Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2016, Vol. 52, No. 4, 1641–1666
DOI: 10.1214/15-AIHP693

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Moment approach for singular values distribution of a large auto-covariance matrix

Qinwen Wang[a] and Jianfeng Yao[b]

[a]*Department of Mathematics, Zhejiang University, Hangzhou 310027, China. E-mail: wqw8813@gmail.com*
[b]*Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam, Hong Kong. E-mail: jeffyao@hku.hk*

Received 28 March 2015; revised 3 June 2015; accepted 15 June 2015

Abstract. Let $(\varepsilon_t)_{t>0}$ be a sequence of independent real random vectors of p-dimension and let $X_T = \sum_{t=s+1}^{s+T} \varepsilon_t \varepsilon_{t-s}^* / T$ be the lag-s (s is a fixed positive integer) auto-covariance matrix of ε_t. Since X_T is not symmetric, we consider its singular values, which are the square roots of the eigenvalues of $X_T X_T^*$. Using the method of moments, we are able to investigate the limiting behaviors of the eigenvalues of $X_T X_T^*$ in two aspects. First, we show that the empirical spectral distribution of its eigenvalues converges to a nonrandom limit F, which is a result previously developed in (*J. Multivariate Anal.* **137** (2015) 119–140) using the Stieltjes transform method. Second, we establish the convergence of its largest eigenvalue to the right edge of F.

Résumé. Soit $(\varepsilon_t)_{t>0}$ une suite de vecteurs aléatoires indépendants de $\mathbb{R}^p$ et $X_T = \sum_{t=s+1}^{s+t} \varepsilon_t \varepsilon_{t-s}^* / T$ la matrice d'autocovariance empirique d'ordre s de la suite (s est un ordre fixé). Comme X_T n'est pas symétrique, nous considérons ses valeurs singulières, c'est-à-dire les racines carrées des valeurs propres de la matrice aléatoire $X_T X_T^*$. En utilisant la méthode des moments, nous établissons les propriétés limites de ces valeurs singulières dans deux directions. D'abord, nous démontrons que leur distribution empirique converge vers une limite déterministe F, retrouvant ainsi un résultat établi dans (*J. Multivariate Anal.* **137** (2015) 119–140) par la méthode de la transformée de Stieltjes. Ensuite, nous montrons que la plus grande de ces valeurs singulières converge vers le point extrémal du support de F.

MSC: 15A52; 60F15

Keywords: Auto-covariance matrix; Singular values; Limiting spectral distribution; Stieltjes transform; Largest eigenvalue; Moment method

1. Introduction

Let $(\varepsilon_t)_{t>0}$ be a sequence of independent real random vectors of p-dimension and let $X_T = \sum_{t=s+1}^{s+T} \varepsilon_t \varepsilon_{t-s}^* / T$ be the lag-s (s is a fixed positive integer) auto-covariance matrix of ε_t. The motivation of the above set up is due to the study of dynamic factor model, see [7]. Set

$$x_t = \Lambda f_t + \varepsilon_t + \mu, \tag{1.1}$$

where x_t is a p-dimensional sequence observed at time t, $\{f_t\}$ a sequence of m-dimensional "latent factor" ($m \ll p$) uncorrelated with the error process $\{\varepsilon_t\}$ and $\mu \in \mathbb{R}^p$ is the general mean. Therefore, the lag-s auto-covariance matrix of the time series x_t can be considered as a finite rank (rank m) perturbation of the lag-s auto-covariance matrix of ε_t. Therefore, the first step is to study the base component, which is the lag-s auto-covariance matrix of the error term. Besides, we are considering the random matrix framework, where the dimension p and the sample size T both tend to infinity with their ratio converging to a constant: $\lim p/T \to y > 0$.

One of the main problems in random matrix theory is to investigate the convergence of the sequence of *empirical spectral distribution* $\{F^{A_n}\}$ for a given sequence of symmetric or Hermitian random matrices $\{A_n\}$, where

$$F^{A_n}(x) := \frac{1}{p}\sum_{j=1}^{p}\delta_{l_j},$$

l_j are the eigenvalues of A_n. The limit distribution F, which is usually nonrandom, is called the *limiting spectral distribution* (LSD) of the sequence $\{A_n\}$. The study of spectral analysis of large dimensional random matrices dates back to the Wigner's famous semicircular law [18] for Wigner matrix, which is further extended in various aspects: Marčenko–Pastur (M–P) law [10] for large dimensional sample covariance matrix; and circular law for complex random matrix [4]. Another aspect is the bound on extreme eigenvalues. The literature dates back to [3], who proved the almost sure convergence of the largest eigenvalue of a sample covariance matrix under however some moment restrictions, which is later improved by [19]. For Wigner matrix, [2] found the sufficient and necessary condition for the almost sure convergence of its largest eigenvalue. Vu [15] presented an upper bound for the spectral norm of symmetric random matrices with independent entries and [11] derived the lower bound. Vershynin [14] studied the sharp upper bound of the spectral norm of products of random and deterministic matrices, which behave similarly to random matrices with independent entries, etc.

Notice that lag-0 auto-covariance matrix of ε_t reduces to the standard sample covariance matrix $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t^*$ and its property in large-dimension has been well developed in the literature. In contrast, very little is known for the lag-s auto-covariance matrix X_T. Recent related work include [8,9] and [6] for the LSD of the symmetrized auto-covariance matrix and [16] for its exact separation, which also ensures the convergence of its largest eigenvalue.

Since X_T is not symmetric, its singular values are the square roots of the p nonnegative eigenvalues of

$$A_T := X_T X_T^*. \tag{1.2}$$

Therefore, the main purpose of this paper is on the limiting behaviors of the eigenvalues of A_T. First, the LSD of A_T has been found in [8] using the method of Stieltjes transform. However, no results on the largest eigenvalue of A_T has been so far found. The main contribution of the paper is using moment method to prove the convergence of this largest eigenvalue to the right edge of the LSD under an appropriate moment condition. As a by product of the moment approach, we provide a new proof of the convergence of ESD of A_T to its LSD. A distinctive feature here is that the matrix A_T can be considered as the product of four matrices involving ε_t, new methodology is needed with respect to the existing literature on moment method in random matrix theory. In particular, we provide in Section 5 some complex recursion formulas related to enumeration of a particular family of "walk paths" on nonnegative integers, which further leads to our moment result, and these formulas may be of independent interest.

The rest of the paper is organized as follows. Preliminary introduction on the related graph theory is provided in Section 2. Section 3 derives the exact moment formula for the limiting spectral distribution of A_T using graph theory, which further leads to the expression of its corresponding Stieltjes transform. Section 4 gives details of the convergence of the largest eigenvalue of A_T. In Section 5, we provide some techniques to derive a system of recursion formulas for two families of "walk paths", which further leads to the limiting moments in Section 3.

2. Some graph theory

In order to enumerate the moments of the LSD of A_T by moment method, we need some information from graph theory. The concepts and notations are close to those used in [1].

For a pair of vectors of indexes $i = (i_1, \dots, i_{2k})$ $(1 \le i_l \le T, l \le 2k)$ and $j = (j_1, \dots, j_{2k})$ $(1 \le j_l \le p, l \le 2k)$, construct a graph $Q(i, j)$ in the following way. Draw two parallel lines, referred to as the I-line and J-line. Plot $i_1, \dots, i_{2k}$ on the I-line and $j_1, \dots, j_{2k}$ on the J-line, called the I-vertices and J-vertices, respectively. Draw k down edges from i_{2u-1} to j_{2u-1}, k down edges from $i_{2u} + s$ to j_{2u}, k up edges from j_{2u-1} to i_{2u}, k up edges from j_{2u} to $i_{2u+1} + s$ (all these up and down edges are called vertical edges) and k horizontal edges from i_{2u} to $i_{2u} + s$, k horizontal edges from $i_{2u-1} + s$ to i_{2u-1} (with the convention that $i_{2k+1} = i_1$), where all the u's are in the region: $1 \le u \le k$. An example of a Q graph with $k = 3$ is shown in Figure 1.

Definition 2.1. *The subgraph of all I-vertices is called the* roof *of Q and is denoted by $H(Q)$ (see the subgraph inside the dashed line in Figure 1 for illustration of $H(Q)$).*

Definition 2.2. *The M-*minor or pillar *of Q is defined as the* minor *of Q by contracting all horizontal edges, which means that all horizontal edges are removed from Q and all I-vertices connected through horizontal edges are glued together. We denote the M-*minor or pillar *of Q by $M(Q)$ (see Figure 2).*

Definition 2.3. *For a given $M(Q)$, glue all coincident vertical edges; namely, we regard all vertical edges with a common I-vertex and J-vertex as one edge. Then we get an undirectional connected graph. We call the resulting graph the* base *of the graph Q, and denote it by $B(Q)$ (see Figure 3).*

Definition 2.4. *For a vertical edge e of $B(Q)$, the number of up (down) vertical edges of Q coincident with e is called the* up (down) multiplicity *of e.*

Definition 2.5. *The* degree *of a vertex i_l is the number of edges incident to this vertex.*

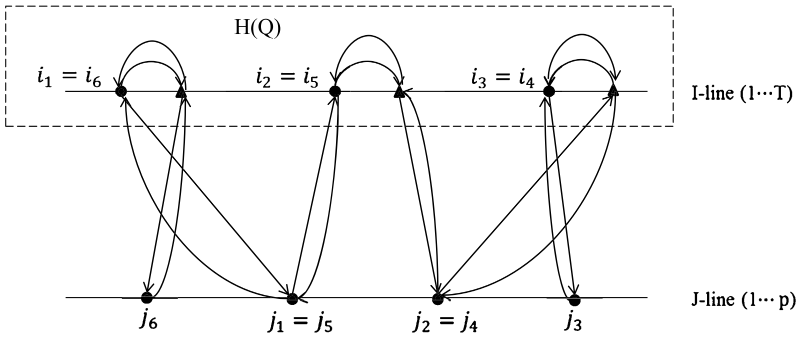

Fig. 1. An example of a Q graph with $k = 3$.

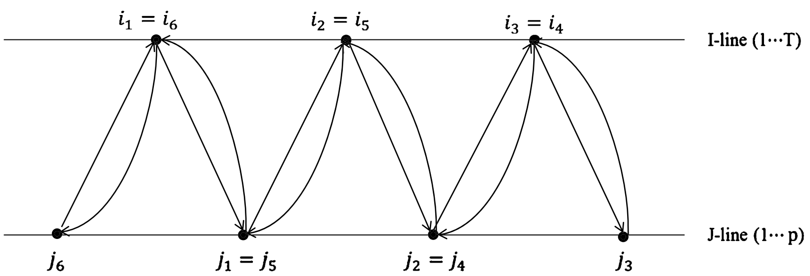

Fig. 2. The M-minor or pillar $M(Q)$ of the graph Q in Figure 1.

