)$ -graphs such that the number of the edges between two vertices is 0 or 2. If we glue the coincident edges, the resulting graph is a connected graph with exactly one cycle. The graph has p edges and p vertices, which implies $r + s = p$.
- $\Delta_4(p, s; \alpha)$. $\Delta(p; \alpha)$ -graphs having two up edges and two down edges with the same endpoints, and all other down edges coincide with one and only one up edge. If we glue the coincident edges, the resulting graph is a tree of $p - 1$ edges and p vertices, implying $r + s = p$.
- $\Delta_5(p, r, s; \alpha)$. $\Delta(p; \alpha)$ -graphs that do not belong to the categories above. In this case, the graph has at most $p - 1$ vertices, since the in-degree equals to out-degree and at least 2 for all vertices. Thus, $r + s \leq p - 1$.

Next, we determine the number of sequence $i_1 \in \mathcal{C}_{p+1-s,p}$ such that $g(i_1, \alpha) \in \Delta_1(p, s; \alpha)$ for a given sequence $\alpha \in \mathcal{C}_{s,p}$. We have the following lemma, which is motivated by [2, Lemma 3.4].

Lemma 3.1. *For any sequence $\alpha \in \mathcal{C}_{s,p}$, there is at most one sequence $i_1 \in \mathcal{C}_{p+1-s,p}$ such that $g(i_1, \alpha) \in \Delta_1(p, s; \alpha)$. We denote by $\mathcal{C}_{s,p}^{(1)}$ the set of such canonical sequences α .*

Spectral distribution of sample covariance matrices from large tensor products

Then the number of the elements in $C_{s,p}^{(1)}$ is

$$\frac{1}{p} \binom{p}{s-1} \binom{p}{s}.$$

Proof. Let

$$\Delta_1(p, s) = \bigcup_{\alpha \in C_{s,p}} \Delta_1(p, s; \alpha).$$

We define a pair of characteristic sequences $\{u_1, \dots, u_p\}$ and $\{d_1, \dots, d_p\}$ by

$$u_l = \begin{cases} 1, & \alpha_{l+1} = \max\{\alpha_1, \dots, \alpha_l\} + 1, \\ 0, & \text{otherwise,} \end{cases} \quad (3.2)$$

and

$$d_l = \begin{cases} -1, & \alpha_l \notin \{1, \alpha_{l+1}, \dots, \alpha_p\}, \\ 0, & \text{otherwise.} \end{cases} \quad (3.3)$$

By definition, we always have $u_p = 0$ and $d_1 = 0$ because $\alpha_1 = 1$.

For a graph in $\Delta_1(p, s)$, there are exactly $s - 1$ up innovations and hence there are $s - 1$ u -variables equal to 1 and $s - 1$ d -variables equal to -1 . From its definition, one sees that $d_l = -1$ means that after plotting the l -th down edge $(\alpha_l, i_1^{(l)})$, the future path will never revisit the vertices α_l , which means that the edge $(\alpha_l, i_1^{(l)})$ must coincide with the up innovation leading to the vertex α_l . Since there are $r = p + 1 - s$ down innovations to lead out the r i_1 -vertices, $d_l = 0$ implies that the edge $(\alpha_l, i_1^{(l)})$ must be a down innovation. Therefore, $d_l = -1$ must follow a $u_j = 1$ for some $j < l$, which leads to the restriction of the pair of characteristic sequences to satisfy

$$u_1 + \dots + u_{l-1} + d_2 + \dots + d_l \geq 0, \quad 2 \leq l \leq p. \quad (3.4)$$

Next, we show that each pair of characteristic sequences satisfying (3.4) defines a graph in $\Delta_1(p, s)$ uniquely.

Firstly, we have $i_1^{(1)} = 1$. Besides, $\alpha_2 = 1$ if $u_1 = 0$ and $\alpha_2 = 2$ if $u_1 = 1$. We use induction to determine the unique graph in $\Delta_1(p, s)$. Suppose that the first l pairs of the down and up edges are uniquely determined by the two sequences $\{u_1, \dots, u_l\}$ and $\{d_1, \dots, d_l\}$, and the subgraph of the first l pairs of down and up edges satisfies the following properties:

- (a1) The subgraph is connected, and the unidirectional noncoincident edges form a tree;
- (a2) If $\alpha_{l+1} = 1$, then each down edge coincides with an up edge, which means that the subgraph has not any single innovation;
- (a3) If $\alpha_{l+1} \neq 1$, then from the vertex α_1 to α_{l+1} , there is only one path (a chain without cycles) of down-up-down-up single innovations and all other down edges coincide with an up edge.

We consider the following four cases to determine the $(l + 1)$ -th pair of down and up edges.

Case 1. $d_{l+1} = 0$ and $u_{l+1} = 1$. Then both edges of the $(l + 1)$ -th pair are innovations, which implies that $i_1^{(l+1)} = |\{j \leq l + 1 : d_j = 0\}|$ and $\alpha_{l+2} = 1 + |\{j \leq l + 1 : u_j = 1\}|$. After adding the two edges, the subgraph with the first $l + 1$ pairs of down and up edges satisfies the properties (a1)–(a3).

Spectral distribution of sample covariance matrices from large tensor products

Case 2. $d_{l+1} = 0$ and $u_{l+1} = 0$. Then the down edge $(\alpha_{l+1}, i_1^{(l+1)})$ is an innovation so $i_1^{(l+1)} = |\{j \leq l + 1 : d_j = 0\}|$. The up edge $(i_1^{(l+1)}, \alpha_{l+2})$ is not an innovation so α_{l+2} coincides with some vertex α_j for $1 \leq j \leq l + 1$. If $\alpha_j \neq \alpha_{l+1}$, then there is a path $i_1^{(l)} \rightarrow \alpha_{l+1} \rightarrow i_1^{(l+1)} \rightarrow \alpha_j$. Also note that there should be a path connecting α_j and $i_1^{(l)}$ in the subgraph of first l pairs of down and up edges. The two paths with undirected noncoincident edges will lead to a circle, which is a contradiction. Hence, $\alpha_{l+2} = \alpha_j = \alpha_{l+1}$ and the $(l + 1)$ -th up edge coincide with the $(l + 1)$ -th down edge. After adding the two edges, the subgraph with the first $l + 1$ pairs of down and up edges satisfies the properties (a1)–(a3).

Case 3. $d_{l+1} = -1$ and $u_{l+1} = 1$. By (3.4), we have

$$u_1 + \dots + u_l + d_2 + \dots + d_l \geq 1.$$

Hence, there is at least one vertex in $\{\alpha_2, \dots, \alpha_{l+1}\}$ which is not α_1 , such that the vertex will be visited in the last $p - (l + 1)$ pairs of down and up edges. Thus, by property (a1) and (a2), we have $\alpha_{l+1} \neq \alpha_1$. Hence, there must be a single up innovation leading to the vertex α_{l+1} . We denote the up innovation by $(i_1^{(j)}, \alpha_{l+1})$ for some $1 \leq j \leq l$. Hence, we choose $i_1^{(l+1)} = i_1^{(j)}$, which means that the down edge starting from α_{l+1} coincides with the up innovation leading to α_{l+1} . Besides, the $(l + 1)$ -th up edge is an innovation, which starts from $i_1^{(l+1)}$ and ends in $\alpha_{l+2} = 1 + |\{j \leq l + 1 : u_j = 1\}|$. After adding the two edges, the subgraph with the first $l + 1$ pairs of down and up edges satisfies the properties (a1)–(a3).

