On spectrum of sample covariance matrices from large tensor vectors

Wangjun Yuan*

July 3, 2023

Abstract

In this paper, we study the limiting spectral distribution of sums of independent rank-one large k-fold tensor products of large n-dimensional vectors. In the literature, the limiting moment sequence is obtained for the case $k = o(n)$ and $k = O(n)$. Under appropriate moment conditions on base vectors, it has been showed that the eigenvalue empirical distribution converges to the celebrated Marčenko-Pastur law if $k = O(n)$ and the components of base vectors have unit modulus, or $k = o(n)$. In this paper, we study the limiting spectral distribution by allowing k to grow much faster, whenever the components of base vectors are complex random variables on the unit circle. It turns out that the limiting spectral distribution is Marčenko-Pastur law. Comparing with the existing results, our limiting setting only requires $k \to \infty$. Our approach is based on the moment method.

AMS 2000 subject classifications: Primary 60B20; Secondary 15B52.

Keywords and phrases: Large k-fold tensors; Eigenvalue spectral distribution; Marčenko-Pastur law; Quantum information theory.

1 Introduction

For $n \in \mathbb{N}$, let $\mathbf{y} = \frac{1}{\sqrt{n}}(\xi_1, \ldots, \xi_n) \in \mathbb{C}^n$, where $\{\xi_1, \ldots, \xi_n\}$ is a family of i.i.d. centered random variables with unit variance, and let $\{\mathbf{y}_\alpha^{(l)} : 1 \leq \alpha \leq m, 1 \leq l \leq k\}$ be a family of i.i.d. copies of $\mathbf{y}$. Let $Y_\alpha = \mathbf{y}_\alpha^{(1)} \otimes \cdots \otimes \mathbf{y}_\alpha^{(k)}$ be k-fold tensor products of n-dimensional i.i.d. vectors for $1 \leq \alpha \leq m$. Since k-fold tensor product of n-dimensional vectors can be identify as a n^k-dimensional vector, the family $\{Y_\alpha : 1 \leq \alpha \leq m\}$ is an i.i.d. family of n^k-dimensional vectors. Let $\{\tau_1, \tau_2, \ldots\}$ be a sequence of real numbers, and $Y = (Y_1, \ldots, Y_m)$ be a $n^k \times m$ matrix. Consider the $n^k \times n^k$ Hermitian matrix

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

*Department of Mathematics, University of Luxembourg. E-mail: ywangjun@connect.hku.hk

Thus, $M_{n,k,m}$ is a sum of m independent rank-1 Hermitian matrices of dimension n^k.

We would like to point out that the tensor Y_α introduced above is the non-symmetric random tensor model. Instead of considering the tensor product of i.i.d. vectors, the k-fold tensor product $\mathbf{y}^{\otimes k}$ of the same random vector $\mathbf{y} \in \mathbb{C}^n$, which can be identified as a vector of dimension $\binom{n}{k}$ with components $(Y_\alpha)_{j_1\ldots j_k} = \mathbf{y}_{j_1}\ldots\mathbf{y}_{j_k}$ for distinct $j_1, \ldots, j_k$, is known as the symmetric random tensor model. The limiting spectral distribution of the Hermitian matrix (1.1) constructed by i.i.d. copy of $\mathbf{y}^{\otimes k}$ was studied in [5, 15].

In the simplest case $k = 1$, where the column vector $Y_\alpha = \mathbf{y}_\alpha^{(1)}$ has i.i.d. entries, was studied in the paper [11] under appropriate moment conditions on ξ_1. When $\tau_\alpha = 1$ for all α, the limiting spectral distribution (LSD) is the famous Marčenko-Pastur law, which is a probability distribution with density function

$$p(x) = \frac{\sqrt{((1+\sqrt{c})^2 - x)(x - (1-\sqrt{c})^2)}}{2\pi x}\mathbf{1}_{[(1-\sqrt{c})^2,(1+\sqrt{c})^2]}(x) + (1-c)\delta_0(dx)\mathbf{1}_{0<c<1}, \quad (1.2)$$

as $n \to +\infty$ and $m/n^k \to c$ with some positive constant c. We refer the interested readers to [2, 3]. Since then, a lot of works on the Marčenko-Pastur law appeared in the literature. See, for example, [4], [12], [13] and the reference therein. Later, the necessary and sufficient conditions on Y_1 such that the Marčenko-Pastur law serves as the LSD of $M_{n,1,m}$ were carried out in [14] when $\tau_\alpha = 1$ for all α.

The case $k \geq 2$ is very different from the case $k = 1$. Actually, comparing with the case $k = 1$, the tensor structure appears when $k \geq 2$, which results in the dependence of the entries of Y_α and non-unitary invariant. Recently, [9] obtained the LSD for the k-fold tensor model $M_{n,k,m}$ when k is large. For the special case $\tau_\alpha \equiv 1$, the LSD is exactly the Marčenko-Pastur law (1.2). A central limit theorem (CLT) is also established for a class of linear spectral statistics following the approach of [10] when $k = 2$. The main setup [9] is that the number of tensor fold k must be small enough compared to the space dimension n. More precisely, $k/n \to 0$ is required for the validity of the LSD. Very recently, [7] dealt with the case $k = O(n)$ for the model (1.1). Under the setting $k/n \to d$, the limiting moment sequence of the empirical spectral distribution (ESD) of $M_{n,k,m}$ was obtained assuming the finiteness of the all moments of ξ_1. It is interesting that, in the limiting moment sequence, the fourth moment of ξ_1 appears. If $\tau_\alpha \equiv 1$, the limiting moment sequence corresponds to the Marčenko-Pastur law (1.2) if and only if either $d = 0$ or $d > 0$ and the fourth moment of ξ_1 is 1.

In the present paper, we study the model (1.1) with ξ_1 being chosen from the unit circle on the complex plane. Under the limiting setting $k/n \to d \geq 0$ of [7], since we have $|\xi_1|^4 = 1$, a direct application of [7, Theorem 2.1] shows that the limiting moment sequence coincides with the moment sequence of Marčenko-Pastur law (1.2). We are interested in the limiting setting that k goes to infinity much faster than the setting in [7]. The present paper shows that the limiting moment sequence of the ESD of $M_{n,k,m}$ is exactly the Marčenko-Pastur law (1.2) when n, k goes to infinity. From the point of view of probability theory, our results extend the results in [7, 9] by removing the restriction on the speed of k approaching infinity. Our results even allows that k/n does not have a limit when $n \to \infty$. We would like to add one more sentence on the limiting setting in application. In the real world, one would expect that the dimension of a system is fixed and is reused for many times, which leads to the setting that k tends to infinity while n is fixed. We do not know how to study the model (1.1) with n fixed, and we plan to consider this case as our next model. As an variant of this real-world scenario, the results in this paper apply to the setting that $k \gg n \gg 1$.

Another motivation to study the model (1.1) is from quantum information theory. We consider a classical probability problem of allocating m balls randomly into n bins. Using the terminology of matrices, this classical probability problem is equivalent to the spectrum of the $M_{n,1,m}$ when Y_α is chosen randomly from the canonical basis $\{e_1, \ldots, e_n\}$ of $\mathbb{C}^n$. The eigenvalues of the $M_{n,1,m}$ are the frequency of the vectors $e_1, \ldots, e_n$ among of family $\{Y_\alpha : 1 \le \alpha \le m\}$. It was considered as a quantum analog for $M_{n,1,m}$ if the vector Y_α is chosen randomly from the unit sphere $\mathbb{C}^n$ on the complex plane. A modified version of the quantum problem, which is introduced in [1], is to choose the vector Y_1 from the random product states in $(\mathbb{C}^n)^{\otimes k}$. When k and m/n^k are fixed, and the base vector $\mathbf{y}$ is Gaussian, [1] established the convergence in expectation of the normalized trace of moments $n^{-k}\mathrm{Tr}M_{n,k,m}^p$ when $n \to +\infty$. The limiting moments coincide with the corresponding moments of the Marčenko-Pastur law (1.2). In quantum physics and quantum information theory, it is natural to investigate the behavior of a system with a large number of quantum states. The paper [16] 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. Besides, the asymptotic behavior of the average entropy of entanglement for elements of an ensemble of random states associated with a graph was studied in [6] when the dimension of the quantum subsystem is large.

In this paper, we consider the limiting spectrum distribution of $M_{n,k,m}$ where m, n, k grows to infinity under the following ratio

$$\frac{m}{n^k} \to c \tag{1.3}$$

for some constant $c \in (0, \infty)$. Under appropriate moment conditions on the sequence of coefficients $\{\tau_\alpha\}$, we derive the limits for the spectral distribution of $M_{n,k,m}$ under the limiting scheme (1.3).

The following theorem is the first the main results of the paper, where we establish the convergence in expectation of the moments.