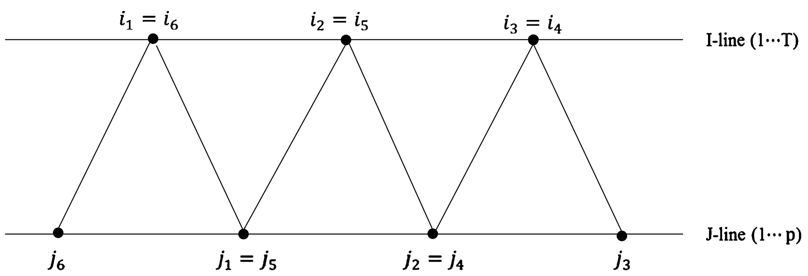

Fig. 3. The base $B(Q)$ of the graph Q in Figure 1.

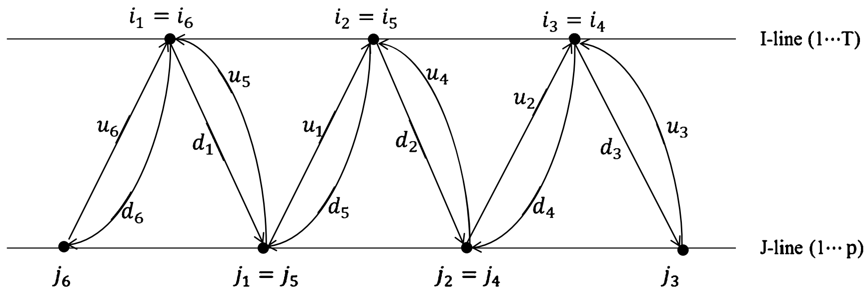

Fig. 4. Characteristic sequence associated with the pillar $M(Q)$ in Figure 2.

Definition 2.6. *For a cycle in $M(Q)$ from i_1 to j_1, then from j_1 to i_2, and so forth until we finally return to i_1 from j_{2k}. On each leg of this journey, say from i_l to j_l, if j_l has not been visited before, i.e. for all $k < l$, we have $j_k \neq j_l$, then this edge is called an* innovation. *An up (down)* innovation *e is an up (down) vertical edge that is an innovation.*

Definition 2.7. *Two graphs are said to be* isomorphic *if one becomes the other by a suitable permutation on $(1, \dots, T)$ and a suitable permutation on $(1, \dots, p)$.*

Definition 2.8. *Define a* characteristic sequence *as $(d_1 u_1 \cdots d_{2k} u_{2k})$, where $\{u_1, \dots, u_{2k}\}$ and $\{d_1, \dots, d_{2k}\}$ are associated with the $2k$ up edges and $2k$ down edges of a pillar $M(Q)$ according to the following rule:*

$$u_l = \begin{cases} 1, & \text{the } l\text{th up edge is an up innovation}, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$d_l = \begin{cases} 0, & \text{the } l\text{th down edge is an down innovation}, \\ -1, & \text{otherwise}. \end{cases}$$

An example of the characteristic sequence associated with the pillar in Figure 2 is given in Figure 4 with

$$\{u_1, u_2, u_3, u_4, u_5, u_6\} = \{1, 1, 0, 0, 0, 0\},$$
$$\{d_1, d_2, d_3, d_4, d_5, d_6\} = \{0, 0, 0, -1, -1, 0\};$$

that is, the corresponding characteristic sequence is $(0\ 1\ 0\ 1\ 0\ 0\ -1\ 0\ -1\ 0\ 0\ 0)$. Conversely, it can be verified that any characteristic sequence $(d_1 u_1 \cdots d_{2k} u_{2k})$ uniquely defines a pillar $M(Q)$.

3. LSD of A_T using moment method

The problem of showing the convergence of the ESD of A_T reduces to showing that the sequence of its moments $m_k(A_T) := \operatorname{tr} A_T^k / p$ ($k \geq 1$ is a constant number) tends to a limit $(m_k)_k$, and this limit determines properly a probability distribution. For example, the later property is guaranteed if the moment sequence $(m_k)_k$ satisfies the Carleman condition:

$$\sum_{k=1}^{\infty} m_{2k}^{-1/2k} = \infty. \tag{3.1}$$

The following theorem gives the exact formula for the limiting moments $(m_k)_k$.

Theorem 3.1. *Suppose the following conditions hold:*

(a) *The* $p\times(T+s)$ *table* $(\varepsilon_{ij})_{1\le i\le p,1\le j\le T+s}$ *are made with an array of independent real random variables satisfying*

$$\mathbb{E}(\varepsilon_{ij})=0,\qquad \mathbb{E}\big(\varepsilon_{ij}^2\big)=1,\qquad \sup_{it}\mathbb{E}\big(\varepsilon_{it}^4\big)<\infty.$$

(b) p *and* T *tend to infinity proportionally*, *that is*,

$$p\to\infty,\qquad T\to\infty,\qquad y_T:=p/T\to y\in(0,\infty).$$

Then, *with probability one*, *the empirical spectral distribution* F^{A_T} *of the matrix* A_T *in* (1.2) *tends to a limiting distribution* F *whose* k*th moment* (k *is a fixed positive number*) *is given by*:

$$m_k=\sum_{i=0}^{k-1}\frac{1}{k}\binom{2k}{i}\binom{k}{i+1}y^{2k-1-i}.$$

Remark 3.1. *Using the expression of the limiting moment above*, *we are able to derive that the Stieltjes transform of* F:

$$s(z):=\int\frac{1}{x-z}\,dF(x)$$

satisfies the following equation:

$$y^2z^2s^3(z)+y^2zs^2(z)-yzs^2(z)-zs(z)-1=0,\tag{3.2}$$

which coincides with an earlier result in [8] *found by using the Stieltjes transform method.*

Indeed, *by the series expansion of the function* $\frac{1}{1-x}$, *the Stieltjes transform of a LSD can be expanded using its moments*:

$$\begin{aligned}s(z)&=\int\frac{1}{x-z}\,dF(x)=-\frac{1}{z}-\sum_{i=1}^{\infty}\frac{1}{z^{i+1}}\cdot m_i\\&=-\frac{1}{z}-\frac{1}{z}\cdot\sum_{i=1}^{\infty}\frac{m_i}{z^i}.\end{aligned}$$

Let $h(z)$ *be the moment generating function of* m_i:

$$h(z)=\sum_{i=0}^{\infty}m_iz^i,$$

then the part $\sum_{i=1}^{\infty}m_i/z^i$ *equals to* $h(1/z)-1$. *Therefore*, *we have the relationship between the Stieltjes transform* $s(z)$ *and the moment generating function* $h(z)$:

$$s(z)=-\frac{1}{z}h\left(\frac{1}{z}\right).\tag{3.3}$$

In the proof of Theorem 3.1, *we will see that* h *satisfies the equation*:

$$xy^2h^3(x)+x\big(y-y^2\big)h^2(x)-h(x)+1=0,$$

which is detailed in Section 5, *see* (5.32). *Let* $x=1/z$ *in it and combine with* (3.3) *leads to* (3.2).

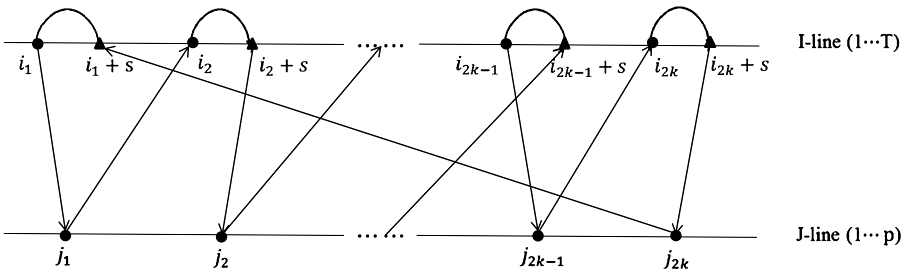

Fig. 5. The $Q(i, j)$-graph that corresponds to (3.4).

Proof of Theorem 3.1. After truncation and centralization, see Appendix A in [8], we may assume in all the following that

$$|\varepsilon_{ij}| \le \eta p^{1/4}, \qquad \mathbb{E}(\varepsilon_{ij}) = 0, \qquad \operatorname{Var}(\varepsilon_{ij}) = 1,$$

where η is chosen such that $\eta \to 0$ but $\eta p^{1/4} \to \infty$. And for convenience, we denote $M = \eta p^{1/4}$ below. With a little bit calculation, we have

$$\begin{aligned} m_k(A_T) &= \frac{1}{p} \sum_{\mathbf{i}=1}^{T} \sum_{\mathbf{j}=1}^{p} \frac{1}{T^{2k}} [\varepsilon_{j_1 i_1} \varepsilon_{j_1 i_2} \varepsilon_{j_2 s+i_2} \varepsilon_{j_2 s+i_3} \varepsilon_{j_3 i_3} \varepsilon_{j_3 i_4} \varepsilon_{j_4 s+i_4} \varepsilon_{j_4 s+i_5} \\ &\quad \cdots \varepsilon_{j_{2k-1} i_{2k-1}} \varepsilon_{j_{2k-1} i_{2k}} \varepsilon_{j_{2k} s+i_{2k}} \varepsilon_{j_{2k} s+i_1}] \\ &= \frac{1}{pT^{2k}} \sum_{\mathbf{i},\mathbf{j}} E_{Q(\mathbf{i},\mathbf{j})}, \end{aligned}$$

where the summation runs over all $Q(i, j)$-graph of length $4k$ (see Section 2 for the definition of the Q graph). The indices in $\mathbf{i} = (i_1, \dots, i_{2k})$ run over $1, 2, \dots, T$ and the indices in $\mathbf{j} = (j_1, \dots, j_{2k})$ run over $1, 2, \dots, p$. See the following Figure 5 for illustration.

Now suppose the pillar of the Q-graph in Figure 5 has t noncoincident I-vertices and s noncoincident J-vertices, which results in s down innovation and $t-1$ up innovation (we make the convention that the first down edge is always a down innovation and the last up edge is not an innovation). Then in the corresponding characteristic sequence $(d_1 u_1 \cdots d_{2k} u_{2k})$, the number of "1" ("1" only appears in the even position as it corresponds to the up edge) is $t-1$, the number of "-1" ("-1" only appears in the odd position) is $2k - s$ and the sequence starts and ends with "0."