Case 4. $d_{l+1} = -1$ and $u_{l+1} = 0$. As discussed in Case 3, the $(l + 1)$ -th down edge coincides with the only up innovation ended at α_{l+1} . Before this up innovation, there must be a single down innovation by property (a3). Then the up edge can be drawn to coincide with this down innovation. If the path of single innovations of the subgraph with the first l pairs of down and up edges has only one pair of down-up innovations, then $\alpha_{l+2} = 1$, and hence the subgraph with the first $l + 1$ pairs of down and up edges has no single innovations, which implies that the properties (a1) and (a2) hold. Otherwise, $\alpha_{l+2} \neq 1$ and the subgraph with first $l + 1$ pairs of down and up edges satisfies properties (a1) and (a3).

By induction, it is shown that two characteristic sequences $\{u_1, \dots, u_p\}$ and $\{d_1, \dots, d_p\}$ satisfying (3.4) determine a graph in $\Delta_1(p, s)$ uniquely.

For a given sequence $\alpha = (\alpha_1, \dots, \alpha_p) \in \mathcal{C}_{s,p}$, the two characteristic sequences are uniquely determined by (3.2) and (3.3). This shows that there is at most one sequence $i_1 \in \mathcal{C}_{p+1-s,p}$ such that $g(i_1, \alpha) \in \Delta_1(p, s; \alpha)$. More precisely, there is exactly one sequence $i_1 \in \mathcal{C}_{p+1-s,p}$ such that $g(i_1, \alpha) \in \Delta_1(p, s; \alpha)$ if and only if the pair of two characteristic sequences satisfies (3.4).

Moreover, by [2, Lemma 3.5], the number of the elements in $\mathcal{C}_{s,p}^{(1)}$ equals to the number of graphs in

$$\bigcup_{\alpha \in \mathcal{C}_{s,p}} \Delta_1(p, s; \alpha),$$

which is

$$\frac{1}{s} \binom{p}{s-1} \binom{p-1}{s-1} = \frac{1}{p} \binom{p}{s-1} \binom{p}{s} \square$$

Remark 3.2. There exists $\alpha \in \mathcal{C}_{s,p} \setminus \mathcal{C}_{s,p}^{(1)}$, such that the u -sequence and the d -sequence defined by (3.2) and (3.3) respectively satisfy the restriction (3.4). Consider the canonical

Spectral distribution of sample covariance matrices from large tensor products

2-sequence $\alpha = (1, 2, 1, 2)$, the corresponding u -sequence is $\{1, 0, 0, 0\}$ and the corresponding d -sequence is $\{0, 0, 0, -1\}$, which satisfy (3.4). However, the graph of α belongs to $\Delta_3(4, 2; \alpha)$, which means that $\alpha \in \mathcal{C}_{2,4} \setminus \mathcal{C}_{2,4}^{(1)}$. Indeed, the α -sequence of the graph in $\Delta_1(4, 2; \alpha)$ corresponding to the pair of characteristic sequences is $(1, 2, 2, 2)$.

3.1 Proof of (i)

We will use $\alpha, \beta \in [m]$ to index the columns of Y . We will also use a multiple index \mathbf{i} of the form $\mathbf{i} = (i_1, \dots, i_k)$ for the rows of Y . Then \mathbf{i} should be in $[n^k]$. For any $p \in \mathbb{N}$, we compute the moment

$$\frac{1}{n^k} \mathbb{E}[\text{Tr} M_{n,k,m}^p].$$

By convention, $\alpha_{p+1} = \alpha_1$. We have

$$\begin{aligned} \frac{1}{n^k} \mathbb{E}[\text{Tr} M_{n,k,m}^p] &= \frac{1}{n^k} \sum_{\alpha_1, \dots, \alpha_p=1}^m \left(\prod_{t=1}^p \tau_{\alpha_t} \right) \mathbb{E}[\text{Tr}(Y_{\alpha_{p+1}} Y_{\alpha_1}^* \cdots Y_{\alpha_p} Y_{\alpha_p}^*)] \\ &= \frac{1}{n^k} \sum_{\alpha_1, \dots, \alpha_p=1}^m \left(\prod_{t=1}^p \tau_{\alpha_t} \right) \mathbb{E} \left[\sum_{\mathbf{i}^{(1)}, \dots, \mathbf{i}^{(p)}=1}^{n^k} \prod_{t=1}^p \overline{Y_{\mathbf{i}^{(t)} \alpha_t}} Y_{\mathbf{i}^{(t)} \alpha_{t+1}} \right] \\ &= \frac{1}{n^k} \sum_{\alpha_1, \dots, \alpha_p=1}^m \left(\prod_{t=1}^p \tau_{\alpha_t} \right) \mathbb{E} \left[\sum_{\mathbf{i}^{(1)}, \dots, \mathbf{i}^{(p)}=1}^{n^k} \prod_{t=1}^p \prod_{l=1}^k \left(\overline{(\mathbf{y}_{\alpha_t}^{(l)})_{i_l^{(t)}}} (\mathbf{y}_{\alpha_{t+1}}^{(l)})_{i_l^{(t)}} \right) \right] \\ &= \frac{1}{n^k} \sum_{\alpha_1, \dots, \alpha_p=1}^m \left(\prod_{t=1}^p \tau_{\alpha_t} \right) \mathbb{E} \left[\prod_{l=1}^k \sum_{i_l^{(1)}, \dots, i_l^{(p)}=1}^n \prod_{t=1}^p \left(\overline{(\mathbf{y}_{\alpha_t}^{(l)})_{i_l^{(t)}}} (\mathbf{y}_{\alpha_{t+1}}^{(l)})_{i_l^{(t)}} \right) \right] \\ &= \frac{1}{n^k} \sum_{\alpha_1, \dots, \alpha_p=1}^m \left(\prod_{t=1}^p \tau_{\alpha_t} \right) \left(\mathbb{E} \left[\sum_{i_1^{(1)}, \dots, i_1^{(p)}=1}^n \prod_{t=1}^p \left(\overline{(\mathbf{y}_{\alpha_t}^{(1)})_{i_1^{(t)}}} (\mathbf{y}_{\alpha_{t+1}}^{(1)})_{i_1^{(t)}} \right) \right] \right)^k. \end{aligned} \tag{3.5}$$

We used the i.i.d. setting in the last equality.

For two sequences $\alpha = (\alpha_1, \dots, \alpha_p) \in [m]^p$ and $i_1 = (i_1^{(1)}, \dots, i_1^{(p)}) \in [n]^p$, let

$$E(i_1, \alpha) = \mathbb{E} \left[\prod_{t=1}^p \left(\overline{(\mathbf{y}_{\alpha_t}^{(1)})_{i_1^{(t)}}} (\mathbf{y}_{\alpha_{t+1}}^{(1)})_{i_1^{(t)}} \right) \right]. \tag{3.6}$$

Observe that $E(i_1, \alpha) = E(i_1', \alpha')$ if the two sequences i_1 and α are equivalent to i_1' and α' , respectively. By (3.5) and (3.1), we have

$$\begin{aligned} \frac{1}{n^k} \mathbb{E}[\text{Tr} M_{n,k,m}^p] &= \frac{1}{n^k} \sum_{s=1}^p \sum_{\alpha \in \mathcal{J}_{s,p}(m)} \left(\prod_{t=1}^p \tau_{\alpha_t} \right) \left(\sum_{r=1}^p \sum_{i_1 \in \mathcal{J}_{r,p}(n)} E(i_1, \alpha) \right)^k \\ &= \frac{1}{n^k} \sum_{s=1}^p \sum_{\alpha \in \mathcal{C}_{s,p}} \left(\sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^p \tau_{\varphi(\alpha_t)} \right) \left(\sum_{r=1}^p n \cdots (n-r+1) \sum_{i_1 \in \mathcal{C}_{r,p}} E(i_1, \alpha) \right)^k. \end{aligned} \tag{3.7}$$

We first compute the sum on i_1 in (3.7). For any sequences $\alpha \in \mathcal{C}_{s,p}$, we can split the sum according to the category of the graph $g(i_1, \alpha)$ as follows:

$$\sum_{i_1 \in \mathcal{C}_{r,p}} E(i_1, \alpha) = \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_1(p, s; \alpha)}} E(i_1, \alpha) + \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_3(p, s; \alpha)}} E(i_1, \alpha)$$