Theorem 1.1. *Let $M_{n,k,m}$ be in (1.1) with $|\xi_1| = 1$. Suppose that for all $q \in \mathbb{N}$,*

$$\frac{1}{m}\sum_{j=1}^{m} \tau_j^q \to m_q^{(\tau)}, \quad m \to +\infty. \tag{1.4}$$

Then for any fixed $p \in \mathbb{N}_+$, we have

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

Here, $\deg_t(\alpha)$ *is the frequency of* t *in the sequence* α *and is given by* (2.2)*, and* $\mathcal{C}_{s,p}^{(1)}$ *is a set of sequences that is defined in Lemma 2.1.*

Our approach is based on the method of moments. We associate graphs to the addends in the sum of the moment, and compute the moment by distinguishing the graphs that contributes to the limit. The class $\mathcal{C}_{s,p}^{(1)}$ of sequences corresponds to the addends with the largest leading terms, while the addends of $\alpha \notin \mathcal{C}_{s,p}^{(1)}$ are negligible. Note that the tensor structure results in the power k of the terms. For the case $k = 1$, the limit of the moment sequence can be obtained by counting the size of $\mathcal{C}_{s,p}^{(1)}$, since only the leading terms in the sum of $\alpha \in \mathcal{C}_{s,p}^{(1)}$ contribute to the limit. We refer the interested readers to [3, Section 3.1.3]. This is also true whenever $k = o(n)$. See [7, Section 4]. For the case $k = O(n)$ studied in [7], the second leading term also contributes to the limit due to the order of the power $k = O(n)$ for $\alpha \in \mathcal{C}_{s,p}^{(1)}$. In our limiting setting, that k can grow much faster, for $\alpha \in \mathcal{C}_{s,p}^{(1)}$, all terms may contribute to the limit. Hence, we need to characterize all the possible graph associate to $\alpha \in \mathcal{C}_{s,p}^{(1)}$. This is the main novelty in methodology of the present paper.

In the next theorem, we strengthen Theorem 1.1 from convergence in expectation to almost sure convergence.

Theorem 1.2. *Assume that the conditions in Theorem 1.1 hold. Suppose that* $k = k(n)$ *is a function of* n *and tends to infinity as* $n \to \infty$*. Then for any fixed* $p \in \mathbb{N}_+$*,*

$$\lim_{n\to+\infty} \frac{1}{n^k} \mathrm{Tr} M_{n,k,m}^p = \sum_{s=1}^{p} c^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\prod_{t=1}^{s} m_{\deg_t(\alpha)}^{(\tau)} \right),$$

almost surely.

As we have the limiting moment sequence from Theorem 1.2, it is nature to expect that the moment sequence can determine a probability measure uniquely. If so, then we can establish the almost sure convergence of the ESD of $M_{n,k,m}$. The following corollary provides a condition that guarantees the uniqueness of the probability measure corresponding to the moment sequence.

Corollary 1.1. *Assume that the conditions in Theorem 1.1 hold. Suppose that there exists a positive constant* A*, such that* $|m_q^{(\tau)}| \le A^q q^q$ *for all* $q \in \mathbb{N}$*. Then there exists a probability measure* μ *whose moment sequence is*

$$\int_{\mathbb{R}} x^p \mu(dx) = \sum_{s=1}^{p} c^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\prod_{t=1}^{s} m_{\deg_t(\alpha)}^{(\tau)} \right), \quad \forall p \in \mathbb{N}_+. \tag{1.5}$$

Moreover, suppose that $k = k(n)$ *is a function of* n *and tends to infinity when* $n \to \infty$*, then the ESD of* $M_{n,k,m}$ *converges almost surely to* μ*.*

In particular, if $\tau_\alpha = 1$ *for all* $1 \le \alpha \le m$*, then the ESD of* $M_{n,k,m}$ *converges almost surely to the Marčenko-Pastur law* (1.2)*.*

The rest of the paper is organized as follow. We develop the theory of graph combinatories in Section 2. We first introduce some results from literature on the graph combinatorics in Section 2.1. Then we characterize the set $\mathcal{C}_{s,p}^{(1)}$, which corresponding to the leading term in the moment computation. The paired graph, which is the graph that contributes to the limit for $\alpha \in \mathcal{C}_{s,p}^{(1)}$, is studied in Section 2.3. Then we prove the main theorems in Section 3. The proofs of Theorem 1.1, Theorem 1.2 and Corollary 1.1 are presented in Section 3.1, Section 3.2 and Section 3.3, respectively.

2 Graph combinatorics

In this section, we study the graph combinatorics, which will be used in the proof in Section 3.

2.1 Preliminaries

In this subsection, we introduce some preliminaries on graph combinatorics, which can be found in [3, 7].

For a positive integer s, we denote by $[s]$ the set of integers from 1 to s. We call $\alpha = (\alpha_1, \dots, \alpha_p) \in [m]^p$ a sequence of length p with vertices α_j for $1 \le j \le p$. We denote by $|\alpha|$ the 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

$$[m]^p = \bigcup_{s=1}^{p} \mathcal{J}_{s,p}(m). \tag{2.1}$$

For a sequence $\alpha = (\alpha_1, \dots, \alpha_p)$ and each value t in α, we count its frequency by

$$\deg_t(\alpha) = \#\{j \in [p] : \alpha_j = t\}. \tag{2.2}$$

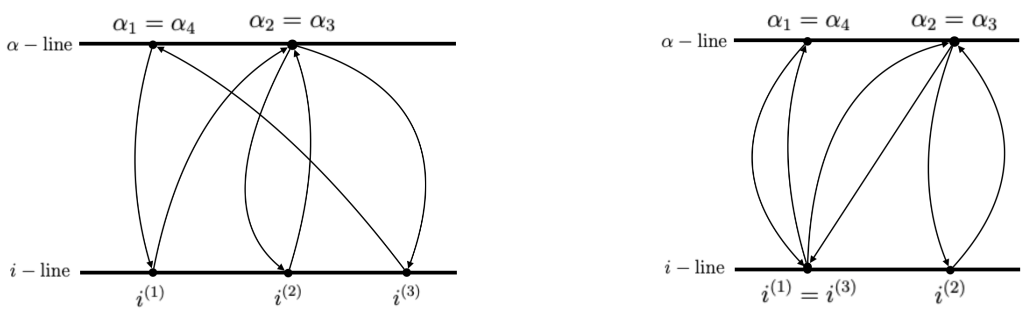
where we use the notation $\#S$ for the number of elements in the set S.

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 $\alpha_u \le \max\{\alpha_1, \dots, \alpha_{u-1}\} + 1$ for $u \ge 2$. 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.

We fixed a canonical s-sequence $\alpha = (\alpha_1, \dots, \alpha_p) \in [m]^p$. For $i = (i^{(1)}, \dots, i^{(p)}) \in [n]^p$, draw two parallel lines referred as the α-line and the i-line, respectively. Plot $i^{(1)}, \dots, i^{(p)}$ on the i-line and $\alpha_1, \dots, \alpha_p$ on the α-line. Draw p down edges from α_u to $i^{(u)}$ and p up edges from $i^{(u)}$ to α_{u+1} for $1 \le u \le p$ with the convention that $\alpha_1 = \alpha_{p+1}$. We denote the graph by $g(i, \alpha)$ and call such graph a $\Delta(p; \alpha)$-graph. From the definition, one can easily see that the graph $g(i, \alpha)$ is a connected directed graph with up edges and down edges appear alternatively. An example of the $\Delta(p; \alpha)$-graph is given in (a) of the Figure 1.

Two graphs $g(i, \alpha)$ and $g(i', \alpha)$ are called equivalent if the two sequences i and i' are equivalent, and we write $g(i, \alpha) \sim g(i', \alpha)$ for this equivalence. For each equivalent class, we choose the canonical graph such that $i = (i^{(1)}, \ldots, i^{(p)}) \in [n]^p$ 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 call a graph $\Delta_1(p, s; \alpha)$-graph if it is a $\Delta(p; \alpha)$-graphs such that each down edge coincide with exactly one up edge and if we glue the pair of coincident edges and remove the orientation, the resulting graph is a tree with p edges and $p+1$ vertices. Hence, we have $r + s = p + 1$. We give an example of $\Delta_1(p, s; \alpha)$-graph in (b) of Figure 1.



(a) $\Delta(p, \alpha)$-graph with $p = 3$, $\alpha = (1, 2, 2)$, $i = (1, 2, 3)$.

(b) $\Delta_1(p, s; \alpha)$-graph with $p = 3, s = 2$, $\alpha = (1, 2, 2)$, $i = (1, 2, 1)$.

Figure 1

For a given sequence $\alpha \in \mathcal{C}_{s,p}$, the following lemma determines the number of sequences $i \in \mathcal{C}_{p+1-s,p}$ such that $g(i, \alpha) \in \Delta_1(p, s; \alpha)$.

Lemma 2.1. *([7, Lemma 3.1]) For any $1 \le s \le p$ and any sequence $\alpha \in \mathcal{C}_{s,p}$, there is at most one sequence $i \in \mathcal{C}_{p+1-s,p}$ such that $g(i, \alpha) \in \Delta_1(p, s; \alpha)$. We denote by $\mathcal{C}^{(1)}_{s,p}$ the set of such canonical sequences α. Then the number of the elements in $\mathcal{C}^{(1)}_{s,p}$ is*

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

2.2 Characterization of $\mathcal{C}^{(1)}_{s,p}$

In this subsection, we establish some properties of the sequences in $\mathcal{C}^{(1)}_{s,p}$. We start with the following definition.

Definition 1. *A sequence $\alpha = (\alpha_1, \ldots, \alpha_p)$ is called a* crossing *sequence if there exist $j_1 < j_2 < j_3 < j_4$, such that $\alpha_{j_1} = \alpha_{j_3} \neq \alpha_{j_2} = \alpha_{j_4}$. A sequence is called* non-crossing *sequence if it is not a crossing sequence.*

The following theorem is a characterization of the set $\mathcal{C}^{(1)}_{s,p}$.

Theorem 2.1. *For any $1 \le s \le p$, the set of all non-crossing sequences $\alpha \in \mathcal{C}_{s,p}$ is $\mathcal{C}^{(1)}_{s,p}$.*

The proof of Theorem 2.1 follows directly from the two lemmas below.