We classify the Q-graphs in Figure 5 into three categories:

Category 1 (denoted by Q_1) contains all the Q-graphs that in its pillar $M(Q)$, each down edge must coincide with one and only one up edge and its base $B(Q)$ is a tree of $2k$ edges. In this category, $t + s - 1 = 2k$ and thus s is suppressed for simplicity. Figure 6 shows an example of Q_1 (see the left panel) with $k = 3$, $t = 2$ and $s = 5$. Its corresponding pillar and base are in the middle and right panel.

Category 2 (denoted by Q_2) contains all the Q-graphs that have at least one single vertical edge.

Category 3 (denoted by Q_3) contains all other Q-graphs. See Figure 7 for an example of Q_3 (the left panel) with $k = 3$, $t = 1$ and $s = 3$. In this example, the multiplicities of each edge in its pillar is four (the middle panel).

The almost sure convergence of the ESD of A_T will result from the following two assertions:

$$\begin{aligned} \mathbb{E}\big(m_k(A_T)\big) &= \frac{1}{p} \sum_{\mathbf{i}=1}^{T} \sum_{\mathbf{j}=1}^{p} \frac{1}{T^{2k}} \mathbb{E}[\varepsilon_{j_1 i_1} \varepsilon_{j_1 i_2} \varepsilon_{j_2 s+i_2} \varepsilon_{j_2 s+i_3} \varepsilon_{j_3 i_3} \varepsilon_{j_3 i_4} \varepsilon_{j_4 s+i_4} \varepsilon_{j_4 s+i_5} \\ &\quad \cdots \varepsilon_{j_{2k-1} i_{2k-1}} \varepsilon_{j_{2k-1} i_{2k}} \varepsilon_{j_{2k} s+i_{2k}} \varepsilon_{j_{2k} s+i_1}] \end{aligned} \tag{3.4}$$

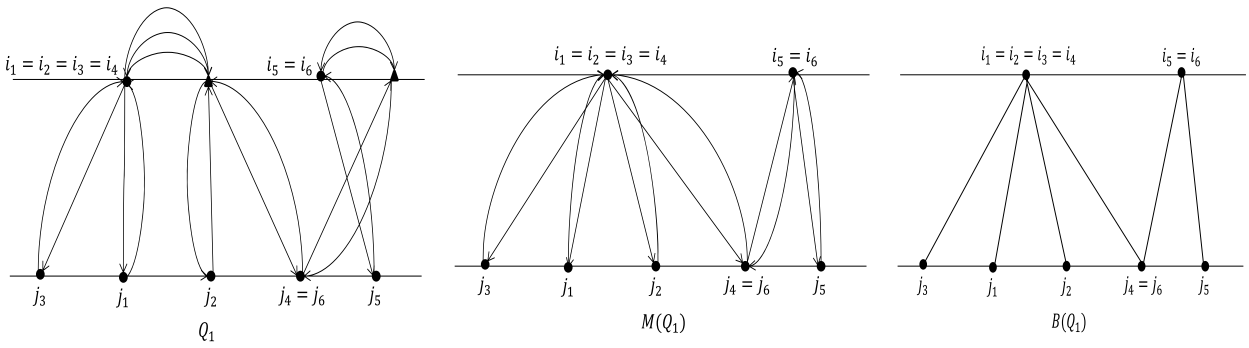

Fig. 6. An example of Q_1 (left panel) with its pillar (middle panel) and base (right panel).

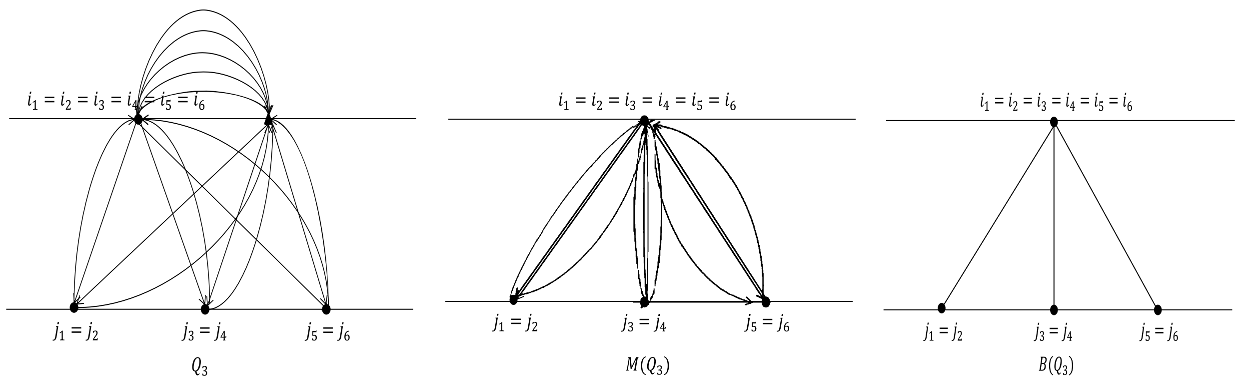

Fig. 7. An example of Q_3 (left panel) with its pillar (middle panel) and base (right panel).

$$
\begin{aligned}
&= \frac{1}{pT^{2k}} \sum_{\mathbf{i},\mathbf{j}} \mathbb{E}(E_{Q(\mathbf{i},\mathbf{j})}) \\
&= \sum_{i=0}^{k-1} \frac{1}{k} \binom{2k}{i} \binom{k}{i+1} y_T^{2k-1-i} + o(1),
\end{aligned} \tag{3.5}
$$

and

$$
\begin{aligned}
\operatorname{Var}\bigl(m_k(A_T)\bigr) &= \frac{1}{p^2T^{4k}} \sum_{\mathbf{i_1},\mathbf{j_1},\mathbf{i_2},\mathbf{j_2}} \bigl[\mathbb{E}(E_{Q_1(\mathbf{i_1},\mathbf{nj_1})} E_{Q_2(\mathbf{i_2},\mathbf{j_2})}) - \mathbb{E}(E_{Q_1(\mathbf{i_1},\mathbf{j_1})})\mathbb{E}(E_{Q_2(\mathbf{ni_2},\mathbf{nj_2})})\bigr] \\
&= O\bigl(p^{-2}\bigr).
\end{aligned} \tag{3.6}
$$

Proof of (3.5). Since $\mathbb{E}\varepsilon_{ij} = 0$, the only non-vanishing terms in (3.4) are those for which each edge in the Q-graph occurs at least twice. So the contribution of Category 2 is zero.

Next, we only consider those Q-graphs that fall in Category 1 and 3. Denote b_l the degree associated to the I-vertex i_l $(1 \le l \le t)$ in its corresponding M-pillar, then we have $b_1 + \cdots + b_t = 4k$, which is the total number of edges. On the other hand, since we glue the I-vertex i_l and $i_l + s$ in the definition of M-pillar, each degree b_l should be no less than 4; otherwise, there will be some single vertical edges in the Q-graph, which results in Category 2. Therefore, we have

$$4k = b_1 + \cdots + b_t \ge 4t,$$

which is $t \le k$.

Category 1: In Category 1, since in the Q-graph, each down edge must coincide with one and only one up edge, the total number of non-coincident edges is $2k$. Besides, due to the restriction that the base $B(Q)$ is a tree, we have $s+t-1=2k$ (according to the definition of a tree that $\#\{\text{edges}\}=\#\{\text{vertices}\}-1$), and this means that the summation over the elements in the corresponding characteristic sequence is zero (it is because we have the number of "1" equals $t-1$ and the number of "-1" equals $2k-s$).

The characteristic sequence of a M-pillar whose corresponding Q-graph lies in Category 1 has the following features (see Remarks 3.4 and 3.5):

(1) The total length of the characteristic sequence is $4k$;
(2) The sequence starts with a zero and ends with a zero (the first down edge is a down innovation and the last up edge is not an innovation);
(3) The number "1" appears only in the even position in the sequence and the number "-1" only in the odd position (down edges are in the odd position while up edges are in the even);
(4) The total number of "1" in the characteristic sequence is $t-1$ ($t-1$ up innovation), which equals the total number of "-1";
(5) The sequences are made with the following subsequence structure:

$$\begin{aligned}
&1\underbrace{00}_{\text{two}}-1\\
&1\underbrace{000000}_{\text{six}}-1\\
&1\underbrace{0000000000}_{\text{ten}}-1\\
&\cdots\\
&1\underbrace{0000\cdots00000}_{4k-6}-1.
\end{aligned}\tag{3.7}$$

Denote $f_{t-1}(k)$ as the number of Q-graphs whose M-pillar satisfies the above conditions (1)–(5) (here, we have two index: $t-1$, which is the number of up innovations and k, which is a quarter of the total length of the sequence), then we have the contribution of Category 1 to (3.4):

$$\begin{aligned}
&\frac{1}{pT^{2k}}\cdot\sum_{t=1}^{k}T(T-1)\cdots(T-t+1)p(p-1)\cdots(p-s+1)f_{t-1}(k)\\
&\quad=\sum_{t=1}^{k}y_T^{2k-t}f_{t-1}(k)+O\left(\frac{1}{p}\right).
\end{aligned}\tag{3.8}$$

Category 3: Category 3 consists two situations, see the following lemma.

Lemma 3.1 (Lemma 4.5 in [1]). *Denote the coincident multiplicities of the lth noncoincident vertical edge by a_l, $l=1,2,\dots,m$, where m is the number of noncoincident vertical edges. If $Q\in Q_3$, then* (a) *either there is a $a_l\geq 3$ with $t+s-1\leq m<2k$ or* (b) *all $a_l=2$ with $t+s-1<m=2k$.*

Remark 3.2. *An example that illustrates situation* (a) *in Lemma 3.1 is presented in Figure 8. In this example, $t=2$, $s=4$, $k=3$, $m=5$, thus we have $t+s-1=5=m<2k=6$ and there exists one vertical edge with multiplicity four. Another example is the left panel in Figure 7, which falls into situation* (b). *Since all its vertical edges are repeated exactly twice. Besides, in this case $m=6$, therefore we have $t+s-1=3<m=2k=6$.*

First, we see the contribution of (a). By the moment assumption that the moment $\mathbb{E}|\varepsilon_{ij}|^a$ is bounded by M^{a-2} for $a\geq 2$ and $a_1+\cdots+a_m=4k$, we conclude that the expectation

$$\mathbb{E}[\varepsilon_{j_1i_1}\varepsilon_{j_1i_2}\varepsilon_{j_2s+i_2}\varepsilon_{j_2s+i_3}\varepsilon_{j_3i_3}\varepsilon_{j_3i_4}\varepsilon_{j_4s+i_4}\varepsilon_{j_4s+i_5}\cdots\varepsilon_{j_{2k}s+i_{2k}}\varepsilon_{j_{2k}s+i_1}]$$