Spectral distribution of sample covariance matrices from large tensor products

$$+ \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_4(p, s; \alpha)}} E(i_1, \alpha) + \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_5(p, r, s; \alpha)}} E(i_1, \alpha) \quad (3.8)$$

We deal with the four terms on the right hand side, respectively. For $i_1 \in \mathcal{C}_{r,p}$ with $g(i_1, \alpha) \in \Delta_1(p, s; \alpha)$, by the definition of $\Delta_1(p, s; \alpha)$, we have $E(i_1, \alpha) = n^{-p}$ and $p = r + s - 1$. Hence,

$$\sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_1(p, s; \alpha)}} E(i_1, \alpha) = n^{1-r-s} \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_1(p, s; \alpha)}} 1. \quad (3.9)$$

For $i_1 \in \mathcal{C}_{r,p}$ with $g(i_1, \alpha) \in \Delta_3(p, s; \alpha)$, by the definition of $\Delta_3(p, s; \alpha)$, we have $E(i_1, \alpha) = n^{-p}$ and $p = r + s$. Hence,

$$\sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_3(p, s; \alpha)}} E(i_1, \alpha) = n^{-r-s} \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_3(p, s; \alpha)}} 1. \quad (3.10)$$

For $i_1 \in \mathcal{C}_{r,p}$ with $g(i_1, \alpha) \in \Delta_4(p, s; \alpha)$, by the definition of $\Delta_4(p, s; \alpha)$, we have $E(i_1, \alpha) = n^{-p}m_4$ and $p = r + s$. Hence,

$$\sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_4(p, s; \alpha)}} E(i_1, \alpha) = n^{-r-s}m_4 \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_4(p, s; \alpha)}} 1. \quad (3.11)$$

For $i_1 \in \mathcal{C}_{r,p}$ with $g(i_1, \alpha) \in \Delta_5(p, r, s; \alpha)$, by the definition of $\Delta_5(p, r, s; \alpha)$, we have $E(i_1, \alpha) = O(n^{-p})$ and $p \geq r + s + 1$. Hence,

$$\sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_5(p, r, s; \alpha)}} E(i_1, \alpha) = O(n^{-p}) \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_5(p, r, s; \alpha)}} 1. \quad (3.12)$$

Substituting (3.9), (3.10), (3.11) and (3.12) to (3.8), and noting that

$$n \cdots (n - r + 1) = n^r \left(1 - \frac{r(r-1)}{2n} + O(n^{-2}) \right),$$

we have

$$\begin{aligned} & \sum_{r=1}^p n \cdots (n - r + 1) \sum_{i_1 \in \mathcal{C}_{r,p}} E(i_1, \alpha) \\ &= n^{1-s} \left(1 - \frac{(p+1-s)(p-s)}{2n} + O(n^{-2}) \right) \sum_{\substack{i_1 \in \mathcal{C}_{p+1-s,p} \\ g(i_1, \alpha) \in \Delta_1(p, s; \alpha)}} 1 \\ & \quad + n^{-s} \left(1 - \frac{(p-s)(p-s-1)}{2n} + O(n^{-2}) \right) \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_3(p, s; \alpha)}} 1 \\ & \quad + n^{-s} \left(1 - \frac{(p-s)(p-s-1)}{2n} + O(n^{-2}) \right) m_4 \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_4(p, s; \alpha)}} 1 \\ & \quad + \sum_{r=1}^{p-s-1} n^r \left(1 - \frac{r(r-1)}{2n} + O(n^{-2}) \right) \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_5(p, r, s; \alpha)}} O(n^{-p}) \\ &= n^{1-s} \left(1 - \frac{(p+1-s)(p-s)}{2n} \right) \sum_{\substack{i_1 \in \mathcal{C}_{p+1-s,p} \\ g(i_1, \alpha) \in \Delta_1(p, s; \alpha)}} 1 \end{aligned}$$

Spectral distribution of sample covariance matrices from large tensor products

$$+ n^{-s} \left(\sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_3(p, s; \alpha)}} 1 + m_4 \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_4(p, s; \alpha)}} 1 \right) + O(n^{-s-1}) \quad (3.13)$$

$$= \begin{cases} O(n^{1-s}), & \alpha \in \mathcal{C}_{s,p}^{(1)}, \\ O(n^{-s}), & \alpha \in \mathcal{C}_{s,p} \setminus \mathcal{C}_{s,p}^{(1)}, \end{cases} \quad (3.14)$$

where the last equality follows from Lemma 3.1. Hence, by (3.13), (3.14) and Lemma 3.1, (3.7) can be written as

$$\begin{aligned} & \frac{1}{n^k} \mathbb{E}[\text{Tr} M_{n,k,m}^p] \\ &= \frac{1}{n^k} \sum_{s=1}^p \sum_{\alpha \in \mathcal{C}_{s,p}} \left(\sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^p \tau_{\varphi(\alpha_t)} \right) \left(\sum_{r=1}^p n \cdots (n-r+1) \sum_{i_1 \in \mathcal{C}_{r,p}} E(i_1, \alpha) \right)^k \\ &= \sum_{s=1}^p \left(\frac{m}{n^k} \right)^s (1 + o(1)) \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\frac{1}{m^s} \sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^p \tau_{\varphi(\alpha_t)} \right) \\ & \quad \left(1 - \frac{(p+1-s)(p-s)}{2n} + \frac{1}{n} \left(\sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_3(p, s; \alpha)}} 1 + m_4 \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_4(p, s; \alpha)}} 1 \right) \right)^k. \end{aligned} \quad (3.15)$$

Recall the definition of $\text{deg}_t(\alpha)$ given in (2.1). Then the sequence α has exactly $\text{deg}_t(\alpha)$ vertices that equals to t . Thus, we have the following identity

$$\sum_{t=1}^s \text{deg}_t(\alpha) = p.$$

The following lemma computes the size of $\Delta_4(p, s; \alpha)$ for given $\alpha \in \mathcal{C}_{s,p}^{(1)}$.

Lemma 3.3. For fixed $\alpha = (\alpha_1, \dots, \alpha_p) \in \mathcal{C}_{s,p}^{(1)}$, the number of $i_1 \in \mathcal{C}_{r,p}$ such that $g(i_1, \alpha) \in \Delta_4(p, s; \alpha)$ is

$$\sum_{t=1}^s \binom{\text{deg}_t(\alpha)}{2}. \quad (3.16)$$

Proof. We sort the graphs in $\Delta_4(p, s; \alpha)$ into two classes. We will show that graphs in $\Delta_4(p, s; \alpha)$ can be transferred to graphs in $\Delta_1(p, s; \alpha)$ by splitting the vertex in the sequence i_1 which has multiple edges.

The first class consists of graphs whose first $l-1$ up edges are distinct, and the l -th down edge $(\alpha_l, i_1^{(l)})$ coincide with the j -th down edge for some $1 \leq j < l \leq p$. Hence, $\alpha_l = \alpha_j$ and $i_1^{(l)} = i_1^{(j)}$. Since the subgraph of the path from α_j to α_l does not have a cycle, one can find $1 \leq j' < l$, such that the j' -th up edge is from $i_1^{(l)}$ to α_l . Another up edge from $i_1^{(l)}$ to α_l belongs to the set of the last $p - (l-1)$ up edges. Hence, the edge $(\alpha_l, i_1^{(l)})$ is not a down innovation. Now we split the vertex $i_1^{(l)}$ into two vertices $i_1^{(l,1)}$ and $i_1^{(l,2)}$. The new vertex $i_1^{(l,1)}$ is still labeled by the value of $i_1^{(l)}$ and the first $l-1$ pairs of up and down edges which connect $i_1^{(l)}$ are plotted to connect the new vertex $i_1^{(l,1)}$. The l -th down edge connect $i_1^{(l)}$ is plotted to connect the new vertex $i_1^{(l,2)}$, which is a down innovation. The rest of the edges can be plotted such that the new graph belongs to $\Delta_1(p, s; \alpha)$. See Figure 1.