Lemma 2.2. *([7, Lemma 3.4]) For any $1 \le s \le p$ and any sequence $\alpha \in \mathcal{C}_{s,p}$, if α is a crossing sequence, then $\alpha \notin \mathcal{C}^{(1)}_{s,p}$.*

Lemma 2.3. *For any $1 \le s \le p$ and any sequence $\alpha \in \mathcal{C}_{s,p}$, if α is a non-crossing sequence, then $\alpha \in \mathcal{C}^{(1)}_{s,p}$.*

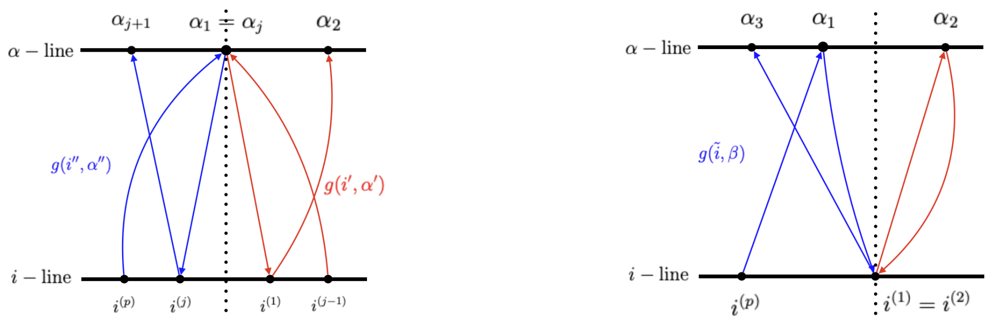
Proof. (of Lemma 2.3) We prove by induction on p. The case $p = 1$ is trivial, since $\alpha = (\alpha_1)$ and $i = (i^{(1)})$, so $g(i, \alpha)$ is the graph with exactly one up edge from $i^{(1)}$ to α_1 and one down edge from α_1 to $i^{(1)}$.

Assume that Lemma 2.3 holds for sequence of length at most $p - 1$ for $p \ge 2$. We need to show Lemma 2.3 holds for any non-crossing sequence $\alpha = (\alpha_1, \ldots, \alpha_p) \in \mathcal{C}_{s,p}$. We consider the following two cases according to whether α_1 coincide with other vertices.
Case 1. There exists $1 < j < p + 1$, such that $\alpha_j = \alpha_1$.

In this case, we split the sequence α to two subsequences $\alpha' = (\alpha_1, \ldots, \alpha_{j-1})$ and $\alpha'' = (\alpha_j, \ldots, \alpha_p)$. For the subsequence α', it is canonical s'-sequence for some $s' < s$. Moreover, α' is non-crossing and has length $j - 1 \le p - 1$, so by induction hypothesis, we have $\alpha' \in \mathcal{C}^{(1)}_{s',j-1}$, and thus, there exists a canonical $(j - s')$-sequence i' of length $j - 1$, such that $g(i', \alpha') \in \Delta_1(j - 1, s'; \alpha')$.

For the subsequence α'', it is also non-crossing but not canonical. The non-crossing property allows us to identify α'' to a canonical sequence. Note that the vertices of α' take values in $[s']$, so the vertices of α'' take values in $\{1\} \cup \{s' + 1, \ldots, s\}$. We define $\beta'' = (\beta_j, \ldots, \beta_p)$ by setting $\beta_k = \alpha_k$ if $\alpha_k = \alpha_1$, and $\beta_k = \alpha_k - s' + 1$ if $\alpha_k \ne \alpha_1$. Now β'' is canonical non-crossing $(s - s' + 1)$-sequence of length $p - j + 1 \le p - 1$. Hence, by induction hypothesis, we have $\beta'' \in \mathcal{C}^{(1)}_{s-s'+1,p-j+1}$ and there exists a canonical $(p - j - s + s' + 1)$-sequence i'' of length $p - j + 1$, such that $g(i'', \beta'') \in \Delta_1(p - j + 1, s - s' + 1; \beta'')$.

The sequence i which satisfies $g(i, \alpha) \in \Delta_1(p, s; \alpha)$ can be obtain by 'gluing' the two sequence i' and i''. We provide the Figure 2 part (a) for the idea of gluing two graphs. More precisely, we define a canonical $(p - s + 1)$-sequence $i = (i^{(1)}, \ldots, i^{(p)})$ by $i^{(k)} = i'^{(k)}$ for $1 \le k \le j - 1$, and $i^{(k)} = i''^{(k)} + j - s'$ for $j \le k \le p$. One can easily check that i is a canonical $(p - s + 1)$ sequence, and the subsequence $(i^{(1)}, \ldots, i^{(j-1)})$ and $(i^{(j)}, \ldots, i^{(p)})$ has no common vertex. Thus, the subgraph of $g(i, \alpha)$ from α_1 to α_j and the subgraph from α_j to α_{p+1} do not have coincide edge and satisfy the definition of $\Delta_1(p, s; \alpha)$-graph. Therefore, we can conclude that the graph $g(i, \alpha) \in \Delta_1(p, s; \alpha)$ so $\alpha \in \mathcal{C}^{(1)}_{s,p}$.
Case 2. For any $1 < j < p+1$, $\alpha_j \ne \alpha_1$. In this case, we have $\alpha_1 = 1, \alpha_2 = 2$. We consider the following two subcases.
Case 2(a). If for any $2 < k < p+1$, $\alpha_k \ne \alpha_2$. We consider the sequence $\beta = (\beta_1, \ldots, \beta_{p-1})$ given by $\beta_1 = \alpha_1$ and $\beta_k = \alpha_{k+1} - 1$. Then β is a non-crossing canonical $(s - 1)$-sequence of length $p - 1$. By induction hypothesis, $\beta \in \mathcal{C}^{(1)}_{s-1,p-1}$, and $g(\tilde{i}, \beta) \in \Delta_1(p - 1, s - 1; \beta)$



(a) Combine two $\Delta_1(p, s; \alpha)$-graphs $g(i', \alpha')$ and $g(i'', \alpha'')$.

(b) Insert the coincident edges $i^{(1)} \to \alpha_2$ and $\alpha_2 \to i^{(2)}$.

Figure 2

for some canonical $(p-s+1)$-sequence $\tilde{i} = (\tilde{i}^{(1)}, \ldots, \tilde{i}^{(p-1)})$. The sequence i which satisfies $g(i, \alpha) \in \Delta_1(p, s; \alpha)$ can be obtained by 'inserting' the vertex of value 1 to the sequence $\tilde{i}$ between $\tilde{i}^{(1)}$ and $\tilde{i}^{(2)}$. The part (b) of Figure 2 is provided for the idea of inserting the coincident edges between $i^{(1)} = i^{(2)}$ and α_2. More precisely, we define a canonical $(p-s+1)$-sequence $i = (i^{(1)}, \ldots, i^{(p)})$ by $i^{(1)} = i^{(2)} = 1$ and $i^{(k)} = \tilde{i}^{(k-1)}$ for $3 \le k \le p$. One can easily check that i is a canonical $(p-s+1)$-sequence of length p, and there are exactly two edges with vertex α_2: an up edge $i^{(1)} \to \alpha_2$ and a down edge $\alpha_2 \to i^{(2)}$. Noting that $i^{(2)} = i^{(1)}$, the up edge and down edge coincide, but they do not coincide with other edges. Thus, we have $g(i, \alpha) \in \Delta_1(p, s; \alpha)$, which means that $\alpha \in \mathcal{C}_{s,p}^{(1)}$.

Case 2(b). If there exist $2 < k \le p$, such that $\alpha_k = \alpha_2$. Then we can split the sequence α into two subsequences α', α'', where $\alpha' = (\alpha_2, \ldots, \alpha_{k-1})$, and $\alpha'' = (\alpha_1, \alpha_{k+1}, \ldots, \alpha_p)$. One can use the argument of Case 1 to deduce that there are two canonical sequences i_1, i_2 of length $k-2$ and $p-k+1$ respectively, such that the graph $g(i_1, \alpha')$ and $g(i_2, \alpha'')$ satisfy the definition of Δ_1-graph. We shift the sequence i_1 by adding 1 to the value of each vertex, and denote by i_1' the sequence after shifting. We also shift the sequence i_2 by adding $|i_1|$ to all vertices that do not have value 1. We write i_2' for the sequence after shifting. Then one can glue the two sequences i_1' and i_2' using the argument in Case 1. More precisely, the canonical sequence $i = (i^{(1)}, \ldots, i^{(p)})$ can be defined by $i^{(1)} = 1$, $i^{(j)} = i_1'^{(j-1)}$ for $2 \le j \le k-1$, and $i^{(j)} = i_2'^{(j-k+1)}$ for $k \le j \le p$. The non-crossing of the sequence α ensure that the graph $g(i, \alpha) \in \Delta_1(p, s; \alpha)$. $\square$

In the following, we study the graph $g(i, \alpha)$ for non-crossing sequence α. We first introduce the conception of paired graph and single graph.