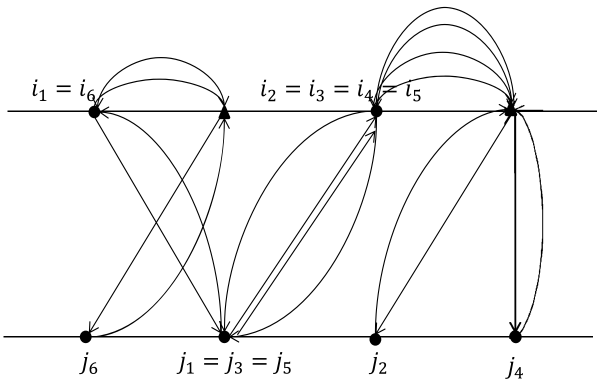

Fig. 8. An example of Q_3 that satisfies situation (a) in Lemma 3.1.

in (3.4) has magnitude at most M^{4k-2m}. Then we have (3.4) bounded by

$$\frac{1}{pT^{2k}} \cdot \sum_{t=1}^{k} T^t p^s M^{4k-2m} \#\{\text{isomorphism class in } Q_3\}$$

$$= O\left(\sum_{t=1}^{k} p^{t+s-k-m/2-1} \cdot \eta^{4k-2m}\right), \tag{3.9}$$

where the equality is due to the fact that p and T are in the same order and also for fixed k, the part $\#\{\text{isomorphism class in } Q_3\}$ is of order $O(1)$. Then

$$\sum_{t=1}^{k} p^{t+s-k-m/2-1} \le \sum_{t=1}^{k} p^{m-k-m/2} \le \sum_{t=1}^{k} p^{(2k-1)/2-k} = O\big(p^{-1/2}\big), \tag{3.10}$$

which is due to the assumption that $t+s-1 \le m < 2k$, so the contribution of (3.9) is $o(p^{-1/2})$.

Next, we consider the contribution of (b). Since all $a_l = 2$, we have the part of expectation equals 1. Therefore, (3.4) is bounded by

$$\frac{1}{pT^{2k}} \cdot \sum_{t=1}^{k} T^t p^s \#\{\text{isomorphism class in } Q_3\} = O\left(\sum_{t=1}^{k} p^{t+s-2k-1}\right) = O(1/p), \tag{3.11}$$

where the equation is due to the fact that $t+s \le 2k$.

Therefore, combine (3.8), (3.10) and (3.11), we finally have

$$\mathbb{E}\big(m_k(A_T)\big) = \sum_{t=1}^{k} y_T^{2k-t} f_{t-1}(k) + o(1). \tag{3.12}$$

To end the proof of (3.5), we need to determine the value of $f_{t-1}(k)$. This involves complex combinatorics and analytic arguments and the details of the derivation is given in Section 5. Finally, using (5.34) derived in Remark 5.1 that

$$f_m(k) = \frac{1}{k}\binom{2k}{m}\binom{k}{m+1},$$

we have

$$\mathbb{E}\big(m_k(A_T)\big) = \sum_{i=0}^{k-1} \frac{1}{k} \binom{2k}{i} \binom{k}{i+1} y_T^{2k-1-i} + o(1),$$

which is (3.5). □

Proof of (3.6). Recall

$$\begin{aligned}
&\operatorname{Var}\big(m_k(A_T)\big) \\
&\quad = \frac{1}{p^2 T^{4k}} \sum_{\mathbf{i}_1,\mathbf{j}_1,\mathbf{i}_2,\mathbf{j}_2} \big[\mathbb{E}(E_{Q_1(\mathbf{i}_1,\mathbf{j}_1)} E_{Q_2(\mathbf{i}_2,\mathbf{j}_2)}) - \mathbb{E}(E_{Q_1(\mathbf{i}_1,\mathbf{j}_1)})\mathbb{E}(E_{Q_2(\mathbf{i}_2,\mathbf{j}_2)})\big].
\end{aligned}$$

If Q_1 has no edges coincident with edges of Q_2, then due to the independence between Q_1 and Q_2, we have

$$\mathbb{E}(E_{Q_1(\mathbf{i}_1,\mathbf{j}_1)} E_{Q_2(\mathbf{i}_2,\mathbf{j}_2)}) - \mathbb{E}(E_{Q_1(\mathbf{i}_1,\mathbf{j}_1)})\mathbb{E}(E_{Q_2(\mathbf{i}_2,\mathbf{j}_2)}) = 0.$$

If $Q = Q_1 \cup Q_2$ has an overall single edge, then

$$\mathbb{E}(E_{Q_1(\mathbf{i}_1,\mathbf{j}_1)} E_{Q_2(\mathbf{i}_2,\mathbf{j}_2)}) = \mathbb{E}(E_{Q_1(\mathbf{i}_1,\mathbf{j}_1)})\mathbb{E}(E_{Q_2(\mathbf{i}_2,\mathbf{j}_2)}) = 0,$$

so the contribution to $\operatorname{Var}(m_k(A_T))$ is also zero.

Now, suppose $Q = Q_1 \cup Q_2$ contains no single edges and there's at least one edge in Q_1 coincident with one in Q_2, then the number of non-coincident I-vertices in Q is at least $t_1 + t_2 - 1$ and J-vertices is $s_1 + s_2 - 1$. Since $t_1 + s_1 - 1 \le 2k$ and $t_2 + s_2 - 1 \le 2k$, we have

$$\begin{aligned}
&\operatorname{Var}\big(m_k(A_T)\big) \\
&\quad = \frac{1}{p^2 T^{4k}} \sum_{\mathbf{i}_1,\mathbf{j}_1,\mathbf{i}_2,\mathbf{j}_2} \big[\mathbb{E}(E_{Q_1(\mathbf{i}_1,\mathbf{j}_1)} E_{Q_2(\mathbf{i}_2,\mathbf{j}_2)}) - \mathbb{E}(E_{Q_1(\mathbf{i}_1,\mathbf{j}_1)})\mathbb{E}(E_{Q_2(\mathbf{i}_2,\mathbf{j}_2)})\big] \\
&\quad = O\left(\frac{1}{p^2 T^{4k}} T^{t_1+t_2-1} p^{s_1+s_2-1}\right) \\
&\quad = O\big(p^{-2}\big),
\end{aligned}$$

which is (3.6). □

Carleman condition. In [8], the density function that corresponds to the Stieltjes transform $s(z)$ in (3.2) has been derived and it has compact support $[a, b]$, where

$$\begin{aligned}
a &= \frac{1}{8}\big(-1 + 20y + 8y^2 - (1+8y)^{3/2}\big) \cdot \mathbb{1}_{\{y \ge 1\}}, \\
b &= \frac{1}{8}\big(-1 + 20y + 8y^2 + (1+8y)^{3/2}\big).
\end{aligned} \tag{3.13}$$

Therefore, we have

$$m_k = \sum_{i=0}^{k-1} \frac{1}{k} \binom{2k}{i} \binom{k}{i+1} y^{2k-1-i} \le b^k. \tag{3.14}$$

From this, it is easy to see that the Carleman condition (3.1) is satisfied.

The proof of Theorem 3.1 is complete. □

Remark 3.3. *For the verification of Carleman condition, it would be enough to use the Stirling's formula in m_k to derive a less sharp bound $m_k \le A^k$ for some $A \ge b$. Since the sharp bound* (3.14) *will also be used in Section* 4, *its early introduction is thus preferred.*

Remark 3.4. *The explanation of* (5) *in* (3.7) *is that in the original Q-graph, each vertical edge is repeated exactly twice, and then we glue the I-vertex i_l and $i_l + s$ in its pillar $M(Q)$, therefore, the degree of each I-vertex is multiple of four, which implies that the length of each subsequence in* (3.7) *is multiple of four.*

Remark 3.5. *Note that in a characteristic sequence, subsequences in* (3.7) *cannot intersect each other; for example, if we arrange two subsequences* $1\ 0\ 0\ {-1}$ *and* $1\ 0\ 0\ 0\ 0\ 0\ 0\ {-1}$ *in the characteristic sequence of length* $4k = 16$, *then the following two structures are allowed*:

$$0\quad 1\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad -1\quad 1\quad 0\quad 0\quad -1\quad 0\quad 0\quad 0 \qquad (\textit{two subsequences are parallel}),$$

$$0\quad 1\quad 0\quad 1\quad 0\quad 0\quad -1\quad 0\quad -1\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0 \qquad (\textit{one is completely contained in another});$$

while

$$0\quad 1\quad 0\quad 0\quad 0\quad 0\quad 0\quad 1\quad -1\quad 0\quad -1\quad 0\quad 0\quad 0\quad 0\quad 0 \qquad (\textit{two subsequences intersect each other})$$

is not.

4. Convergence of the largest eigenvalue of A_T

Recall that due to (3.5) and (3.14) in the previous section, we have the following:

$$\begin{aligned}\mathbb{E}\big(m_k(A_T)\big) &= \sum_{i=0}^{k-1} \frac{1}{k}\binom{2k}{i}\binom{k}{i+1} y_T^{2k-1-i} + o(1) \\ &\le b(y_T)^k + o(1)\end{aligned}$$

for bounded k, where $b(y_T)$ is the value of b in (3.14) while substituting y_T for y. The main point in this section is to improve this estimation in order to allow a *growing* k such that:

$$\mathbb{E}\big(m_k(A_T)\big) = \sum_{i=0}^{k-1} \frac{1}{k}\binom{2k}{i}\binom{k}{i+1} y_T^{2k-1-i} \cdot \big(1 + o_k(1)\big), \tag{4.1}$$

where this $o_k(1)$ now (depending on k) tends to zero when $k \to \infty$. The derivation of the convergence of the largest eigenvalue in Wigner case can be referred to [13].

Proposition 4.1. *Suppose the following conditions hold*:

(a) *The $p \times (T+s)$ table $(\varepsilon_{ij})_{1\le i\le p, 1\le j\le T+s}$ are made with an array of independent real random variables satisfying*

$$\mathbb{E}(\varepsilon_{ij}) = 0, \qquad \mathbb{E}\big(\varepsilon_{ij}^2\big) = 1, \qquad \sup_{it} \mathbb{E}\big(|\varepsilon_{it}|^4\big) < \infty.$$

(b) *p and T tend to infinity proportionally, that is,*

$$p \to \infty, \qquad T \to \infty, \qquad y_T := p/T \to y \in (0, \infty).$$

(c) *k is an integer of satisfying $k = (\log p)^{1.01}$ (say).*

Then we have

$$\mathbb{E}\big(m_k(A_T)\big) = \sum_{i=0}^{k-1} \frac{1}{k} \binom{2k}{i} \binom{k}{i+1} y_T^{2k-1-i} \cdot \big(1 + o_k(1)\big).$$

Proof. After truncation, centralization and rescaling as the same route in Appendix B in [17], we may assume that the ε_{it}'s satisfy the condition that

$$\mathbb{E}(\varepsilon_{it}) = 0, \qquad \mathrm{Var}(\varepsilon_{it}) = 1, \qquad |\varepsilon_{it}| \le \delta T^{1/2}, \tag{4.2}$$

where δ is chosen such that

$$\begin{cases} \delta \to 0, \\ \delta T^{1/4+\beta} \to 0, \quad \text{for any } \beta > 0, \\ \delta T^{1/2} \to \infty. \end{cases} \tag{4.3}$$

When $k \to \infty$, the term #{isomorphism class} in (3.9) is no more a constant order. Then the main task is to show that the contribution of Category 3 to the kth moment of A_T when $k \to \infty$ can still be negligible compared with Category 1. And this will be achieved by counting precisely the number of isomorphism graphs in Category 3 as a function of k.