The second class consists of graphs whose first l down edges are distinct, and the l -th up edge $(i_1^{(l)}, \alpha_{l+1})$ coincides with the j -th up edge for some $1 \leq j < l \leq p-1$. This

Spectral distribution of sample covariance matrices from large tensor products

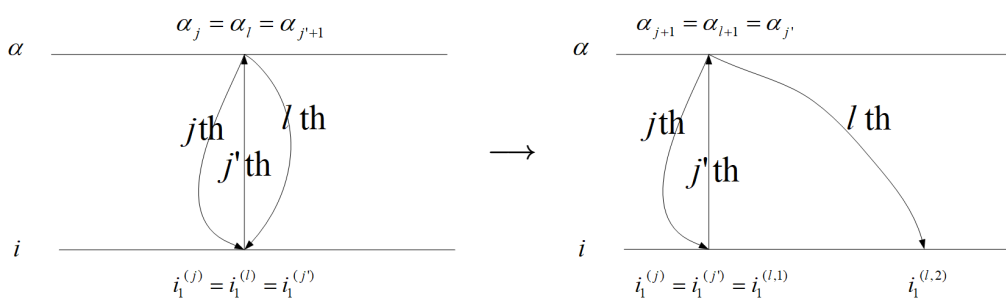


Figure 1: A graph in the first class of $\Delta_4(p, s; \alpha)$ (left column) and corresponding $\Delta_1(p, s; \alpha)$ graph (right column).

means that $i_1^{(l)} = i_1^{(j)}$ and $\alpha_{l+1} = \alpha_{j+1}$. Since the subgraph of the path from $i_1^{(j)}$ to $i_1^{(l)}$ does not have a cycle, one can find $j < j' \leq l$, such that the j' -th down edge is from α_{l+1} to $i_1^{(l)}$, which means that $\alpha_{j'} = \alpha_{l+1}$, $i_1^{(j')} = i_1^{(l)}$ and the j' -th down edge is not a down innovation. Another down edge from α_{l+1} to $i_1^{(l)}$ belongs to the set of the last $p - l$ down edges. Now we split the vertex $i_1^{(l)}$ into two vertices $i_1^{(l,1)}$ and $i_1^{(l,2)}$. The new vertex $i_1^{(l,1)}$ is still labeled by the value of $i_1^{(l)}$ and the first $j' - 1$ pairs of up and down edges which connect $i_1^{(l)}$ are plotted to connect the new vertex $i_1^{(l,1)}$. The j' -th down edge connect $i_1^{(j')} = i_1^{(l)}$ is plotted to connect the new vertex $i_1^{(l,2)}$, which is a down innovation. The l -th up edge starts from $i_1^{(l,2)}$. The rest of the edges can be plotted such that the new graph belongs to $\Delta_1(p, s; \alpha)$. See Figure 2.

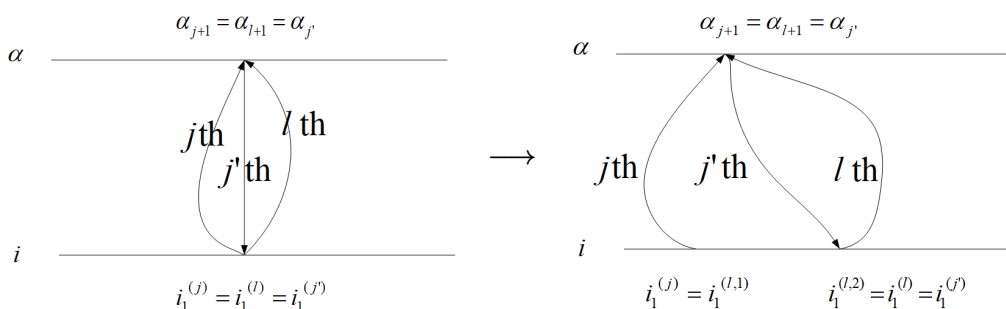


Figure 2: A graph in the second class of $\Delta_4(p, s; \alpha)$ (left column) and corresponding $\Delta_1(p, s; \alpha)$ graph (right column).

Therefore, for any graphs in $\Delta_4(p, s; \alpha)$, one can split the vertex in the sequence i_1 associated to the multiple edges to get a graph in $\Delta_1(p, s; \alpha)$. Equivalently, any graphs in $\Delta_4(p, s; \alpha)$ can be obtained from graphs in $\Delta_1(p, s; \alpha)$ by gluing two vertices of the sequence i_1 that have a same neighborhood. Moreover, gluing different pairs of vertices of the sequence i_1 leads to different graphs in $\Delta_4(p, s; \alpha)$.

For fixed $\alpha \in \mathcal{C}_{s,p}^{(1)}$, by Lemma 3.1, there exists a unique sequence i_1 , such that the graph $g(i_1, \alpha) \in \Delta_1(p, s; \alpha)$. Note that for a vertex t on the α -line, the number of its neighborhoods on the i -line is $\deg_t(\alpha)$. Hence, the number of choices of gluing two vertices of the sequence i_1 with the same neighborhood t on the α -line is

$$\binom{\deg_t(\alpha)}{2}.$$

Thus, for fixed $\alpha \in \mathcal{C}_{s,p}^{(1)}$, the number of $i_1 \in \mathcal{C}_{r,p}$ such that $g(i_1, \alpha) \in \Delta_4(p, s; \alpha)$ is given

Spectral distribution of sample covariance matrices from large tensor products

by (3.16). □

The following lemma is a property of the sequences in $\mathcal{C}_{s,p}^{(1)}$.

Lemma 3.4. For $\alpha \in \mathcal{C}_{s,p}$, if there exist different integers $t_1, t_2 \in [s]$, such that $\alpha_{j_1} = \alpha_{j'_1} = t_1$ and $\alpha_{j_2} = \alpha_{j'_2} = t_2$ for some $1 \leq j_1 < j_2 < j'_1 < j'_2 \leq p$, then $\alpha \notin \mathcal{C}_{s,p}^{(1)}$.

Proof. We prove by contradiction. Suppose $\alpha \in \mathcal{C}_{s,p}^{(1)}$. Then there exists a $w \in \mathcal{C}_{p+1-s,p}$, such that $g(w, \alpha) \in \Delta_1(p, s; \alpha)$. Note that the path from α_{j_1} to α_{j_2} and the path from $\alpha_{j'_1}$ to $\alpha_{j'_2}$ are two paths from the vertex on the α -line with label t_1 to the vertex on the α -line with label t_2 . The two paths will lead to multiple directed edges or undirected cycle, which contradicts the definition of $\Delta_1(p, s; \alpha)$. □

The following corollary is a direct consequence of Lemma 3.4.

Corollary 3.5. For $\alpha \in \mathcal{C}_{s,p}^{(1)}$, graphs in $\Delta_3(p, s; \alpha)$ are the $\Delta(p; \alpha)$ -graphs in which each down edge must coincide with one and only one up edge. If we glue the coincident edges, the resulting graph is a connected graph with exactly one cycle.

The following lemma computes the size of $\Delta_3(p, s; \alpha)$ for a given $\alpha \in \mathcal{C}_{s,p}^{(1)}$.

Lemma 3.6. For $\alpha = (\alpha_1, \dots, \alpha_p) \in \mathcal{C}_{s,p}^{(1)}$, the number of $i_1 \in \mathcal{C}_{r,p}$ such that $g(i_1, \alpha) \in \Delta_3(p, s; \alpha)$ is

$$\binom{p+1-s}{2} - \sum_{t=1}^s \binom{\deg_t(\alpha)}{2}. \tag{3.17}$$

Proof. Recall that in a graph from $\Delta_3(p, s; \alpha)$, each down edge coincides with exactly one up edge, and the graph is connected with exactly one cycle after gluing coincident edges.