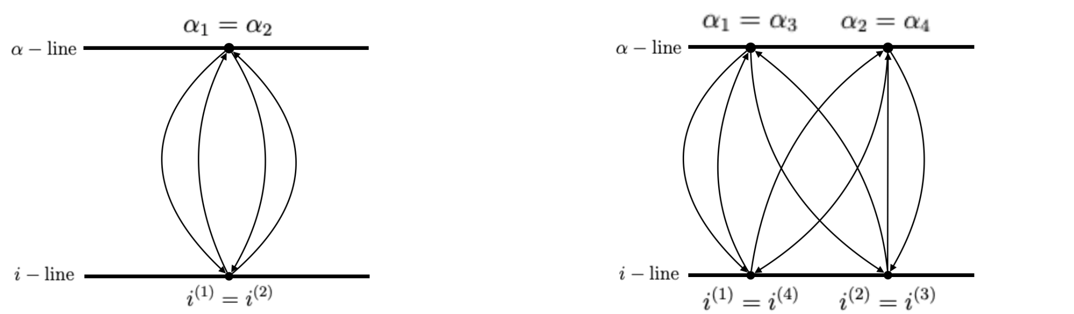
Definition 2. *Let α, i be two sequences. The $\Delta(p; \alpha)$-graph $g(i, \alpha)$ is called a* paired *graph if for any two vertices, between which the number of up edges equals to the number of down edges. The graph $g(i, \alpha)$ is called a* single *graph if there exist two vertices, such that difference of the number of up edges and down edges between the two vertices is exactly one.*

Remark 2.1. *For a $\Delta(p;\alpha)$-graph $g(i,\alpha)$, if one reduces the graph by removing an up edges with one of the coincident down edges at the same time (but keep the vertices), then a paired graph is the graph which can be reduced to a graph without edges, while a single graph is the graph that can be reduced to a graph with at least one single edge.*

Remark 2.2. *1. A $\Delta_1(p,s;\alpha)$-graph is always a paired graph. Figure 3 provides two examples of paired graphs that are not $\Delta_1(p,s;\alpha)$-graph.*

2. Single graphs exist for any sequence α. Figure 1 (a) is an example of single graph. Indeed, one only need to choose i to have distinct vertices.

3. There are $\Delta(p;\alpha)$-graphs which is neither a paired graph nor a single graph. See for example Figure 4 where the multiple edges in $g(i,\alpha)$ have the same orientation.



(a) paired graph with $p = 2$, $\alpha = i = (1,1)$.

(b) paired graph with $p = 4$, $\alpha = (1,2,1,2)$, $i = (1,2,2,1)$.

Figure 3

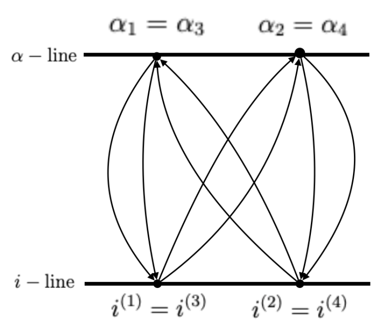


Figure 4: $g(i,\alpha)$ with $p = 4$ and $\alpha = i = (1,2,1,2)$.

Next, we establish the following proposition for the $\Delta(p;\alpha)$-graph for non-crossing sequence α.

Proposition 2.1. *For any $1 \le s \le p$, and $\alpha \in \mathcal{C}^{(1)}_{s,p}$, for any canonical sequence i of length p, the graph $g(i,\alpha)$ is either a paired graph or a single graph.*

In order to prove Proposition 2.1, we need to introduce the conception of *consecutive* down (resp. up) edges. Let α, i be two canonical sequences. For any two coincide down edges $\alpha_{j_1} \to i^{(j_1)}$ and $\alpha_{j_2} \to i^{(j_2)}$ with some $1 \le j_1 < j_2 \le p$, if all up edges $\{i^{(j)} \to \alpha_{j+1} : j_1 \le j < j_2\}$ between the two down edges do not coincide with them (without considering the orientation), then we call the two down edges $\alpha_{j_1} \to i^{(j_1)}$ and $\alpha_{j_2} \to i^{(j_2)}$ are *consecutive* down edges with *distance* $j_2 - j_1$. Similarly, for any two coincide up edges $i^{(j_1)} \to \alpha_{j_1+1}$ and $i^{(j_2)} \to \alpha_{j_2+1}$ with $1 \le j_1 < j_2 \le p$, if down edge $\alpha_j \to i^{(j)}$ does not coincide with them for any $j_1 + 1 \le j \le j_2$, then we call the two up edges *consecutive* up edges with *distance* $j_2 - j_1$. In the graph given by Figure 4, the two coincident down edges $\alpha_1 \to i^{(1)}$ and $\alpha_3 \to i^{(3)}$ are consecutive down edges with distance 2, while the two coincident up edges $i^{(1)} \to \alpha_2$ and $i^{(3)} \to \alpha_4$ are consecutive up edges with distance 2.

Proof. (of Proposition 2.1) We prove by contradiction. We fix the sequence $\alpha \in \mathcal{C}^{(1)}_{s,p}$. Assume that there exists a canonical sequence $i = (i^{(1)}, \dots, i^{(p)})$, such that the graph $g(i, \alpha)$ is neither a paired graph nor a single graph. By definition, there exist two vertices, such that the numbers of the up and down edges between the two vertices are different by at lease two. Thus, there are consecutive up edges or consecutive down edges. We choose the pair of consecutive edges with the smallest distance and consider the case that they are up edges and down edges separately. If there are more than one pair of consecutive edges with the smallest distance, we can choose any one of them.

Case 1. The pair of consecutive edges with smallest distance are down edges $\alpha_{j_1} \to i^{(j_1)}$ and $\alpha_{j_2} \to i^{(j_2)}$ with some $1 \le j_1 < j_2 \le p$.

We restrict out attention to the path $P : \alpha_{j_1} \to i^{(j_1)} \to \alpha_{j_1+1} \to \dots \to i^{(j_2-1)} \to \alpha_{j_2}$. For all vertices that coincides with $i^{(j_1)}$, we denote by A the collection of their neighbourhoods among the collection $\{\alpha_j : j_1 + 1 \le j \le j_2 - 1\}$ and E the corresponding collection of edges. We denote by B the collection $\{\alpha_1, \dots, \alpha_{j_1}, \alpha_{j_2}, \dots, \alpha_{p+1}\}$. We keep the multiplicity for coincide vertices (resp. edges) for A (resp. E).

Note that vertices in A do not coincide with α_{j_1} by the definition of consecutive down edges, and do not coincide with any vertex in B since α is non-crossing. Thus, A and B are disjoint. Since the vertex $i^{(j_1)}$ is not the endpoint of the path P, the numbers of the up edges and down edges within P associated with $i^{(j_1)}$ are the same. Noting that on the path P, the edge $\alpha_{j_1} \to i^{(j_1)}$ is the only edge associated with $i^{(j_1)}$ that are not in E, so the number of edges in E is odd. Hence, there exist the coincide edges in E consist of different number of up edges and down edges. If the difference of up and down coincident edges is exactly one, then the graph is a single graph, which is a contradiction. If the up and down coincident edges differ by at least two, then there is another pair of consecutive edges. This also leads to a contradiction since the consecutive down edges $\alpha_{j_1} \to i^{(j_1)}$ and $\alpha_{j_2} \to i^{(j_2)}$ should have the smallest distance.

Case 2. The pair of consecutive edges with smallest distance are up edges $i^{(j_1)} \to \alpha_{j_1+1}$ and $i^{(j_2)} \to \alpha_{j_2+1}$ with $1 \le j_1 < j_2 \le p$.

The argument is similar to the Case 1, and is sketched below. We consider the path $P' : \alpha_{j_1+1} \to \dots \to \alpha_{j_2} \to i^{(j_2)} \to \alpha_{j_2+1}$. For all vertices that coincides with $i^{(j_2)}$, we denote by A' the collection of their neighbourhoods among $\{\alpha_j : j_1 + 2 \le j \le j_2\}$ and E' the corresponding collection of edges. We also denote $B' = \{\alpha_1, \ldots, \alpha_{j_1+1}, \alpha_{j_2+1}, \ldots, \alpha_{p+1}\}$. Then by the definition of consecutive up edges and the fact that α is non-crossing, one can deduce that A' and B' are disjoint. Besides, by analyzing the neighbourhood of $i^{(j_2)}$ in the path P', one can deduce that the number of edges in E' is odd. This contradicts to either the condition that $g(i, \alpha)$ is not a single graph or the assumption that consecutive edges have distance at least $j_2 - j_1$. □

2.3 Paired graph

In this subsection, we study the paired graph, which contributes to the moments in Section 3. For graphs that are not single graph, we have the following proposition for the number of vertices.

Proposition 2.2. *For any $1 \le r, s \le p$, for any $\alpha \in \mathcal{C}_{s,p}$ and $i \in \mathcal{C}_{r,p}$, we have the following statements:*

1. *If $g(i, \alpha)$ is a paired graph, then $r + s \le p + 1$. The equality holds if and only if $g(i, \alpha)$ is a $\Delta_1(p, s; \alpha)$-graph.*

2. *If $g(i, \alpha)$ is neither a paired graph nor a single graph, then $r + s \le p$.*

Proof. If $g(i, \alpha)$ is not a single graph, then all edges must coincide with at least one other edges. If we remove the orientation and glue all the coincide edges, it results in non-directed connected graph with at most p edges and exactly $r + s$ vertices, which implies that $r + s \le p + 1$. The equality holds if and only if the resulting graph is a tree with exactly p edges. In this case, all edges in the graph $g(i, \alpha)$ must coincide with exactly one other edge. If there are two coincident edges that have the same orientation, then directed graph $g(i, \alpha)$ is disconnected, which is a contradiction. Thus, the equality only happens when the graph $g(i, \alpha)$ is a $\Delta_1(p, s; \alpha)$-graph. □

Next, we introduce the *Stirling number of the second kind* with the notation $S(n, k)$, which is defined as the number of ways to partition a set of n objects into k non-empty subsets. For non-crossing sequence α, the following proposition counts the number of paired graph associate to α.

Proposition 2.3. *For any $1 \le s \le p$ and any sequence $\alpha \in \mathcal{C}^{(1)}_{s,p}$, the number of sequence $i \in \mathcal{C}_{r,p}$ such that $g(i, \alpha)$ is a paired graph is $S(p + 1 - s, r)$ if $r \le p + 1 - s$, and is 0 if $r > p + 1 - s$.*

Proof. The case $r > p + 1 - s$ is straightforward from Proposition 2.2. In the following, we only consider the case $r \le p + 1 - s$. We fix a sequence $\alpha \in \mathcal{C}^{(1)}_{s,p}$.