Since in Category 3, we have $t + s - 1 < 2k$. Using the previous notion of the characteristic sequence, we have $\#\{1\} = t - 1$ and $\#\{-1\} = 2k - s$, and this is equivalent to the fact that $\#\{1\} < \#\{-1\}$.

First, we consider the case that $t = 1$, which is to say $\#\{1\} = 0$ and $\#\{-1\} = 2k - s$. The way we choose $2k - s$ positions from the total $2k$ (the total number of length is $4k$, only the odd ones are allowed for "-1") is

$$\binom{2k}{2k-s}.$$

Then since we have s noncoincident J-vertices on the J-line, the noncoincident vertical edges is at most $2s$ (since we have each edge repeated at least twice). Therefore, the expectation

$$\mathbb{E}[\varepsilon_{j_1 i_1}\varepsilon_{j_1 i_2}\varepsilon_{j_2 s+i_2}\varepsilon_{j_2 s+i_3}\varepsilon_{j_3 i_3}\varepsilon_{j_3 i_4}\varepsilon_{j_4 s+i_4}\varepsilon_{j_4 s+i_5}\cdots\varepsilon_{j_{2k} s+i_{2k}}\varepsilon_{j_{2k} s+i_1}]$$

is bounded by $(\delta T^{1/2})^{4k-2s}$. Therefore, we have the contribution to (3.4):

$$\frac{1}{pT^{2k}} \sum_{s=1}^{2k-1} T p^s \big(\delta T^{1/2}\big)^{4k-2s} \binom{2k}{2k-s} = \frac{T}{p} \delta^{4k} \sum_{s=1}^{2k-1} \left(\frac{p}{\delta^2 T}\right)^s \binom{2k}{s}. \tag{4.4}$$

Since

$$\left(\frac{p}{\delta^2 T}\right)^s \binom{2k}{s} \le \left(\frac{2kp}{\delta^2 T}\right)^s \quad \text{and} \quad \frac{2kp}{\delta^2 T} \to \infty,$$

the dominating term in the summation at the right-hand side of (4.4) is when $s = 2k - 1$. Therefore, the left-hand side of (4.4) can be bounded as

$$O\left(2k \frac{T}{p} \delta^{4k} \left(\frac{p}{\delta^2 T}\right)^{2k-1}\right) = O\big(2k y^{2k-2} \delta^2\big).$$

Then consider the term when $t = 1$ in

$$\sum_{t=0}^{k-1} \frac{1}{k} \binom{2k}{t} \binom{k}{t+1} y_T^{2k-t-1},$$

which is

$$\frac{1}{k}\binom{2k}{1}\binom{k}{2}y_T^{2k-2}=O\big(k^2y_T^{2k-2}\big).$$

We have

$$\begin{aligned}2ky^{2k-2}\delta^2&=k^2y_T^{2k-2}\cdot o_k(1)\\&=\sum_{t=0}^{k-1}\frac{1}{k}\binom{2k}{t}\binom{k}{t+1}y_T^{2k-t-1}\cdot o_k(1).\end{aligned}\tag{4.5}$$

Then consider the case that $t>1$. Since $\#\{1\}=t-1<\#\{-1\}=2k-s$, we can first construct a characteristic sequence that satisfies (1)–(5) in (3.7). Therefore, we have the degree of each I-vertex at least four and each edge in the Q-graph repeated exactly twice, which ensures that the Q-graph will not fall in Category 2. And the possible ways for constructing such a characteristic sequence is $f_{t-1}(k)$ by definition. Since in the characteristic sequence, $2(t-1)$ positions have been taken to place the "1" and "−1," there leaves $4k-2(t-1)-2/2$ (the sequence starts and ends with a zero, so we should exclude the two positions at the beginning and at the end, and also "−1" appears in the odd positions, so we should divide it by two) positions to place the remaining "−1," whose number is $2k-t-s+1$, so the choice is bounded by

$$\binom{2k-t}{2k-t-s+1}.$$

Let m be the number of noncoincident vertical edges, which is no less than $t+s-1$, see Lemma 3.1, then the expectation

$$\mathbb{E}[\varepsilon_{j_1i_1}\varepsilon_{j_1i_2}\varepsilon_{j_2s+i_2}\varepsilon_{j_2s+i_3}\varepsilon_{j_3i_3}\varepsilon_{j_3i_4}\varepsilon_{j_4s+i_4}\varepsilon_{j_4s+i_5}\cdots\varepsilon_{j_{2k}s+i_{2k}}\varepsilon_{j_{2k}s+i_1}]$$

is bounded by $(\delta T^{1/2})^{4k-2m}\le(\delta T^{1/2})^{4k-2(t+s-1)}$. Finally, the contribution to (3.4) is bounded by:

$$\frac{1}{pT^{2k}}\sum_s\sum_t\big(\delta T^{1/2}\big)^{4k-2(t+s-1)}f_{t-1}(k)T^tp^s\binom{2k-t}{s-1}\tag{4.6}$$

$$=\frac{1}{pT^{2k}}\sum_{t=1}^{k}\big(\delta T^{1/2}\big)^{4k-2t+2}f_{t-1}(k)T^t\sum_{s=1}^{2k-t}\bigg(\frac{p}{\delta^2T}\bigg)^s\binom{2k-t}{s-1}.\tag{4.7}$$

Since

$$\bigg(\frac{p}{\delta^2T}\bigg)^s\binom{2k-t}{s-1}\le\bigg(\frac{2kp}{\delta^2T}\bigg)^s\quad\text{and}\quad\frac{2kp}{\delta^2T}\to\infty,$$

the dominating term in

$$\sum_{s=1}^{2k-t}\bigg(\frac{p}{\delta^2T}\bigg)^s\binom{2k-t}{s-1}$$

is when $s=2k-t$. Then (4.7) can be bounded as

$$\begin{aligned}&O\Bigg(\frac{2k}{pT^{2k}}\sum_{t=1}^{k}\big(\delta T^{1/2}\big)^{4k-2t+2}f_{t-1}(k)T^t\bigg(\frac{p}{\delta^2T}\bigg)^{2k-t}\Bigg)\\&\quad=O\Bigg(2k\delta^2\sum_{t=1}^{k}f_{t-1}(k)y_T^{2k-1-t}\Bigg)\end{aligned}$$

$$= O\left(2k\delta^2 \sum_{t=1}^{k} \frac{1}{k} \binom{2k}{t-1} \binom{k}{t} y_T^{2k-1-t}\right)$$

$$= O\left(2k\delta^2 y_T^{-1} \sum_{t=0}^{k-1} \frac{1}{k} \binom{2k}{t} \binom{k}{t+1} y_T^{2k-1-t}\right)$$

$$= \sum_{t=0}^{k-1} \frac{1}{k} \binom{2k}{t} \binom{k}{t+1} y_T^{2k-t-1} \cdot o_k(1), \tag{4.8}$$

where the last equality is due to the choice of k that

$$k\delta^2 = (\log p)^{1.01}\delta^2 = o\left(\left(\delta p^{1/4+\beta}\right)^2\right) \to 0, \quad \text{as } p \to \infty.$$

Finally, combine (4.5) and (4.8) leads to the fact that the contribution of Q_3 to (3.4) is

$$\sum_{t=0}^{k-1} \frac{1}{k} \binom{2k}{t} \binom{k}{t+1} y_T^{2k-t-1} \cdot o_k(1).$$

The proof of Proposition 4.1 is complete. □

Using the estimate in Proposition 4.1, we are able to prove the main result of this section, that is, the convergence of the largest eigenvalue of A_T to the right edge point of its support.

Theorem 4.1. *Under the same conditions as in Proposition 4.1, the largest eigenvalue of A_T converges to the right endpoint b defined in (3.14) almost surely.*

Proof. First we show that almost surely,

$$\liminf l_1 \geq b. \tag{4.9}$$

Indeed on the set $\{\liminf l_1 < b\}$, we have $\liminf l_1 < b - \delta$ for some $\delta = \delta(\omega) > 0$. Let $g : \mathbb{R} \to \mathbb{R}_+$ be a continuous and positive function supported on $[b-\delta, b]$, with $\int g(x)\,dF(x) = 1$, where F is the LSD of F^{A_T}. Then

$$\liminf \int g(x)\,dF^{A_T}(x) \leq 0. \tag{4.10}$$

For such ω, F^{A_T} will not converge weakly to F. Since this convergence occurs almost surely by Theorem 3.1, the claim (4.9) is proved.

Next, we claim that for any $\Delta > 0$,

$$\sum_{p=1}^{\infty} P(l_1 > b + \Delta) < \infty. \tag{4.11}$$

Write

$$l_1 - b = (l_1 - b)^+ - (l_1 - b)^-,$$

with its positive and negative parts. Claim (4.11) implies that almost surely, $(l_1 - b)^+ \to 0$. Therefore, a.s.

$$\limsup(l_1 - b) \leq \limsup(l_1 - b)^+ = 0.$$

Combine with (4.9), we have a.s. $\lim(l_1 - b) = 0$. It remains to prove claim (4.11).