We also sort the graphs in $\Delta_3(p, s; \alpha)$ into two classes. We will show that graphs in $\Delta_3(p, s; \alpha)$ can be transferred to graphs in $\Delta_1(p, s; \alpha)$ by splitting an appropriate vertex in the sequence i_1 .

The first class consists of graphs whose subgraph with first $l-1$ up and down edges has no cycle when orientation is removed and multiple edges are glued, but has a cycle when adding the l -th down edge. Hence, we have $3 \leq l \leq p$, and both α_l and $i_1^{(l)}$ are not new vertices. Now we split the vertex $i_1^{(l)}$ into two vertices $i_1^{(l,1)}$ and $i_1^{(l,2)}$. The new vertex $i_1^{(l,1)}$ is still labeled by the value of $i_1^{(l)}$, and the first $l-1$ pairs of up and down edges which connect $i_1^{(l)}$ are plotted to connect the new vertex $i_1^{(l,1)}$. The l -th down edge $(\alpha_l, i_1^{(l)})$ is replaced by the new down edge $(\alpha_l, i_1^{(l,2)})$, which is a down innovation. The l -th up edge starts from $i_1^{(l,2)}$. The rest of the edges can be plotted such that the new graph belongs to $\Delta_1(p, s; \alpha)$, see Figure 3.

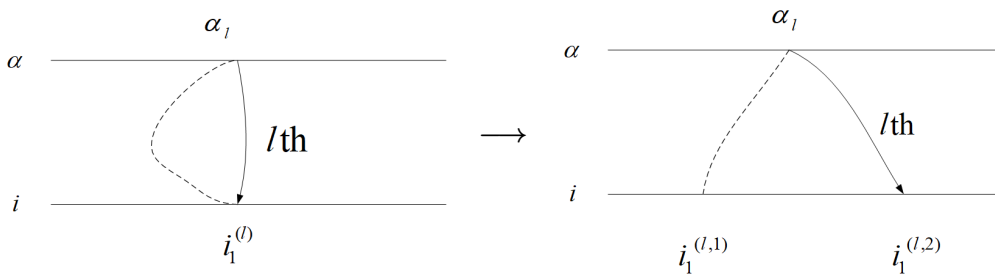


Figure 3: A graph in the first class of $\Delta_3(p, s; \alpha)$ (left column) and corresponding $\Delta_1(p, s; \alpha)$ graph (right column). The dashed line describes the existence of a cycle.

Spectral distribution of sample covariance matrices from large tensor products

The second class consists of graphs whose subgraph with first $l - 1$ up edges and first l down edges has no cycle when orientation is removed and multiple edges are glued, but has a cycle when adding the l -th up edge. Thus, there exists a $1 \leq j < l$, such that $\alpha_{l+1} = \alpha_j$ and the j -th down edge belongs to the cycle. Since the subgraph with first $l - 1$ up edges and first l down edges has no cycle, we have $\alpha_u \neq \alpha_j$ for all $j < u < l + 1$. By Lemma 3.4, the first $j - 1$ pairs as well as the last $p - l$ pairs of down and up edges never visit the vertex α_u for all $j < u < l + 1$. Hence, along the path from $i_1^{(j)}$ to $i_1^{(l)}$, all down edges are paired with exactly one up edge. Note that along the path from α_j to $i_1^{(l)}$, there must be a single edge, which can only be $(\alpha_j, i_1^{(j)})$. If $i_1^{(j)} \neq i_1^{(l)}$, then in the subgraph of path from α_j to α_{l+1} , the degree of the vertices $i_1^{(j)}$ and $i_1^{(l)}$ are both odd number. This is impossible. However, $i_1^{(j)} = i_1^{(l)}$ implies that the l -th up edge $(i_1^{(l)}, \alpha_{l+1})$ coincides with the j -th down edge $(\alpha_j, i_1^{(j)})$. This contradicts the assumption that the subgraph with first l pairs of down and up edges has a cycle but has no cycle if the l -th up edge is deleted while removing the orientation and gluing multiple edges. Thus, the second class is empty.

Therefore, for any graphs in $\Delta_3(p, s; \alpha)$, one can split an appropriate vertex in the sequence i_1 to get a graph in $\Delta_1(p, s; \alpha)$. Equivalently, any graphs in $\Delta_3(p, s; \alpha)$ can be obtained from graphs in $\Delta_1(p, s; \alpha)$ by gluing two vertices of the sequence i_1 without common neighborhood. Moreover, gluing different pairs of vertices of the sequence i_1 leads to different graphs in $\Delta_3(p, s; \alpha)$.

For a fixed $\alpha \in \mathcal{C}_{s,p}^{(1)}$, by Lemma 3.1, there exists a unique sequence i_1 , such that the graph $g(i_1, \alpha) \in \Delta_1(p, s; \alpha)$. Moreover, the sequence i_1 has $p + 1 - s$ different vertices. Hence, the number of choices of gluing two vertices in the sequence i_1 is $\binom{p+1-s}{2}$. By Lemma 3.3, the number of choices of gluing two vertices in the sequence i_1 that have different neighborhoods is given by (3.17). \square

By Lemmas 3.3, 3.6 and Equation (3.15), we have

$$\begin{aligned} & \frac{1}{n^k} \mathbb{E}[\text{Tr} M_{n,k,m}^p] \\ &= \sum_{s=1}^p \left(\frac{m}{n^k}\right)^s (1 + o(1)) \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^s \frac{1}{m} \tau_{\varphi(t)}^{\text{deg}_t(\alpha)} \right) \times \\ & \quad \left[1 - \frac{(p+1-s)(p-s)}{2n} + \frac{1}{n} \left(\binom{p+1-s}{2} - \sum_{t=1}^s \binom{\text{deg}_t(\alpha)}{2} + m_4 \sum_{t=1}^s \binom{\text{deg}_t(\alpha)}{2} \right) \right]^k \\ &= \sum_{s=1}^p \left(\frac{m}{n^k}\right)^s (1 + o(1)) \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^s \frac{1}{m} \tau_{\varphi(t)}^{\text{deg}_t(\alpha)} \right) \left[1 + \frac{m_4 - 1}{n} \sum_{t=1}^s \binom{\text{deg}_t(\alpha)}{2} \right]^k \\ &= \sum_{s=1}^p \left(\frac{m}{n^k}\right)^s (1 + o(1)) \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \prod_{t=1}^s \left(\frac{1}{m} \sum_{j=1}^m \tau_j^{\text{deg}_t(\alpha)} \right) \left[1 + \frac{m_4 - 1}{n} \sum_{t=1}^s \binom{\text{deg}_t(\alpha)}{2} \right]^k, \quad (3.18) \end{aligned}$$

where we use the following equality in the last equality

$$\sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^s \frac{1}{m} \tau_{\varphi(t)}^{\text{deg}_t(\alpha)} = \sum_{\substack{j_1, \dots, j_s=1 \\ \forall p \neq q, j_p \neq j_q}}^m \prod_{t=1}^s \frac{1}{m} \tau_{j_t}^{\text{deg}_t(\alpha)} = \prod_{t=1}^s \left(\frac{1}{m} \sum_{j=1}^m \tau_j^{\text{deg}_t(\alpha)} \right) (1 + o(1)).$$

Spectral distribution of sample covariance matrices from large tensor products

Letting $m, n, k \rightarrow \infty$ in (3.18) with the ratio (1.2) and the limit assumption (2.2), we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^k} \mathbb{E}[\text{Tr} M_{n,k,m}^p] = \sum_{s=1}^p c^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\prod_{t=1}^s m_{\deg_t(\alpha)}^{(\tau)} \right) \exp \left(d(m_4 - 1) \sum_{t=1}^s \binom{\deg_t(\alpha)}{2} \right). \tag{3.19}$$