Firstly, we will show that any paired graph $g(i, \alpha)$ can be transferred to a $\Delta_1(p, s; \alpha)$-graph by splitting the vertices in the sequence i. By definition, one can easily see that paired graphs may have more than one pair of up and down edges between two vertices, and may have cycles if the multiple edges are glued and orientation are removed. Thus, our strategy is to remove the multiple pairs of up and down edges in the first step, and then remove the cycles in the second step.

Step 1. For any two vertices v_1, v_2 in the paired graph $g(i, \alpha)$, we denote by m_{v_1,v_2} the number of the up edges between v_1 and v_2. We define

$$K(g(i, \alpha)) = \sum_{v_1,v_2} (m_{v_1,v_2} - 1)$$

where the sum $\sum_{v_1,v_2}$ is over all pairs of vertices (v_1, v_2) that are neighbourhood in $g(i, \alpha)$. One can easily check by definition that $K(g(i, \alpha)) = 0$ if and only if every edge in $g(i, \alpha)$ coincides with exactly one edge, and the two coincident edges have different orientation.

For the case $K(g(i, \alpha)) = 1$, there exists two vertices, between which there are two up edges and two down edges. We will split the corresponding i-vertex can into two vertices and resulting in a new i-sequence i', such that the the graph $g(i', \alpha)$ is a paired graph without coincident edges of the same orientation. The argument is similar to [7, Lemma 3.3], and is sketched below in two cases.

Case 1. If we scan the edges from α_1 to α_{p+1}, the first appearance of the four coincident edges is an down edges. In this case, the coincident edges are the jth down edge $\alpha_j \to i^{(j)}$, the lth down edge $\alpha_l \to i^{(l)}$, the j'th up edge $i^{(j')} \to \alpha_{j'+1}$ and the l'th up edge $i^{(l')} \to \alpha_{l'+1}$ for some $j < j' + 1 \leq l < l' + 1$. We split the vertex $i^{(l)}$ into two vertices $i^{(l,1)}$ and $i^{(l,2)}$. The edges from $\alpha_1 \to i^{(1)}$ to $i^{(l-1)} \to \alpha_l$ that connects $i^{(l)}$ are plotted to connect $i^{(l,1)}$, while the edges from $\alpha_l \to i^{(l)}$ to $i^{(p)} \to \alpha_{p+1}$ that connects $i^{(l)}$ are plotted to connect $i^{(l,2)}$. See Figure 5 below.

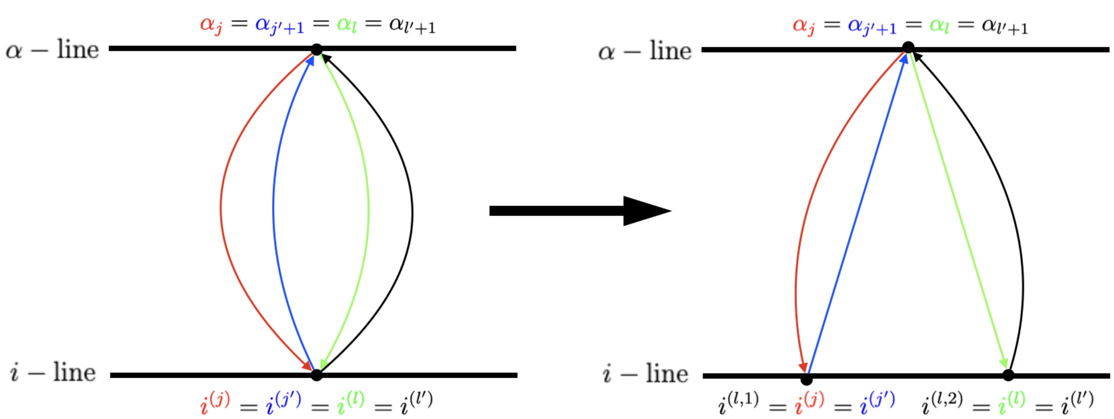


Figure 5: Case 1-Split an i vertex to cancel multiple pairs of edges.

Case 2. If we scan the edges starting from $\alpha_1 \to i^{(1)}$, the first appearance of the four coincident edges is an up edges. In this case, the coincident edges are the jth up edge $i^{(j)} \to \alpha_{j+1}$, the lth up edge $i^{(l)} \to \alpha_{l+1}$, the j'th down edge $\alpha_{j'} \to i^{(j')}$ and the l'th down edge $\alpha_{l'} \to i^{(l')}$ for some $j < j' \leq l < l'$. We split the vertex $i^{(l)}$ into two vertices $i^{(l,1)}$ and $i^{(l,2)}$. The edges from $\alpha_{j'} \to i^{(j')}$ to $i^{(l)} \to \alpha_{l+1}$ that connects $i^{(l)}$ are plotted to connect $i^{(l,2)}$, while the rest of the edges that connects $i^{(l)}$ are plotted to connect $i^{(l,1)}$. See Figure 6 below.

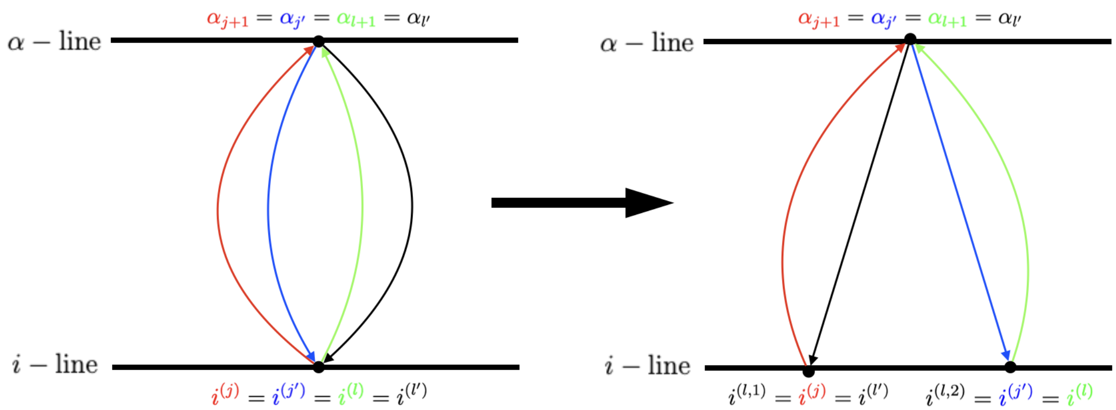


Figure 6: Case 2-Split an i vertex to cancel multiple pairs of edges.

In both cases, one could check that after splitting the vertex $i^{(l)}$, the graph is still connected, and is paired graph. Moreover, the number of edges between any two vertices is either 0 or 2, which implies that $K(g(i', \alpha)) = 0$.

One can use induction to show that there exists a sequence i', such that $K(g(i', \alpha)) = 0$ and the paired graph $g(i', \alpha)$ can be obtained from $g(i, \alpha)$ by splitting some of the vertices in i. Indeed, by scanning all edges starting from $\alpha_1 \to i^{(1)}$, we can find the first coincident directed edges. Then we can apply the argument above to split the i-vertex associated to the coincident directed edges into two i-vertices. We denote the resulting i-sequence by $\tilde{i}'$. Then we have $K(g(\tilde{i}', \alpha)) = K(g(i, \alpha)) - 1$. By the induction hypothesis, we can find a sequence i' by splitting $\tilde{i}'$, such that $K(g(i', \alpha)) = 0$. Moreover, the i-sequence i' can also be obtained by splitting the sequence i.

Step 2. Let $g(i', \alpha)$ be a paired graph such that $K(g(i', \alpha)) = 0$. Then every up edge coincides with exactly one down edge. Denote by $C(g(i', \alpha))$ the number of cycles when gluing all the pairs of coincident up edge and down edge and removing the orientation. By definition, $C(g(i', \alpha)) = 0$ if and only if $g(i', \alpha)$ is a $\Delta_1(p, s; \alpha)$-graph.

If $C(g(i', \alpha)) = 1$, then there is exactly one cycle when gluing the pair of coincident up edge and down edge. We will split one vertex in i' and denote by i'' the new i-sequence, such that $g(i'', \alpha)$ is still a paired graph without any cycle when gluing all pairs of coincident up edge and down edge, and $g(i'', \alpha)$ does not have coincident edges of the same orientation. That is, $g(i'', \alpha)$ is a $\Delta_1(p, s; \alpha)$-graph. The argument is similar to [7, Lemma 3.6], and is sketched below.

We scan the edges starting from $\alpha_1 \to i^{(1)}$, and find the first edge that results in a cycle without considering orientation and removing the multiplicity of the edges. Using the non-crossing property of α, one can show that this edge must be a down edge. We denote the down edge by $\alpha_j \to i^{(j)}$. We split the vertex $i^{(j)}$ into two vertices $i^{(j,1)}$ and $i^{(j,2)}$. The edges in the path $\alpha_1 \to i^{(1)} \to \ldots \to \alpha_j$ that connects $i^{(j)}$ are plotted to connect $i^{(j,1)}$, while the edges in the path $\alpha_j \to i^{(j)} \to \ldots \to i^{(p)} \to \alpha_{p+1}$ that connects $i^{(j)}$ are plotted to connect $i^{(j,2)}$. See Figure 7. Once could check that after splitting the vertex $i^{(j)}$, there is no multiple up edge or multiple down edge. Besides, there is no cycle without considering the orientation and gluing pairs of coincident up edge and down edge. Hence, if we denote by i'' the new i-sequence, then $K(g(i', \alpha)) = 0 = C(g(i', \alpha))$.