Since we have

$$P(l_1 > b + \Delta) \le P\big(\operatorname{tr} A_T^k \ge (b + \Delta)^k\big) \le \frac{\mathbb{E} \operatorname{tr} A_T^k}{(b + \Delta)^k} = \frac{p \cdot \mathbb{E}(m_k(A_T))}{(b + \Delta)^k}, \tag{4.12}$$

and by (4.1), we have

$$p \cdot \mathbb{E}\big(m_k(A_T)\big) = p \sum_{i=0}^{k-1} \frac{1}{k} \binom{2k}{i} \binom{k}{i+1} y_T^{2k-1-i} \cdot \big(1 + o_k(1)\big). \tag{4.13}$$

Combining (3.14), (4.12) and (4.13) leads to:

$$\begin{aligned} P(l_1 > b + \Delta) &\le \frac{p \sum_{i=0}^{k-1} \frac{1}{k} \binom{2k}{i} \binom{k}{i+1} y_T^{2k-1-i} \cdot (1 + o_k(1))}{(b + \Delta)^k} \\ &\le p \left(\frac{b(y_T)}{b + \Delta} \right)^k \cdot \big(1 + o_k(1)\big). \end{aligned} \tag{4.14}$$

Since $y_T \to y$, for T sufficiently large, we have $b(y_T) \le b + \Delta/2$. Hence, (4.14) could be bounded by

$$p \left(\frac{b + \Delta/2}{b + \Delta} \right)^{(\log p)^{1.01}} \cdot \big(1 + o_p(1)\big) := p\alpha^{(\log p)^{1.01}} \cdot \big(1 + o_p(1)\big) := a_p.$$

Then due to the fact that $\alpha = \frac{b+\Delta/2}{b+\Delta} < 1$, we have

$$\lim_{p\to\infty} a_p^{1/p} = \lim_{p\to\infty} \exp\left(\frac{1}{p} \big(\log p + (\log p)^{1.01} \cdot \log \alpha + \log\big(1 + o_p(1)\big)\big) \right) < 1,$$

combining with (4.14) leads to

$$\sum_p P(l_1 > b + \Delta) \le \sum_p a_p < \infty.$$

The proof of Theorem 4.1 is complete. □

5. Determination of the sequence $\{f_m(k)\}$

One of the most challenging points in the proof of Theorem 3.1 is to determine the value of the sequence $\{f_m(k)\}$, i.e.

$$f_m(k) = \frac{1}{k} \binom{2k}{m} \binom{k}{m+1}. \tag{5.1}$$

Indeed, this enumeration cannot be done directly; rather it depends on the enumeration of a family of closely-related graphs, see the sequence $\{g_m(k)\}$ below. We first establish a system of two recursion formulas on the number of these two families in Section 5.1. Then these recursions are transferred in Section 5.2 to the generator functions $F_k(z) = \sum_{m=0}^k f_m(k) z^m$ and $G_k(z) = \sum_{m=0}^k g_m(k) z^m$. Next, we deduce in Section 5.3 two equations satisfied by $F(z, x) = \sum_{k=0}^\infty F_k(z) x^k$ and $G(z, x) = \sum_{k=0}^\infty G_k(z) x^k$. Finally, by solving these equations and taking the derivatives $\frac{\partial F(z,x)}{\partial x^k}|_{x=0}$, we obtain the target formula (5.1).

First recall the definition of $f_m(k)$, which is the number of M-pillars whose characteristic sequence satisfies the following conditions:

(1) The total length of the characteristic sequence is $4k$;
(2) The sequence starts with a zero and ends with a zero;

(3) The number "1" appears only in the even position in the sequence and the number "−1" only in the odd position;
(4) The total number of "1" in the characteristic sequence is m;
(5) The sequences are made with the following subsequence structure:

$$1\underbrace{00}_{\text{two}}-1$$
$$1\underbrace{000000}_{\text{six}}-1$$
$$1\underbrace{0000000000}_{\text{ten}}-1$$
$$\cdots$$
$$1\underbrace{0000\cdots00000}_{4k-6}-1.$$

Also, we define another M-pillar, whose characteristic sequence also satisfies the above condition (1)–(5), but with (2) replaced by the following (2)*:

(2)* The sequence starts with a zero and ends with three zeros.

We denote $g_m(k)$ as the number of such M-pillar satisfying (1), (2)*, (3)–(5).

5.1. *Master recursions on $f_m(k)$ and $g_m(k)$*

In this subsection, we derive a system of two master recursions on $f_m(k)$ and $g_m(k)$.

Once a characteristic sequence with length $2k$ is given, we denote S_n $(1 \le n \le 2k)$ as the partial sums of its first n elements. To ease the understanding of the upcoming proofs, we plot the path $\{(n, S_n)\}_{1\le n\le 2k}$ in Figure 9.

Definition 5.1. *We say that there is a* return to the origin *at time n, if $S_n = 0$.*

Definition 5.2. *A random walk is said to have a* first return to the origin *at time n $(n > 0)$, if $S_m \ne 0$ for all $m < n$ and $S_n = 0$.*

In Figure 9, the first return occurs at time $n = 9$.

5.1.1. *Master recursion 1*

First, we start with $f_m(k)$. Suppose the first return occurs at time i and $\max_{0\le n\le i} S_n = s$, which means that in the corresponding characteristic sequence, the number of "1" is s. We partition the random walk into two parts according to the first return, see Figure 10. For the reason that the length of the subsequence structures list in (5) are all multiplicity

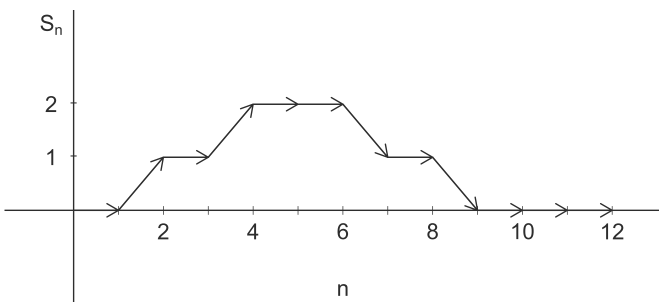

Fig. 9. The path $\{(n, S_n)\}$ that corresponds to the characteristic sequence (0 1 0 1 0 0 −1 0 −1 0 0 0).

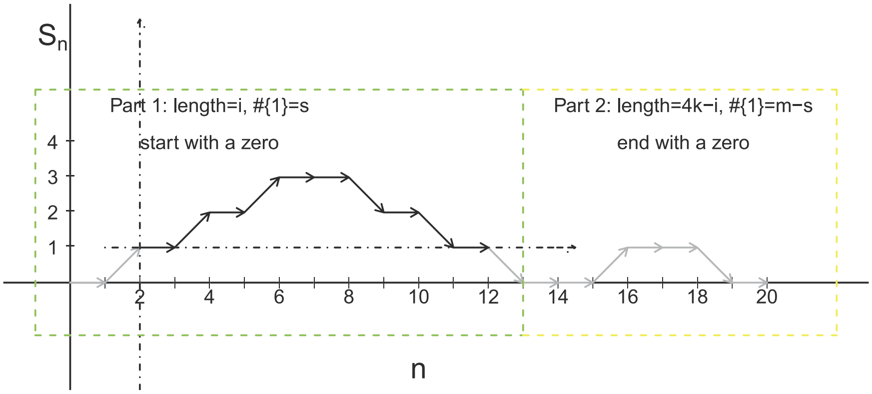

Fig. 10. Illustration of $f_m(k)$.

of four and condition (2), all the possibilities of i are $i = 5, 7, \dots, 4k-3, 4k-1$, we divide it into two cases:

Case 1: $i = i_1 = 4j - 3 \quad 2 \le j \le k$,

Case 2: $i = i_2 = 4j - 1 \quad 2 \le j \le k$.

Case 1: First, we consider the second part, which is of length $4k - i_1 = 4k - 4j + 3$, with the number of "1" being $m - s$ in its corresponding characteristic sequence. But this time, the sequence may not start with a "0" (once it returns to the origin, it can depart immediately, and in this case, the sequence starts with a "1"). Therefore, we add a zero in the front artificially, leading to a total length of $4k - 4j + 4$, which starts and ends with a zero. So the way of constructing such a sequence is $f_{m-s}(k - j + 1)$. Then consider the first part, which has a total length of i_1. Suppose the first departure from the origin is at time $n - 1$, where $n = 2, 6, \dots, i_1 - 7, i_1 - 3$ (also, it means that the first arrival at 1 is at the time n), then if we move the axis to the point $(n, 1)$ (see the black dashed axis in Figure 10) and consider the walk above the new x-axis (see the black parts in Figure 10), which has the length $i_1 - n - 1$. Further, this walk starts and ends with a zero, and if we add two more zeros in its end, it will lead to a walk with a total length of $i_1 - n + 1 = 4j - n - 2$, starts with a zero and ends with three zeros, with the number of "1" being $s - 1$. Therefore, the number of such walk is $g_{s-1}(\frac{4j-n-2}{4})$ according to the definition. So we have got the total contribution of Case 1 is:

$$\sum_{\substack{s=1,\dots,m \\ j=2,\dots,k}} \left(\sum_{n=2}^{4j-6} g_{s-1}\left(\frac{4j-n-2}{4}\right)\right) \cdot f_{m-s}(k-j+1). \tag{5.2}$$

Case 2: We follow the same route as in Case 1. After the first return, the remaining length is $4k - i_2 = 4k - 4j + 1$, then we add a zero in front and two zeros in the end, which leads to a walk of total length $4k - 4j + 4$, starts with a zero and ends with three zeros. And the number of "1" is $m - s$. By definition, the way of constructing such a walk is $g_{m-s}(k - j + 1)$. Then for the first part, also suppose the first departure from the origin is at time $n - 1$ ($n = 4, 8, \dots, i_2 - 7, i_2 - 3$) and consider the part of the walk that is above the new x-axis (with the new origin located at $(n, 1)$), which is of total length $i_2 - n - 1$. Then we add two more zeros in the end, results in a walk of total length $i_2 - n + 1 = 4j - n$, starts with a zero and ends with three zeros, and the number of "1" is $s - 1$. The way of constructing such a walk is $g_{s-1}(\frac{4j-n}{4})$. And combine these two parts, the contribution of Case 2 is:

$$\sum_{\substack{s=1,\dots,m \\ j=2,\dots,k}} \left(\sum_{n=4}^{4j-4} g_{s-1}\left(\frac{4j-n}{4}\right)\right) \cdot g_{m-s}(k-j+1). \tag{5.3}$$

Overall, combine (5.2) and (5.3) leads to the following recursion:

$$\begin{aligned} f_m(k) &= \sum_{\substack{s=1,\dots,m \\ j=2,\dots,k}} \left(\sum_{n=2}^{4j-6} g_{s-1}\left(\frac{4j-n-2}{4}\right)\right) \cdot f_{m-s}(k-j+1) \\ &\quad + \sum_{\substack{s=1,\dots,m \\ j=2,\dots,k}} \left(\sum_{n=4}^{4j-4} g_{s-1}\left(\frac{4j-n}{4}\right)\right) \cdot g_{m-s}(k-j+1) \\ &= \sum_{\substack{s=1,\dots,m \\ j=2,\dots,k}} \left(\sum_{n=4}^{4j-4} g_{s-1}\left(\frac{4j-n}{4}\right)\right) \cdot \left[f_{m-s}(k-j+1) + g_{m-s}(k-j+1)\right]. \end{aligned} \tag{5.4}$$

5.1.2. *Master recursion* 2

Then we start with $g_m(k)$. Also suppose the first return time is i, where $i = 5, 7, \dots, 4k-5, 4k-3$ (in the definition of $g_m(k)$, the characteristic sequence ends with three zeros, so the maximum value of i is $4k-3$ here). We divide these i into two cases:

Case 1: $i = i_1 = 4j-5, \qquad 3 \le j \le k,$

Case 2: $i = i_2 = 4j-3, \qquad 2 \le j \le k.$

As before, we suppose the number of "1" is s in the first part.