3.2 Proof of (ii)

For any $p \in \mathbb{N}$, we compute the variance of the moment

$$\text{Var} \left(\frac{1}{n^k} \text{Tr} M_{n,k,m}^p \right) = \frac{1}{n^{2k}} \mathbb{E} \left[\left(\text{Tr} M_{n,k,m}^p \right)^2 \right] - \frac{1}{n^{2k}} \left(\mathbb{E} \left[\text{Tr} M_{n,k,m}^p \right] \right)^2. \tag{3.20}$$

We use the convention that $\alpha_{p+1} = \alpha_1$ and $\beta_{p+1} = \beta_1$. Recalling the model (1.1), we can write

$$\begin{aligned} & \frac{1}{n^{2k}} \mathbb{E} \left[\left(\text{Tr} M_{n,k,m}^p \right)^2 \right] \\ &= \frac{1}{n^{2k}} \sum_{\alpha_1, \dots, \alpha_p=1}^m \sum_{\beta_1, \dots, \beta_p=1}^m \left(\prod_{t=1}^p \tau_{\alpha_t} \tau_{\beta_t} \right) \mathbb{E} \left[\text{Tr} (Y_{\alpha_{p+1}} Y_{\alpha_1}^* \cdots Y_{\alpha_p} Y_{\alpha_p}^*) \text{Tr} (Y_{\beta_{p+1}} Y_{\beta_1}^* \cdots Y_{\beta_p} Y_{\beta_p}^*) \right] \\ &= \frac{1}{n^{2k}} \sum_{\alpha_1, \dots, \alpha_p=1}^m \sum_{\beta_1, \dots, \beta_p=1}^m \left(\prod_{t=1}^p \tau_{\alpha_t} \tau_{\beta_t} \right) \\ & \quad \times \mathbb{E} \left[\sum_{i^{(1)}, \dots, i^{(p)}=1}^{n^k} \sum_{j^{(1)}, \dots, j^{(p)}=1}^{n^k} \prod_{t=1}^p \overline{Y_{i^{(t)} \alpha_t} Y_{i^{(t)} \alpha_{t+1}}} \overline{Y_{j^{(t)} \beta_t} Y_{j^{(t)} \beta_{t+1}}} \right] \\ &= \frac{1}{n^{2k}} \sum_{\alpha_1, \dots, \alpha_p=1}^m \sum_{\beta_1, \dots, \beta_p=1}^m \left(\prod_{t=1}^p \tau_{\alpha_t} \tau_{\beta_t} \right) \\ & \quad \times \mathbb{E} \left[\sum_{i^{(1)}, \dots, i^{(p)}=1}^{n^k} \sum_{j^{(1)}, \dots, j^{(p)}=1}^{n^k} \prod_{t=1}^p \prod_{l=1}^k \left(\overline{(\mathbf{y}_{\alpha_t}^{(l)})_{i_t^{(t)}} (\mathbf{y}_{\alpha_{t+1}}^{(l)})_{i_t^{(t)}}} \overline{(\mathbf{y}_{\beta_t}^{(l)})_{j_t^{(t)}} (\mathbf{y}_{\beta_{t+1}}^{(l)})_{j_t^{(t)}}} \right) \right]. \tag{3.21} \end{aligned}$$

The expectation above can be factorized as follows, using the i.i.d. setting:

$$\begin{aligned} & \mathbb{E} \left[\prod_{l=1}^k \sum_{i_1^{(1)}, \dots, i_1^{(p)}=1}^n \sum_{j_1^{(1)}, \dots, j_1^{(p)}=1}^n \prod_{t=1}^p \left(\overline{(\mathbf{y}_{\alpha_t}^{(l)})_{i_1^{(t)}} (\mathbf{y}_{\alpha_{t+1}}^{(l)})_{i_1^{(t)}}} \overline{(\mathbf{y}_{\beta_t}^{(l)})_{j_1^{(t)}} (\mathbf{y}_{\beta_{t+1}}^{(l)})_{j_1^{(t)}}} \right) \right] \\ &= \left(\mathbb{E} \left[\sum_{i_1^{(1)}, \dots, i_1^{(p)}=1}^n \sum_{j_1^{(1)}, \dots, j_1^{(p)}=1}^n \prod_{t=1}^p \left(\overline{(\mathbf{y}_{\alpha_t}^{(1)})_{i_1^{(t)}} (\mathbf{y}_{\alpha_{t+1}}^{(1)})_{i_1^{(t)}}} \overline{(\mathbf{y}_{\beta_t}^{(1)})_{j_1^{(t)}} (\mathbf{y}_{\beta_{t+1}}^{(1)})_{j_1^{(t)}}} \right) \right] \right)^k. \tag{3.22} \end{aligned}$$

On the other hand, by (3.5), we have

$$\begin{aligned} & \frac{1}{n^{2k}} \left(\mathbb{E} \left[\text{Tr} M_{n,k,m}^p \right] \right)^2 \\ &= \frac{1}{n^{2k}} \sum_{\alpha_1, \dots, \alpha_p=1}^m \sum_{\beta_1, \dots, \beta_p=1}^m \left(\prod_{t=1}^p \tau_{\alpha_t} \tau_{\beta_t} \right) \left(\mathbb{E} \left[\sum_{i_1^{(1)}, \dots, i_1^{(p)}=1}^n \prod_{t=1}^p \left(\overline{(\mathbf{y}_{\alpha_t}^{(1)})_{i_1^{(t)}} (\mathbf{y}_{\alpha_{t+1}}^{(1)})_{i_1^{(t)}}} \right) \right] \right) \\ & \quad \times \mathbb{E} \left[\sum_{j_1^{(1)}, \dots, j_1^{(p)}=1}^n \prod_{t=1}^p \left(\overline{(\mathbf{y}_{\beta_t}^{(1)})_{j_1^{(t)}} (\mathbf{y}_{\beta_{t+1}}^{(1)})_{j_1^{(t)}}} \right) \right]^k. \tag{3.23} \end{aligned}$$

Spectral distribution of sample covariance matrices from large tensor products

By plugging (3.22) into (3.21) and using (3.23) and (3.20), we have

$$\begin{aligned} & \text{Var} \left(\frac{1}{n^k} \text{Tr} M_{n,k,m}^p \right) \tag{3.24} \\ &= \frac{1}{n^{2k}} \sum_{\alpha_1, \dots, \alpha_p=1}^m \sum_{\beta_1, \dots, \beta_p=1}^m \left(\prod_{t=1}^p \tau_{\alpha_t} \tau_{\beta_t} \right) \\ & \times \left\{ \left(\mathbb{E} \left[\sum_{i_1^{(1)}, \dots, i_1^{(p)}=1}^n \sum_{j_1^{(1)}, \dots, j_1^{(p)}=1}^n \prod_{t=1}^p \left(\overline{(\mathbf{y}_{\alpha_t}^{(1)})_{i_1^{(t)}} (\mathbf{y}_{\alpha_{t+1}}^{(1)})_{i_1^{(t)}} (\mathbf{y}_{\beta_t}^{(1)})_{j_1^{(t)}} (\mathbf{y}_{\beta_{t+1}}^{(1)})_{j_1^{(t)}}} \right) \right] \right)^k \right. \\ & \left. - \left(\mathbb{E} \left[\sum_{i_1^{(1)}, \dots, i_1^{(p)}=1}^n \prod_{t=1}^p \left(\overline{(\mathbf{y}_{\alpha_t}^{(1)})_{i_1^{(t)}} (\mathbf{y}_{\alpha_{t+1}}^{(1)})_{i_1^{(t)}}} \right) \right] \right) \right. \\ & \left. \times \mathbb{E} \left[\sum_{j_1^{(1)}, \dots, j_1^{(p)}=1}^n \prod_{t=1}^p \left(\overline{(\mathbf{y}_{\beta_t}^{(1)})_{j_1^{(t)}} (\mathbf{y}_{\beta_{t+1}}^{(1)})_{j_1^{(t)}}} \right) \right] \right)^k \right\}. \end{aligned}$$