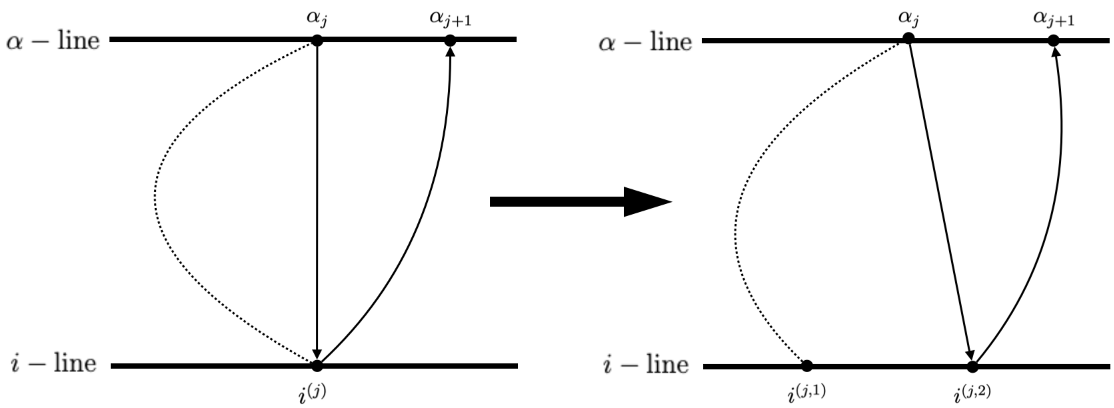


Figure 7: Split an i vertex to cancel cycle.

One can use induction to show that there exists a sequence i'', such that $C(g(i'', \alpha)) = K(g(i'', \alpha)) = 0$ and the paired graph $g(i'', \alpha)$ can be obtained from $g(i', \alpha)$ by splitting some of the vertices in i'. Indeed, we can scan all edges from $\alpha_1 \to i'^{(1)}$ and find the first edge which forms a cycle when gluing coincident edges and removing orientation. Then we can use the argument above to split one of the i-vertex in the cycle into two i-vertices. We denote the resulting i-sequence by $\tilde{i}''$. The splitting procedure will not lead to coincident up edges nor coincident down edges, nor new cycle when gluing all pairs of coincident edges. Thus, we have $C(g(\tilde{i}'', \alpha)) \leq C(g(i', \alpha)) - 1$ and $K(g(\tilde{i}'', \alpha)) = 0$. Then by induction hypothesis, we can split vertices in $\tilde{i}''$ to obtain i'', such that $g(i'', \alpha)$ is paired graph and $C(g(i'', \alpha)) = K(g(i'', \alpha)) = 0$. Moreover, the i-sequence i'' can also be obtained by splitting vertices in the sequence i'.

Therefore, joining the two steps above, we show that for any paired graph $g(i, \alpha)$, we can split vertices on i to obtain i'', such that $g(i'', \alpha)$ is a $\Delta_1(p, s; \alpha)$-graph.

Secondly, we will establish a bijective map from the set of all partitions of $[p + 1 - s]$ to the set of the canonical r-sequence that form a paired graph with α.

By Lemma 2.1, there exists a unique canonical $(p + 1 - s)$-sequence $i = (i^{(1)}, \ldots, i^{(p)})$, such that $g(i, \alpha)$ is a $\Delta_1(p, s; \alpha)$-graph. Let $\mathcal{P}(p + 1 - s)$ be the set of all partitions of $[p + 1 - s]$, and $\mathcal{P}(p + 1 - s, q)$ be the set of all partitions of $[p + 1 - s]$ with q blocks. For a partition $\pi \in \mathcal{P}(p + 1 - s, q)$ with blocks $V_1, V_2, \ldots, V_q$, without loss of generality, we assume that

$$\min\{a : a \in V_1\} < \ldots < \min\{a : a \in V_q\}.$$

We identify the partition π with the mapping $\pi : [p+1-s] \to [q]$ given by $\pi(a) = b$ if $\in V_b$. We abuse the notation for partition and the corresponding mapping. By the definition, one can easily check that π maps canonical sequence to canonical sequence. For fixed $\alpha \in \mathcal{C}^{(1)}_{s,p}$, let $\mathcal{P}(\alpha)$ be the set of all canonical sequence i' such that $g(i', \alpha)$ is a paired graph. We write $\mathcal{P}(\alpha, r)$ for the set of all canonical r-sequence i' in $\mathcal{P}(\alpha)$. We consider the following mapping:

$$\begin{array}{rcl}\Phi : \mathcal{P}(p+1-s) & \longrightarrow & \mathcal{P}(\alpha) \\ \pi & \longrightarrow & \pi(i),\end{array}$$

where $\pi(i) = (\pi(i^{(1)}), \ldots, \pi(i^{(p)}))$.

Note that for any $\pi \in \mathcal{P}(p+1-s)$, $g(\pi(i), \alpha)$ can be obtained from $g(i, \alpha)$ by gluing the i-vertices according to the partition π. Since $g(i, \alpha)$ is a paired graph, so is $g(\pi(i), \alpha)$, which implies that Φ is well-defined. As we have proved in the first part that any paired graph can be transferred to a $\Delta_1(p, s; \alpha)$-graph by appropriately splitting the vertices in the i-sequence, we can conclude that Φ is surjective. Moreover, for two different partitions $\pi_1, \pi_2 \in \mathcal{P}(p+1-s)$, the two canonical sequence $\pi_1(i)$ and $\pi_2(i)$ are different. Thus, Φ is injective. Therefore, Φ is bijective.

To conclude, we consider the following restriction of Φ:

$$\Phi|_r : \mathcal{P}(p+1-s, r) \quad \longrightarrow \quad \mathcal{P}(\alpha, r).$$

Since the bijectivity of $\Phi|_r$ inherites from Φ, by the definition of Stirling number of the second kind, we have

$$\#\mathcal{P}(\alpha, r) = \#\mathcal{P}(p+1-s, r) = S(p+1-s, r).$$

□

We end this subsection by collecting some properties of the Stirling number of the second kind. We refer the readers to [8] for more details.

Lemma 2.4. *1. We have $S(n,n) = S(n,1) = 1$ for $n \geq 1$. For $1 \leq k \leq n$, we have*

$$S(n,k) = \sum_{i=1}^{k} \frac{(-1)^{k-i} i^n}{i!(k-i)!}.$$

2. For positive integers $n \geq k > 1$, we have

$$S(n+1, k) = S(n, k-1) + kS(n,k).$$

3. For positive integers n, we have

$$\sum_{k=1}^{n} S(n,k) \cdot x(x-1)\ldots(x-k+1) = x^n.$$

3 Convergence of spectral moments

3.1 Proof of Theorem 1.1

We compute the moment

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

for any $p \in \mathbb{N}_+$. By convention, $\alpha_{p+1} = \alpha_1$. We have

$$\begin{aligned}
\frac{1}{n^k}\mathbb{E}\big[\mathrm{Tr}M_{n,k,m}^p\big] =& \frac{1}{n^k}\sum_{\alpha_1,\dots,\alpha_p=1}^{m}\left(\prod_{t=1}^{p}\tau_{\alpha_t}\right)\mathbb{E}\Big[\mathrm{Tr}\big(Y_{\alpha_1}^*Y_{\alpha_2}Y_{\alpha_2}^*\cdots Y_{\alpha_p}Y_{\alpha_p}^*Y_{\alpha_{p+1}}\big)\Big]\\
=& \frac{1}{n^k}\sum_{\alpha_1,\dots,\alpha_p=1}^{m}\left(\prod_{t=1}^{p}\tau_{\alpha_t}\right)\mathbb{E}\left[\prod_{l=1}^{k}\mathrm{Tr}\left(\left(\mathbf{y}_{\alpha_1}^{(l)}\right)^*\mathbf{y}_{\alpha_2}^{(l)}\left(\mathbf{y}_{\alpha_1}^{(l)}\right)^*\dots\mathbf{y}_{\alpha_{p+1}}^{(l)}\right)\right]\\
=& \frac{1}{n^k}\sum_{\alpha_1,\dots,\alpha_p=1}^{m}\left(\prod_{t=1}^{p}\tau_{\alpha_t}\right)\left(\mathbb{E}\left[\mathrm{Tr}\left(\left(\mathbf{y}_{\alpha_1}^{(1)}\right)^*\mathbf{y}_{\alpha_2}^{(1)}\left(\mathbf{y}_{\alpha_1}^{(1)}\right)^*\dots\mathbf{y}_{\alpha_{p+1}}^{(1)}\right)\right]\right)^k\\
=& \frac{1}{n^k}\sum_{\alpha_1,\dots,\alpha_p=1}^{m}\left(\prod_{t=1}^{p}\tau_{\alpha_t}\right)\left(\mathbb{E}\left[\sum_{i^{(1)},\dots,i^{(p)}=1}^{n}\prod_{t=1}^{p}\left(\overline{(\mathbf{y}_{\alpha_t}^{(1)})_{i^{(t)}}}(\mathbf{y}_{\alpha_{t+1}}^{(1)})_{i^{(t)}}\right)\right]\right)^k,
\end{aligned}\tag{3.1}$$

where we used the i.i.d. setting in the third equality.