Case 1: First for the second part, which ends with three zeros but may not start with a zero. Since the total length is $4k - i_1 = 4k - 4j + 5$, if we add a zero in its beginning and remove the last two zeros, then it will result in a walk whose characteristic sequence starts and ends with a zero, whose length is $4k-4j+5+1-2 = 4k-4j+4$, with the number of "1" being $m-s$. The total number of constructing such a walk is $f_{m-s}(k-j+1)$. Then for the first part, suppose the first departure from the origin is at time $n-1$, where $n = 4, 8, \dots, i_1-7, i_1-3$. We do the same thing as before, add the new axis whose origin is located at $(n, 1)$. Then we consider the part above this new x-axis, see the black part in Figure 11, whose length is $i_1 - n - 1$. We add two more zeros in the end, it actually becomes the walk with a total length of $i_1 - n + 1 = 4j - n - 4$, starts with a zero and ends with three zeros, with the number of "1" being $s-1$. The way of constructing such a walk is $g_{s-1}(\frac{4j-n-4}{4})$. Combine all this, the contribution of Case 1 is

$$\sum_{\substack{s=1,\dots,m \\ j=3,\dots,k}} \left(\sum_{n=4}^{4j-8} g_{s-1}\left(\frac{4j-n-4}{4}\right)\right) \cdot f_{m-s}(k-j+1). \tag{5.5}$$

Case 2: For the second part, we add a zero in the front, which leads to a walk of total length $4k - i_2 + 1 = 4k - 4j + 4$, starts with a zero and ends with three zeros, with the number of "1" being $m-s$, so the way of constructing such a walk is $g_{m-s}(k-j+1)$. Then for the first part, we also consider the part that above the new x-axe. Since $n = 2, 6, \dots, i_2-7, i_2-3$ this time, we add two more zeros in the end and this leads to a walk with total length of $i_2 - n + 1 = 4j - 3 - n + 1$, starts with a zero and ends with three zeros, with the number of "1" being $s-1$, and the total way of constructing such a walk is $g_{s-1}(\frac{4j-2-n}{4})$. Combine these two parts, the contribution of Case 2 is

$$\sum_{\substack{s=1,\dots,m \\ j=2,\dots,k}} \left(\sum_{n=2}^{4j-6} g_{s-1}\left(\frac{4j-n-2}{4}\right)\right) \cdot g_{m-s}(k-j+1). \tag{5.6}$$

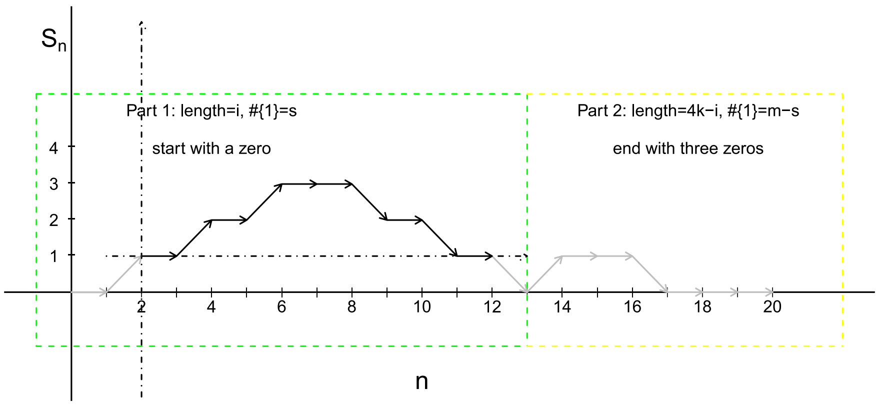

Fig. 11. Illustration of $g_m(k)$.

Finally, combine (5.5) and (5.6) leads to the recursion

$$\begin{aligned} g_m(k) &= \sum_{\substack{s=1,\dots,m\\ j=3,\dots,k}} \left(\sum_{n=4}^{4j-8} g_{s-1}\left(\frac{4j-n-4}{4}\right)\right) \cdot f_{m-s}(k-j+1) \\ &\quad + \sum_{\substack{s=1,\dots,m\\ j=2,\dots,k}} \left(\sum_{n=2}^{4j-6} g_{s-1}\left(\frac{4j-n-2}{4}\right)\right) \cdot g_{m-s}(k-j+1) \\ &= \sum_{\substack{s=1,\dots,m\\ j=3,\dots,k}} \left(\sum_{n=6}^{4j-6} g_{s-1}\left(\frac{4j-n-2}{4}\right)\right) \cdot \big(f_{m-s}(k-j+1)+g_{m-s}(k-j+1)\big) \\ &\quad + \sum_{\substack{s=1,\dots,m\\ j=2,\dots,k}} g_{s-1}(j-1)g_{m-s}(k-j+1) \\ &= \sum_{\substack{s=1,\dots,m\\ j=3,\dots,k}} \left(\sum_{n=8}^{4j-4} g_{s-1}\left(\frac{4j-n}{4}\right)\right) \cdot \big(f_{m-s}(k-j+1)+g_{m-s}(k-j+1)\big) \\ &\quad + \sum_{\substack{s=1,\dots,m\\ j=2,\dots,k}} g_{s-1}(j-1)g_{m-s}(k-j+1). \end{aligned} \tag{5.7}$$

As a result, (5.4) and (5.7) lead to a system of the two recursions on $f_m(k)$ and $g_m(k)$:

$$\begin{cases} f_m(k) - g_m(k) = \sum_{s=1,\dots,m,j=2,\dots,k} g_{s-1}(j-1) f_{m-s}(k-j+1), \\ g_m(k) = \sum_{s=1,\dots,m,j=3,\dots,k} (\sum_{l=2}^{j-1} g_{s-1}(j-l)) \cdot (f_{m-s}(k-j+1) + g_{m-s}(k-j+1)) \\ \qquad + \sum_{s=1,\dots,m,j=2,\dots,k} g_{s-1}(j-1) g_{m-s}(k-j+1). \end{cases} \tag{5.8}$$

5.2. *Recursions related to $F_k(z)$ and $G_k(z)$*

Define $F_k(z) = \sum_{m=0}^k f_m(k)z^m$ and $G_k(z) = \sum_{m=0}^k g_m(k)z^m$. Recall the definition of $f_m(k)$ and $g_m(k)$, where $m \le k-1$, so we make the convention that $f_k(k) = g_k(k) = 0$ and $f_0(k) = g_0(k) = 1$. In this subsection, the main purpose is to derive the following two recursions that related to $F_k(z)$ and $G_k(z)$:

$$\begin{cases} F_k(z) - G_k(z) = z\sum_{j=2}^{k} G_{j-1}(z)\cdot F_{k-j+1}(z), \\ G_k(z) = 1 + z\sum_{j=3}^{k}\sum_{l=2}^{j-1} G_{j-l}(z)(F_{k-j+1}(z) + G_{k-j+1}(z)) \\ \qquad\qquad + z\sum_{j=2}^{k} G_{j-1}(z)G_{k-j+1}(z). \end{cases} \tag{5.9}$$

By multiplying z^m on both sides in the first recursion in (5.8), and then summing from $m = 1$ to k, the left-hand side equals to:

$$\begin{aligned} \sum_{m=1}^{k} f_m(k)z^m - \sum_{m=1}^{k} g_m(k)z^m &= \sum_{m=0}^{k} f_m(k)z^m - \sum_{m=0}^{k} g_m(k)z^m \\ &= F_k(z) - G_k(z), \end{aligned}$$

and the right-hand side equals to

$$z\sum_{m=1}^{k}\sum_{\substack{s=1,\dots,m \\ j=2,\dots,k}} g_{s-1}(j-1)z^{s-1}\cdot f_{m-s}(k-j+1)z^{m-s}. \tag{5.10}$$

Now consider

$$\begin{aligned} G_{j-1}(z)\cdot F_{k-j+1}(z) &= \sum_{m=0}^{j-1} g_m(j-1)z^m \cdot \sum_{n=0}^{k-j+1} f_n(k-j+1)z^n \\ &:= \sum_{s=0}^{k} a_s\cdot z^s, \end{aligned}$$

where

$$\begin{aligned} &s = m+n, \\ &a_s = \sum_{m+n=s} g_m(j-1)f_n(k-j+1). \end{aligned} \tag{5.11}$$

Then (5.10) equals to

$$\begin{aligned} z\sum_{j=2}^{k}\sum_{m=1}^{k} a_{m-1}\cdot z^{m-1} &= z\sum_{j=2}^{k}\sum_{m=0}^{k-1} a_m\cdot z^m = z\sum_{j=2}^{k}\sum_{m=0}^{k} a_m\cdot z^m \\ &= z\sum_{j=2}^{k} G_{j-1}(z)\cdot F_{k-j+1}(z), \end{aligned}$$

where the second equality is due to the fact that $a_k = 0$ (since in (5.11), we have $m \le j-2$ and $n \le k-nj$ by definition, then $s = m+n \le k-2$, so the term $a_{k-1} = a_k = 0$). Therefore, we have got

$$F_k(z) - G_k(z) = z\sum_{j=2}^{k} G_{j-1}(z)\cdot F_{k-j+1}(z),$$

which is the first recursion in (5.9).