Denote the sequences $\alpha = (\alpha_1, \dots, \alpha_p)$, $\beta = (\beta_1, \dots, \beta_p)$, $i_1 = (i_1^{(1)}, \dots, i_1^{(p)})$ and $j_1 = (j_1^{(1)}, \dots, j_1^{(p)})$. We use the notation $\alpha \cap \beta = \emptyset$ if the two sequences α and β have no common vertices. Note that when $\alpha \cap \beta = \emptyset$, the two random variables

$$\prod_{t=1}^p \left(\overline{(\mathbf{y}_{\alpha_t}^{(1)})_{i_1^{(t)}} (\mathbf{y}_{\alpha_{t+1}}^{(1)})_{i_1^{(t)}}} \right), \text{ and } \prod_{t=1}^p \left(\overline{(\mathbf{y}_{\beta_t}^{(1)})_{j_1^{(t)}} (\mathbf{y}_{\beta_{t+1}}^{(1)})_{j_1^{(t)}}} \right)$$

are independent, for any sequences $i_1, j_1 \in [n]^p$; they will not contribute to the variance above. Hence, (3.24) can be written as

$$\begin{aligned} & \text{Var} \left(\frac{1}{n^k} \text{Tr} M_{n,k,m}^p \right) \\ &= \frac{1}{n^{2k}} \sum_{\alpha, \beta \in [m]^p, \alpha \cap \beta \neq \emptyset} \left(\prod_{t=1}^p \tau_{\alpha_t} \tau_{\beta_t} \right) \\ & \times \left\{ \left(\sum_{i_1, j_1 \in [n]^p} \mathbb{E} \left[\prod_{t=1}^p \left(\overline{(\mathbf{y}_{\alpha_t}^{(1)})_{i_1^{(t)}} (\mathbf{y}_{\alpha_{t+1}}^{(1)})_{i_1^{(t)}} (\mathbf{y}_{\beta_t}^{(1)})_{j_1^{(t)}} (\mathbf{y}_{\beta_{t+1}}^{(1)})_{j_1^{(t)}}} \right) \right] \right)^k \right. \\ & \left. - \left(\sum_{i_1, j_1 \in [n]^p} \mathbb{E} \left[\prod_{t=1}^p \left(\overline{(\mathbf{y}_{\alpha_t}^{(1)})_{i_1^{(t)}} (\mathbf{y}_{\alpha_{t+1}}^{(1)})_{i_1^{(t)}}} \right) \right] \mathbb{E} \left[\prod_{t=1}^p \left(\overline{(\mathbf{y}_{\beta_t}^{(1)})_{j_1^{(t)}} (\mathbf{y}_{\beta_{t+1}}^{(1)})_{j_1^{(t)}}} \right) \right] \right)^k \right\} \\ &= \frac{1}{n^{2k}} \sum_{\alpha, \beta \in [m]^p, \alpha \cap \beta \neq \emptyset} \left(\prod_{t=1}^p \tau_{\alpha_t} \tau_{\beta_t} \right) \\ & \times \left\{ \left(\sum_{i_1, j_1 \in [n]^p} E'(i_1, \alpha; j_1, \beta) \right)^k - \left(\sum_{i_1, j_1 \in [n]^p} E(i_1, \alpha) E(j_1, \beta) \right)^k \right\}, \tag{3.25} \end{aligned}$$

where $E(\cdot, \cdot)$ is defined in (3.6), and $E'(i_1, \alpha; j_1, \beta)$ is given by

$$E'(i_1, \alpha; j_1, \beta) = \mathbb{E} \left[\prod_{t=1}^p \left(\overline{(\mathbf{y}_{\alpha_t}^{(1)})_{i_1^{(t)}} (\mathbf{y}_{\alpha_{t+1}}^{(1)})_{i_1^{(t)}} (\mathbf{y}_{\beta_t}^{(1)})_{j_1^{(t)}} (\mathbf{y}_{\beta_{t+1}}^{(1)})_{j_1^{(t)}}} \right) \right].$$

Spectral distribution of sample covariance matrices from large tensor products

We use the notation $g(i_1, \alpha) \cup g(j_1, \beta)$ for the graph obtained from $g(i_1, \alpha)$ and $g(j_1, \beta)$ by joining two graphs together and keep the coincident edges. If the graph $g(i_1, \alpha) \cup g(j_1, \beta)$ has a single edge, then this single edge must belong to one of the graphs $g(i_1, \alpha)$ and $g(j_1, \beta)$. Since the variables have mean zero, only the summation indices i_1, j_1 such that the graph $g(i_1, \alpha) \cup g(j_1, \beta)$ has no single edge contribute to (3.25). Note that the graph $g(i_1, \alpha) \cup g(j_1, \beta)$ is connected with $4p$ edges, the degrees of the vertices are at least 4 with at most two exceptions, whose degrees are at least 2. Denote $s = |(\alpha, \beta)|$ and $r = |(i_1, j_1)|$, then the term $E'(i_1, \alpha; j_1, \beta)$ is non-vanishing when $r + s \leq 2p + 1$.

Besides, if $\alpha \cap \beta \neq \emptyset$, then $s \leq |\alpha| + |\beta| - 1$. Hence, the term $E(i_1, \alpha)E(j_1, \beta)$ is non-vanishing when $|i_1| + |\alpha| \leq p + 1$ and $|j_1| + |\beta| \leq p + 1$, which implies that $r + s \leq |i_1| + |j_1| + (|\alpha| + |\beta| - 1) \leq 2p + 1$. Hence, for n large,

$$\begin{aligned} & \text{Var} \left(\frac{1}{n^k} \text{Tr} M_{n,k,m}^p \right) \\ &= \frac{1}{n^{2k}} \sum_{s=1}^{2p} \sum_{(\alpha, \beta) \in \mathcal{C}_{s,2p}, \alpha \cap \beta \neq \emptyset} \left(\sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^p \tau_{\varphi(\alpha_t)} \tau_{\varphi(\beta_t)} \right) \\ & \quad \times \left\{ \left(\sum_{r=1}^{2p} n^r (1 + O(n^{-1})) \sum_{(i_1, j_1) \in \mathcal{C}_{r,2p}} E'(i_1, \alpha; j_1, \beta) \right)^k \right. \\ & \quad \left. - \left(\sum_{r=1}^{2p} n^r (1 + O(n^{-1})) \sum_{(i_1, j_1) \in \mathcal{C}_{r,2p}} E(i_1, \alpha) E(j_1, \beta) \right)^k \right\} \\ & \leq \frac{1}{n^{2k}} \sum_{s=1}^{2p} \sum_{(\alpha, \beta) \in \mathcal{C}_{s,2p}, \alpha \cap \beta \neq \emptyset} \left(\sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^s \tau_{\varphi(t)}^{\deg_t(\alpha) + \deg_t(\beta)} \right) \times C_p^k n^{(1-s)k} \\ & = \frac{C_p^k}{n^k} \sum_{s=1}^{2p} \left(\frac{m}{n^k} \right)^s \sum_{(\alpha, \beta) \in \mathcal{C}_{s,2p}, \alpha \cap \beta \neq \emptyset} \left(\sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^s \frac{1}{m} \tau_{\varphi(t)}^{\deg_t(\alpha) + \deg_t(\beta)} \right) \\ & \leq \frac{C_p^k}{n^k} \sum_{s=1}^{2p} \left(\frac{m}{n^k} \right)^s \sum_{(\alpha, \beta) \in \mathcal{C}_{s,2p}, \alpha \cap \beta \neq \emptyset} \prod_{t=1}^s \left(\frac{1}{m} \sum_{j=1}^m |\tau_j|^{\deg_t(\alpha) + \deg_t(\beta)} \right) \\ & \leq \frac{C_p^k}{n^k}. \end{aligned} \tag{3.26}$$

where C_p is a positive number that depends only on p and may vary in different places. Therefore, for all fixed $p \in \mathbb{N}_+$, if $k \geq 2$, we have

$$\sum_{n=1}^{\infty} \text{Var} \left(\frac{1}{n^k} \text{Tr} M_{n,k,m}^p \right) < \infty.$$

4 Proof of Proposition 2.4 (the case of $d = 0$)

The variance calculation (ii) is identical to the one given in the proof of Theorem 2.1, see Section 3.2. It remains to establish the moment limit (i).