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

$$E(i,\alpha) = \mathbb{E}\left[\prod_{t=1}^{p}\left(\overline{(\mathbf{y}_{\alpha_t}^{(1)})_{i^{(t)}}}(\mathbf{y}_{\alpha_{t+1}}^{(1)})_{i^{(t)}}\right)\right]. \tag{3.2}$$

By the i.i.d. setting, $E(i,\alpha) = E(i',\alpha')$ if the two sequences i and α are equivalent to i' and α', respectively. By (3.1) and (2.1), we have

$$\begin{aligned}
\frac{1}{n^k}\mathbb{E}\big[\mathrm{Tr}M_{n,k,m}^p\big] =& \frac{1}{n^k}\sum_{s=1}^{p}\sum_{\alpha\in\mathcal{J}_{s,p}(m)}\left(\prod_{t=1}^{p}\tau_{\alpha_t}\right)\left(\sum_{r=1}^{p}\sum_{i\in\mathcal{J}_{r,p}(n)}E(i,\alpha)\right)^k\\
=& \frac{1}{n^k}\sum_{s=1}^{p}\sum_{\alpha\in\mathcal{C}_{s,p}}\left(\sum_{\varphi\in\mathcal{I}_{s,m}}\prod_{t=1}^{p}\tau_{\varphi(\alpha_t)}\right)\left(\sum_{r=1}^{p}n\cdots(n-r+1)\sum_{i\in\mathcal{C}_{r,p}}E(i,\alpha)\right)^k\\
:=& I_1 + I_2,
\end{aligned}\tag{3.3}$$

where

$$I_1 = \frac{1}{n^k}\sum_{s=1}^{p}\sum_{\alpha\in\mathcal{C}_{s,p}^{(1)}}\left(\sum_{\varphi\in\mathcal{I}_{s,m}}\prod_{t=1}^{p}\tau_{\varphi(\alpha_t)}\right)\left(\sum_{r=1}^{p}n\cdots(n-r+1)\sum_{i\in\mathcal{C}_{r,p}}E(i,\alpha)\right)^k,$$

$$I_2 =\frac{1}{n^k}\sum_{s=1}^{p}\sum_{\alpha\in\mathcal{C}_{s,p}\setminus\mathcal{C}_{s,p}^{(1)}}\left(\sum_{\varphi\in\mathcal{I}_{s,m}}\prod_{t=1}^{p}\tau_{\varphi(\alpha_t)}\right)\left(\sum_{r=1}^{p}n\cdots(n-r+1)\sum_{i\in\mathcal{C}_{r,p}}E(i,\alpha)\right)^k.$$

Note that the component of the base vector satisfies

$$\left(\mathbf{y}_\beta^{(l)}\right)_i\overline{\left(\mathbf{y}_\beta^{(l)}\right)_i}=\frac{1}{n},\quad \left|\mathbb{E}\left[\left(\left(\mathbf{y}_\beta^{(l)}\right)_i\right)^p\right]\right|\le\frac{1}{n^{p/2}}. \tag{3.4}$$

Recall the definition of paired graph and single graph in Definition 2. For any sequence $\alpha\in\mathcal{C}_{s,p}$ and $i\in\mathcal{C}_{r,p}$, we have

$$\begin{cases} E(i,\alpha)=n^{-p}, & g(i,\alpha) \text{ is a paired graph,}\\ E(i,\alpha)=0, & g(i,\alpha) \text{ is a single graph,}\\ |E(i,\alpha)|\le n^{-p}, & g(i,\alpha) \text{ otherwise.}\end{cases} \tag{3.5}$$

Firstly, we deal with I_1. For any $\alpha\in\mathcal{C}_{s,p}^{(1)}$, by Proposition 2.1, formula (3.5) and Proposition 2.3, we have

$$\begin{aligned}\sum_{i\in\mathcal{C}_{r,p}}E(i,\alpha)=&n^{-p}\cdot\#\{i\in\mathcal{C}_{r,p}:g(i,\alpha)\text{ is paired graph}\}\\ =&\begin{cases} n^{-p}\cdot S(p+1-s,r), & r\le p+1-s,\\ 0, & r>p+1-s,\end{cases}\end{aligned}$$

where we use the notation $\#S$ for the number of elements in the set S. Thus, by Lemma 2.4, we have

$$\sum_{r=1}^{p}n\cdots(n-r+1)\sum_{i\in\mathcal{C}_{r,p}}E(i,\alpha)=n^{-p}\sum_{r=1}^{p+1-s}n\cdots(n-r+1)\cdot S(p+1-s,r)=n^{1-s}.$$

Thus, we have

$$I_1=\sum_{s=1}^{p}\left(\frac{m}{n^k}\right)^s\sum_{\alpha\in\mathcal{C}_{s,p}^{(1)}}\left(\frac{1}{m^s}\sum_{\varphi\in\mathcal{I}_{s,m}}\prod_{t=1}^{p}\tau_{\varphi(\alpha_t)}\right). \tag{3.6}$$

Next, we deal with I_2. For any $\alpha\in\mathcal{C}_{s,p}\setminus\mathcal{C}_{s,p}^{(1)}$, by Lemma 2.1 and Proposition 2.2, if the graph $g(i,\alpha)$ is not a single graph for $i\in\mathcal{C}_{r,p}$, then $r+s\le p$. Hence, by (3.5), we have

$$\left|\sum_{r=1}^{p}n\cdots(n-r+1)\sum_{i\in\mathcal{C}_{r,p}}E(i,\alpha)\right|\le\sum_{r=1}^{p-s}n^{-p+r}\cdot\#\{i\in\mathcal{C}_{r,p}:g(i,\alpha)\text{ is not a single graph}\}.$$

Thus, we have

$$
\begin{aligned}
|I_2| \leq & \frac{1}{n^k} \sum_{s=1}^{p} \sum_{\alpha \in \mathcal{C}_{s,p} \setminus \mathcal{C}_{s,p}^{(1)}} \left| \sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^{p} \tau_{\varphi(\alpha_t)} \right| \left| \sum_{r=1}^{p} n \cdots (n-r+1) \sum_{i \in \mathcal{C}_{r,p}} E(i,\alpha) \right|^k \\
= & \sum_{s=1}^{p} \left(\frac{m}{n^k} \right)^s \sum_{\alpha \in \mathcal{C}_{s,p} \setminus \mathcal{C}_{s,p}^{(1)}} \left| \frac{1}{m^s} \sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^{p} \tau_{\varphi(\alpha_t)} \right| \\
& \times \left| \sum_{r=1}^{p-s} n^{-p+r+s-1} \cdot \#\{ i \in \mathcal{C}_{r,p} : g(i,\alpha) \text{ is not a single graph} \} \right|^k . \qquad (3.7)
\end{aligned}
$$

By the assumption (1.4), we have the following convergence:

$$
\frac{1}{m^s} \sum_{\varphi \in \mathcal{I}_{s,m}} \prod_{t=1}^{p} \tau_{\varphi(\alpha_t)} \to \prod_{t=1}^{s} m^{(\tau)}_{\deg_t(\alpha)}, \quad m \to \infty.
$$

Hence, under the limiting setting (1.3), when $n, k \to \infty$, we can deduce from (3.6) and (3.7) that

$$
I_1 \to \sum_{s=1}^{p} c^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\prod_{t=1}^{s} m^{(\tau)}_{\deg_t(\alpha)} \right), \quad I_2 \to 0. \qquad (3.8)
$$

Therefore, the proof is concluded by taking limit $n, k \to \infty$ in (3.3) and using (3.8).

3.2 Proof of Theorem 1.2

For any $p \in \mathbb{N}$, for $k \geq 2$, we compute the variance

$$
\mathrm{Var}\left(\frac{1}{n^k} \mathrm{Tr} M^p_{n,k,m} \right).
$$

The idea is similar to [7], and is sketched below. By the computation of [7, Section 3.2], we have

$$
\begin{aligned}
\mathrm{Var}\left(\frac{1}{n^k} \mathrm{Tr} M^p_{n,k,m} \right) = & \frac{1}{n^{2k}} \sum_{\substack{\alpha,\beta \in [m]^p \\ \alpha \cap \beta \neq \emptyset}} \left(\prod_{t=1}^{p} \tau_{\alpha_t} \tau_{\beta_t} \right) \\
& \times \left[\left(\sum_{i,j \in [n]^p} E'(i,\alpha; j,\beta) \right)^k - \left(\sum_{i,j \in [n]^p} E(i,\alpha) E(j,\beta) \right)^k \right],
\end{aligned}
$$

where $E(\cdot,\cdot)$ is given in (3.5), and $E'(i,\alpha;j,\beta)$ is defined by

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

Next, we join the two graphs $g(i,\alpha)$ and $g(j,\beta)$ together and keep the coincident edges. We denote by $g(i,\alpha)\cup g(j,\beta)$ the resulting graph. If there is an edge in the graph $g(i,\alpha)\cup g(j,\beta)$ that does not coincide with any other edges, then this edge must belong to $g(i,\alpha)$ or $g(j,\beta)$, which implies

$$E'(i,\alpha;j,\beta)=E(i,\alpha)E(j,\beta)=0.$$

Thus, we only need to consider the indices such that all edges in $g(i,\alpha)\cup g(j,\beta)$ coincide with other edges. Noting that $\alpha\cap\beta\neq\emptyset$, the graph $g(i,\alpha)\cup g(j,\beta)$ is connected with $4p$ edges. Hence, if we remove orientation and glue coincident edges for the graph $g(i,\alpha)\cup g(j,\beta)$, it results in a non-directed connected graph with at most $2p$ edges, which implies that

$$|(\alpha,\beta)|+|(i,j)|\leq 2p+1.$$

Hence, we have

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{n^k}\mathrm{Tr}M_{n,k,m}^p\right)=&\frac{1}{n^{2k}}\sum_{s=1}^{2p}\sum_{\substack{(\alpha,\beta)\in\mathcal{C}_{s,2p}\\ \alpha\cap\beta\neq\emptyset}}\left(\sum_{\varphi\in\mathcal{I}_{s,m}}\prod_{t=1}^{p}\tau_{\varphi(\alpha_t)}\tau_{\varphi(\beta_t)}\right)\\ &\times\left[\left(\sum_{r=1}^{2p+1-s}n\ldots(n-r+1)\sum_{(i,j)\in\mathcal{C}_{r,2p}}E'(i,\alpha;j,\beta)\right)^k\right.\\ &\left.-\left(\sum_{r=1}^{2p+1-s}n\ldots(n-r+1)\sum_{(i,j)\in\mathcal{C}_{r,2p}}E(i,\alpha)E(j,\beta)\right)^k\right].\end{aligned}$$

By (3.4), for any sequence α,β,i,j, we have

$$|E'(i,\alpha;j,\beta)|,\ |E(i,\alpha)E(j,\beta)|\leq n^{-2p},$$

which implies that

$$\begin{aligned}\max\Bigg\{&\left|\sum_{r=1}^{2p+1-s}n\ldots(n-r+1)\sum_{(i,j)\in\mathcal{C}_{r,2p}}E(i,\alpha)E(j,\beta)\right|,\\ &\left|\sum_{r=1}^{2p+1-s}n\ldots(n-r+1)\sum_{(i,j)\in\mathcal{C}_{r,2p}}E'(i,\alpha;j,\beta)\right|\Bigg\}\end{aligned}$$

$$\leq \sum_{r=1}^{2p+1-s} n^{r-2p} \sum_{(i,j)\in\mathcal{C}_{r,2p}} 1 \leq C_p n^{1-s}(1+o_n(1)),$$

where C_p is a positive number that only depends on p, and $o_n(1)$ is a quantity that tends to 0 as $n \to \infty$. Thus, we have

$$\begin{aligned}
\operatorname{Var}\left(\frac{1}{n^k}\operatorname{Tr}M_{n,k,m}^p\right) \leq& \frac{2C_p^k}{n^k}\sum_{s=1}^{2p}\sum_{\substack{(\alpha,\beta)\in\mathcal{C}_{s,2p}\\ \alpha\cap\beta\neq\emptyset}}\left(\sum_{\varphi\in\mathcal{I}_{s,m}}\prod_{t=1}^{p}\tau_{\varphi(\alpha_t)}\tau_{\varphi(\beta_t)}\right)n^{-ks}(1+o_n(1))^k\\
=& \frac{2C_p^k}{n^k}\sum_{s=1}^{2p}\left(\frac{m}{n^k}\right)^s\sum_{\substack{(\alpha,\beta)\in\mathcal{C}_{s,2p}\\ \alpha\cap\beta\neq\emptyset}}\left(\sum_{\varphi\in\mathcal{I}_{s,m}}\prod_{t=1}^{s}\frac{1}{m}\tau_{\varphi(t)}^{\deg_t(\alpha)+\deg_t(\beta)}\right)(1+o_n(1))^k.
\end{aligned}$$

By assumption (1.4), we have

$$\sum_{\varphi\in\mathcal{I}_{s,m}}\prod_{t=1}^{s}\frac{1}{m}\tau_{\varphi(t)}^{\deg_t(\alpha)+\deg_t(\beta)} \to \prod_{t=1}^{s} m_{\deg_t(\alpha)+\deg_t(\beta)}^{(\tau)}, \quad m\to\infty.$$

Together with (1.3), we have

$$\operatorname{Var}\left(\frac{1}{n^k}\operatorname{Tr}M_{n,k,m}^p\right) \leq \frac{2C_p^k}{n^k}\sum_{s=1}^{2p}c^s\sum_{\substack{(\alpha,\beta)\in\mathcal{C}_{s,2p}\\ \alpha\cap\beta\neq\emptyset}}\left(\prod_{t=1}^{s} m_{\deg_t(\alpha)+\deg_t(\beta)}^{(\tau)}\right)(1+o_n(1))^{k+1+s}.$$

Therefore, for $k\geq 2$, we have

$$\sum_{n\geq 2}\operatorname{Var}\left(\frac{1}{n^k}\operatorname{Tr}M_{n,k,m}^p\right) < +\infty.$$

The proof is concluded by Borel-Cantelli's Lemma, noting that $k=k(n)$ tends to infinity as $n\to\infty$.

3.3 Proof of Corollary 1.1

We start with the uniqueness of μ. We have

$$\begin{aligned}
\left|\sum_{s=1}^{p}c^s\sum_{\alpha\in\mathcal{C}_{s,p}^{(1)}}\left(\prod_{t=1}^{s}m_{\deg_t(\alpha)}^{(\tau)}\right)\right| \leq& \sum_{s=1}^{p}c^s\sum_{\alpha\in\mathcal{C}_{s,p}^{(1)}}\left(\prod_{t=1}^{s}A^{\deg_t(\alpha)}\deg_t(\alpha)^{\deg_t(\alpha)}\right)\\
\leq& \sum_{s=1}^{p}c^s\sum_{\alpha\in\mathcal{C}_{s,p}^{(1)}}A^{\sum_{t=1}^{s}\deg_t(\alpha)}\left(\sum_{s=1}^{p}\deg_t(\alpha)\right)^{\sum_{s=1}^{p}\deg_t(\alpha)}.
\end{aligned}$$

By definition (2.2), for $\alpha \in \mathcal{C}_{s,p}$, we have $\sum_{t=1}^{s} \deg_t(\alpha) = p$. Hence, by Lemma 2.1, we have

$$\left| \sum_{s=1}^{p} c^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\prod_{t=1}^{s} m_{\deg_t(\alpha)}^{(\tau)} \right) \right| \leq A^p p^p \sum_{s=1}^{p} c^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} 1 = A^p p^p \sum_{s=1}^{p} \frac{c^s}{p} \binom{p}{s-1} \binom{p}{s}.$$

Using the inequality $\binom{p}{s} \leq 2^p$ for all $0 \leq s \leq p$, we have

$$\left| \sum_{s=1}^{p} c^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\prod_{t=1}^{s} m_{\deg_t(\alpha)}^{(\tau)} \right) \right| \leq 4^p A^p p^p \sum_{s=1}^{p} \frac{c^s}{p} \leq 4^p A^p (1+c)^p p^p.$$

Hence,

$$\sum_{p=1}^{\infty} \left| \sum_{s=1}^{p} c^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} \left(\prod_{t=1}^{s} m_{\deg_t(\alpha)}^{(\tau)} \right) \right|^{-1/p} \geq \sum_{p=1}^{\infty} \left(4^p A^p (c+1)^p p^p\right)^{-1/p} = \sum_{p=1}^{\infty} \frac{1}{4A(c+1)p} = +\infty.$$

Thus, the Carleman's condition is satisfied, which implies that there exists a unique probability measure μ whose moments are given by (1.5).

The uniqueness of the probability measure μ corresponding to the moments in (1.5) guarantees the almost sure convergence of the ESD of $M_{n,k,m}$ towards μ.

If $\tau_\alpha = 1$ for all $1 \leq \alpha \leq m$, then the condition (1.4) holds with $m_q^{(\tau)} = 1$ for all $q \in \mathbb{N}$. In this case, the moment sequence (1.5) for μ becomes

$$\int_{\mathbb{R}} x^p \mu(dx) = \sum_{s=1}^{p} c^s \sum_{\alpha \in \mathcal{C}_{s,p}^{(1)}} 1 = \sum_{s=1}^{p} \frac{c^s}{p} \binom{p}{s-1} \binom{p}{s}, \quad \forall p \in \mathbb{N}_+,$$

where we use Lemma 2.1 in the last equality. By [3, Lemma 3.1], this moment sequence coincides with the moment sequence of the Marčenko-Pastur law (1.2). Therefore, the uniqueness of μ implies that μ is exactly the Marčenko-Pastur law (1.2).

Acknowledgments

The author gratefully acknowledges the financial support of ERC Consolidator Grant 815703 "STAM- FORD: Statistical Methods for High Dimensional Diffusions". Besides, the author would like to acknowledge Tiefeng Jiang, Felix Parraud, Kevin Schnelli, Jianfeng Yao for the discussion and helpful suggestions.

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] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010. MR2760897

[3] Zhidong Bai and Jack W. Silverstein, *Spectral analysis of large dimensional random matrices*, Second Edition, Springer Series in Statistics, Springer, New York, 2010. MR2567175

[4] Zhidong Bai and Wang Zhou, *Large sample covariance matrices without independence structures in columns*, Statist. Sinica **18** (2008), no. 2, 425–442.

[5] Jennifer Bryson, Roman Vershynin, and Hongkai Zhao, *Marchenko-Pastur law with relaxed independence conditions*, Random Matrices Theory Appl. **10** (2021), no. 4, Paper No. 2150040, 28. MR4379545

[6] Benoît Collins, Ion Nechita, and Karol Życzkowski, *Area law for random graph states*, J. Phys. A **46** (2013), no. 30, 305302, 18. MR3083280

[7] Benoît Collins, Jianfeng Yao, and Wangjun Yuan, *On spectral distribution of sample covariance matrices from large dimensional and large k-fold tensor products*, Electron. J. Probab. **27** (2022), Paper No. 102, 18.

[8] Ralph P. Grimaldi, *Discrete and combinatorial mathematics*, Fifth Edition, An Applied Introduction, Pearson, 2003.

[9] 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

[10] 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.

[11] Vladimir A. Marchenko and Leonid A. Pastur, *Distribution of eigenvalues for some sets of random matrices*, Math. USSR-Sb **1** (1967), 457–483.

[12] 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.

[13] 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.

[14] Pavel Yaskov, *Necessary and sufficient conditions for the Marchenko-Pastur theorem*, Electron. Commun. Probab. **21** (2016), Paper No. 73, 8. MR3568347

[15] ———, *Marchenko-pastur law for a random tensor model*, Electron. Commun. Probab. **28** (2023), 1–17.

[16] Karol Życzkowski, Karol A. Penson, Ion Nechita, and Benoît Collins, *Generating random density matrices*, J. Math. Phys. **52** (2011), no. 6, 062201, 20. MR2841746