Note that the moment formula (3.7) is also valid for the case $d = 0$. We use the graph $g(i_1, \alpha)$ to compute the sum on i_1 in (3.7). For any sequences $\alpha \in \mathcal{C}_{s,p}$ and $i_1 \in \mathcal{C}_{r,p}$ satisfying $g(i_1, \alpha) \in \Delta_1(p, s; \alpha)$, we have $r + s = p + 1$ and $E(i_1, \alpha) = n^{-p}$. Besides, $E(i_1, \alpha) = 0$ if the graph $g(i_1, \alpha) \in \Delta_2(p, r, s; \alpha)$. Moreover, if $g(i_1, \alpha) \notin \Delta_1(p, s; \alpha) \cup$

Spectral distribution of sample covariance matrices from large tensor products

$\Delta_2(p, r, s; \alpha)$, then $r + s \leq p$ and $E(i_1, \alpha) = O(n^{-p})$. Hence,

$$\begin{aligned} & \sum_{r=1}^p n \cdots (n-r+1) \sum_{i_1 \in \mathcal{C}_{r,p}} E(i_1, \alpha) \\ &= \sum_{r=1}^p n^r (1 + O(n^{-1})) \sum_{i_1 \in \mathcal{C}_{r,p}} E(i_1, \alpha) \\ &= \sum_{r=1}^p n^r (1 + O(n^{-1})) \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \in \Delta_1(p, s; \alpha)}} E(i_1, \alpha) \\ & \quad + \sum_{r=1}^p n^r (1 + O(n^{-1})) \sum_{\substack{i_1 \in \mathcal{C}_{r,p} \\ g(i_1, \alpha) \notin \Delta_1(p, s; \alpha) \cup \Delta_2(p, r, s; \alpha)}} E(i_1, \alpha) \\ &= n^{1-s} (1 + O(n^{-1})) \sum_{\substack{i_1 \in \mathcal{C}_{p+1-s,p} \\ g(i_1, \alpha) \in \Delta_1(p, s; \alpha)}} 1 + O(n^{-s}). \end{aligned} \tag{4.1}$$

Therefore, by (3.7), (4.1) and Lemma 3.1, we have

$$\begin{aligned} & \frac{1}{n^k} \mathbb{E}[\text{Tr} M_{n,k,m}^p] \\ &= \frac{1}{n^k} \sum_{s=1}^p \sum_{\alpha \in \mathcal{C}_{s,p}} \left(\sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^p \tau_{\varphi(\alpha_t)} \right) \left(n^{1-s} (1 + O(n^{-1})) \sum_{\substack{i_1 \in \mathcal{C}_{p+1-s,p} \\ g(i_1, \alpha) \in \Delta_1(p, s; \alpha)}} 1 + O(n^{-s}) \right)^k \\ &= \frac{1}{n^k} \sum_{s=1}^p \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^p \tau_{\varphi(\alpha_t)} \right) \left(n^{1-s} (1 + O(n^{-1})) + O(n^{-s}) \right)^k \\ & \quad + \frac{1}{n^k} \sum_{s=1}^p \sum_{\alpha \in \mathcal{C}_{s,p} \setminus \mathcal{C}_{s,p}^{(1)}} \left(\sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^p \tau_{\varphi(\alpha_t)} \right) (O(n^{-s}))^k \\ &= \frac{1}{n^k} \sum_{s=1}^p \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^p \tau_{\varphi(\alpha_t)} \right) \times n^{k(1-s)} (1 + o(1)) \\ &= \sum_{s=1}^p \left(\frac{m}{n^k} \right)^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\frac{1}{m^s} \sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^p \tau_{\varphi(\alpha_t)} \right) (1 + o(1)) \\ &= \sum_{s=1}^p \left(\frac{m}{n^k} \right)^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\prod_{t=1}^s \left(\frac{1}{m} \sum_{j=1}^m \tau_j^{\deg_t(\alpha)} \right) \right) (1 + o(1)), \end{aligned} \tag{4.2}$$

where we used $k = o(n)$ in the third equality. Letting $m, n, k \rightarrow \infty$ with the ratio (1.2), we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^k} \mathbb{E}[\text{Tr} M_{n,k,m}^p] = \sum_{s=1}^p c^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\prod_{t=1}^s m_{\deg_t(\alpha)}^{(\tau)} \right).$$

Remark 4.1. In the case $\tau_1 = \tau_2 = \cdots = 1$, by Lemma 3.1, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^k} \mathbb{E}[\text{Tr} M_{n,k,m}^p] = \sum_{s=1}^p \frac{1}{p} \binom{p}{s-1} \binom{p}{s} c^s,$$

which is the p -th moment of the Marchenko-Pastur law.

Spectral distribution of sample covariance matrices from large tensor products

References

- [1] Andris Ambainis, Aram W. Harrow, and Matthew B. Hastings, *Random tensor theory: extending random matrix theory to mixtures of random product states*, *Comm. Math. Phys.* **310** (2012), no. 1, 25–74. MR2885613
- [2] Zhidong Bai and Jack W. Silverstein, *Spectral analysis of large dimensional random matrices*, Second, Springer Series in Statistics, Springer, New York, 2010. MR2567175
- [3] Zhidong Bai and Wang Zhou, *Large sample covariance matrices without independence structures in columns*, *Statist. Sinica* **18** (2008), no. 2, 425–442. MR2411613
- [4] Benoît Collins, Ion Nechita, and Karol Życzkowski, *Area law for random graph states*, *J. Phys. A* **46** (2013), no. 30, 305302, 18. MR3083280
- [5] Gwo Dong Lin, *Recent developments on the moment problem*, *Journal of Statistical Distributions and Applications* **4** (2017), no. 1.
- [6] Anna Lytova, *Central limit theorem for linear eigenvalue statistics for a tensor product version of sample covariance matrices*, *J. Theoret. Probab.* **31** (2018), no. 2, 1024–1057. MR3803923
- [7] Anna Lytova and Leonid A. Pastur, *Central limit theorem for linear eigenvalue statistics of the Wigner and the sample covariance random matrices*, *Ann. Probab.* **37** (2009), 1778–1840. MR2481919
- [8] Vladimir A. Marchenko and Leonid A. Pastur, *Distribution of eigenvalues for some sets of random matrices*, *Math. USSR-Sb* **1** (1967), 457–483.
- [9] Alain Pajor and Leonid A. Pastur, *On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution*, *Studia Mathematica* **195** (2009), no. 1, 11–29. MR2539559
- [10] Jack W. Silverstein, *Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices*, *J. Multivariate Anal.* **55** (1995), no. 2, 331–339. MR1370408
- [11] Pavel Yaskov, *Necessary and sufficient conditions for the Marchenko-Pastur theorem*, *Electron. Commun. Probab.* **21** (2016), Paper No. 73, 8. MR3568347
- [12] Karol Życzkowski, Karol A. Penson, Ion Nechita, and Benoît Collins, *Generating random density matrices*, *J. Math. Phys.* **52** (2011), no. 6, 062201, 20. MR2841746