Next, from the second recursion in (5.8), we have:

$$g_m(k) = \underbrace{\sum_{\substack{s=1,\dots,m \\ j=3,\dots,k}} \left(\sum_{l=2}^{j-1} g_{s-1}(j-l)\right) \cdot \big(f_{m-s}(k-j+1) + g_{m-s}(k-j+1)\big)}_{(I)}$$
$$+ \underbrace{\sum_{\substack{s=1,\dots,m \\ j=2,\dots,k}} g_{s-1}(j-1) g_{m-s}(k-j+1)}_{(II)}. \tag{5.12}$$

Consider

$$\begin{aligned} G_{j-l}(z)F_{k-j+1}(z) &= \sum_{u=0}^{j-l} g_u(j-l)z^u \cdot \sum_{v=0}^{k-j+1} f_v(k-j+1)z^v \\ &:= \sum_{n=0}^{k+1-l} b_n z^n, \end{aligned} \tag{5.13}$$

where $n = u + v$, $b_n = \sum_{u+v=n} g_u(j-l) f_v(k-j+1)$, also, we have $u \le j-l-1$ and $v \le k-j$, which leads to $n = u + v \le k - l - 1$, so the terms that correspond to $n = k - l + 1, k - l$ equal zero. For the same reason,

$$\begin{aligned} G_{j-l}(z)G_{k-j+1}(z) &= \sum_{u=0}^{j-l} g_u(j-l)z^u \cdot \sum_{v=0}^{k-j+1} g_v(k-j+1)z^v \\ &:= \sum_{n=0}^{k+1-l} c_n z^n, \end{aligned} \tag{5.14}$$

where $n = u + v$, $c_n = \sum_{u+v=n} g_u(j-l) g_v(k-j+1)$ and the terms that correspond to $n = k-l+1, k-l$ equal zero. And

$$G_{j-1}(z)G_{k-j+1}(z) = \sum_{u=0}^{j-l} g_u(j-l)z^u \cdot \sum_{v=0}^{k-j+1} g_v(k-j+1)z^v := \sum_{n=0}^{k} d_n z^n, \tag{5.15}$$

where $n = u + v$, $d_n = \sum_{u+v=n} g_u(j-1) g_v(k-j+1)$ and the terms that correspond to $n = k, k-1$ equal zero since $u \le j-2$ and $v \le k-j$ lead to $n = u + v \le k-2$.

We multiply z^m on both sides of (5.12) and sum from $m = 1$ to k, then the left-hand side equals to

$$G_k(z) - 1. \tag{5.16}$$

Now consider the part (I):

$$\begin{aligned} \sum_{m=1}^{k} (I) \cdot z^m &= \sum_{m=1}^{k} \sum_{j=3}^{k} \sum_{l=2}^{j-1} (b_{m-1} + c_{m-1}) \cdot z^m \\ &= z \sum_{j=3}^{k} \sum_{l=2}^{j-1} \left[\sum_{n=0}^{k-1} b_n z^n + \sum_{n=0}^{k-1} c_n z^n \right] \end{aligned}$$

$$= z\sum_{j=3}^{k}\sum_{l=2}^{j-1}\left[\sum_{n=0}^{k-l-1} b_n z^n + \sum_{n=0}^{k-l-1} c_n z^n\right]$$

$$= z\sum_{j=3}^{k}\sum_{l=2}^{j-1}\left[\sum_{n=0}^{k-l+1} b_n z^n + \sum_{n=0}^{k-l+1} c_n z^n\right]$$

$$= z\sum_{j=3}^{k}\sum_{l=2}^{j-1}\left[G_{j-l}(z)F_{k-j+1}(z) + G_{j-l}(z)G_{k-j+1}(z)\right]. \tag{5.17}$$

And for the part (II),

$$\sum_{m=1}^{k}(II)\cdot z^m = \sum_{m=1}^{k}\sum_{j=2}^{k} d_{m-1}\cdot z^m = z\sum_{j=2}^{k}\sum_{m=0}^{k-1} d_m\cdot z^m$$

$$= z\sum_{j=2}^{k}\sum_{m=0}^{k} d_m\cdot z^m = z\sum_{j=2}^{k} G_{j-1}(z)G_{k-j+1}(z). \tag{5.18}$$

Combining (5.12), (5.16), (5.17) and (5.18), we have got

$$G_k(z) = 1 + z\sum_{j=3}^{k}\sum_{l=2}^{j-1}\left[G_{j-l}(z)F_{k-j+1}(z) + G_{j-l}(z)G_{k-j+1}(z)\right]$$

$$+ z\sum_{j=2}^{k} G_{j-1}(z)G_{k-j+1}(z),$$

which is the second recursion in (5.9).

5.3. *Equations related to $F(z,x)$ and $G(z,x)$*

Define $F(z,x) = \sum_{k=0}^{\infty} F_k(z)x^k$, $G(z,x) = \sum_{k=0}^{\infty} G_k(z)x^k$ and the term that corresponding to $k = 0$ equals 0. In this section, we derive the following equations:

$$\begin{cases} F(z,x) - G(z,x) = zF(z,x)G(z,x), \\ G(z,x) = \sum_{k=1}^{\infty} x^k + z\sum_{k=2}^{\infty} F(z,x)G(z,x)x^{k-1} + z\sum_{k=2}^{\infty} G(z,x)G(z,x)x^{k-1} \\ \quad + zG(z,x)G(z,x), \end{cases} \tag{5.19}$$

which will lead to the solutions of $F(z,x)$ and $G(z,x)$ as functions of z and x.

Since the first recursion in (5.9):

$$F_k(z) - G_k(z) = z\sum_{j=2}^{k} G_{j-1}(z)\cdot F_{k-j+1}(z),$$

we multiply x^k on both sides and do summation from $k = 1$ to ∞, then the left-hand side is exactly $F(z,x) - G(z,x)$. The right-hand side is

$$z\sum_{k=2}^{\infty}\sum_{j=2}^{k} G_{j-1}(z)F_{k-j+1}(z)x^k = zF(z,x)G(z,x),$$

which leads to the first equation in (5.19):

$$F(z,x) - G(z,x) = zF(z,x)G(z,x).$$

Then we denote the second recursion in (5.9) as follows:

$$\begin{aligned} G_k(z) = 1 &+ \underbrace{z\sum_{j=3}^{k}\sum_{l=2}^{j-1} G_{j-l}(z)F_{k-j+1}(z)}_{(I)} + \underbrace{z\sum_{j=3}^{k}\sum_{l=2}^{j-1} G_{j-l}(z)G_{k-j+1}(z)}_{(II)} \\ &+ \underbrace{z\sum_{j=2}^{k} G_{j-1}(z)G_{k-j+1}(z)}_{(III)}. \end{aligned} \tag{5.20}$$

We multiply x^k on both sides and then do summation from $k=1$ to ∞, the left-hand side equals $G(z,x)$. The part (I):

$$\begin{aligned} \sum_{k=1}^{\infty}(I)\cdot x^k &= z\sum_{k=1}^{\infty}\sum_{j=3}^{k}\sum_{l=2}^{j-1} G_{j-l}(z)F_{k-j+1}(z)x^k \\ &= z\sum_{k=3}^{\infty}\sum_{l=2}^{k-1}\left(\sum_{j=l+1}^{k} G_{j-l}(z)x^{j-l}\cdot F_{k-j+1}(z)x^{k-j+1}\right)\cdot x^{l-1} \\ &= z\sum_{l=2}^{\infty}\left(\sum_{k=l+1}^{\infty}\sum_{j=l+1}^{k} G_{j-l}(z)x^{j-l}\cdot F_{k-j+1}(z)x^{k-j+1}\right)\cdot x^{l-1} \\ &= z\sum_{l=2}^{\infty} F(z,x)G(z,x)x^{l-1}, \end{aligned} \tag{5.21}$$

and for the same reason, the contribution of part (II) equals

$$z\sum_{l=2}^{\infty} G(z,x)G(z,x)x^{l-1}. \tag{5.22}$$

For the part (III):

$$\begin{aligned} z\sum_{k=1}^{\infty}\sum_{j=2}^{k} G_{j-1}(z)G_{k-j+1}(z)x^k &= z\sum_{k=2}^{\infty}\sum_{j=2}^{k} G_{j-1}(z)x^{j-1}\cdot G_{k-j+1}(z)x^{k-j+1} \\ &= zG(z,x)G(z,x). \end{aligned} \tag{5.23}$$

Finally, combining (5.21), (5.22) and (5.23) leads to:

$$\begin{aligned} G(z,x) = \sum_{k=1}^{\infty} x^k &+ z\sum_{k=2}^{\infty} F(z,x)G(z,x)x^{k-1} + z\sum_{k=2}^{\infty} G(z,x)G(z,x)x^{k-1} \\ &+ zG(z,x)G(z,x), \end{aligned}$$

which is the second equation in (5.19).

5.4. *Exact formula for m_k*

First, since $F(z,x)=\sum_{k=0}^{\infty} F_k(z)x^k$, we have

$$F_k(z)=\frac{1}{k!}\cdot\frac{\partial F(z,x)}{\partial x^k}\bigg|_{x=0}. \tag{5.24}$$

In the following, we will use the shorthands $F=F(z,x)$, $G=G(z,x)$ and $F_k=F_k(z)$. In (5.19), we can first express G in function of F using the first equation and then derive from the second equation that

$$F=\sum_{l=1}^{\infty} x^l\cdot\left(1+z^2F^2+2zF+zF^2+z^2F^3+zF^2\right).$$

Taking kth derivative on both sides with respect to x and combining with (5.24), we have:

$$F_k=1+2z\sum_{j=1}^{k-1}F_j+\left(z^2+2z\right)\sum_{l=2}^{k-1}\sum_{\substack{a+b=l\\a,b\geq 1}}F_aF_b+z^2\sum_{l=3}^{k-1}\sum_{\substack{a+b+c=l\\a,b,c\geq 1}}F_aF_bF_c. \tag{5.25}$$

Due to the definition of m_k that $m_k=y^{2k-1}F_k(\frac{1}{y})$, substituting $1/y$ for z in (5.25) and multiplying both sides by y^{2k-1}, we get the recursion for m_k:

$$m_k=y^{2k-1}+\frac{2}{y}\sum_{j=1}^{k-1}m_j\cdot y^{2k-2j}+\left(\frac{1}{y^2}+\frac{2}{y}\right)\sum_{l=2}^{k-1}\sum_{\substack{a+b=l\\a,b\geq 1}}m_am_b\cdot y^{2k-2l+1}$$
$$+\frac{1}{y^2}\sum_{l=3}^{k-1}\sum_{\substack{a+b+c=l\\a,b,c\geq 1}}m_am_bm_c\cdot y^{2k-2l+2}. \tag{5.26}$$

Then we substitute $k-1$ for k in (5.26) and multiply both sides by y^2, which leads to the following:

$$y^2\cdot m_{k-1}=y^{2k-1}+\frac{2}{y}\sum_{j=1}^{k-2}m_j\cdot y^{2k-2j}+\left(\frac{1}{y^2}+\frac{2}{y}\right)\sum_{l=2}^{k-2}\sum_{\substack{a+b=l\\a,b\geq 1}}m_am_b\cdot y^{2k-2l+1}$$
$$+\frac{1}{y^2}\sum_{l=3}^{k-2}\sum_{\substack{a+b+c=l\\a,b,c\geq 1}}m_am_bm_c\cdot y^{2k-2l+2}. \tag{5.27}$$

Then by combining (5.26) and (5.27), we have:

$$m_k=\left(2y+y^2\right)m_{k-1}+\left(y+2y^2\right)\cdot\sum_{\substack{a+b=k-1\\a,b\geq 1}}m_am_b+y^2\cdot\sum_{\substack{a+b+c=k-1\\a,b,c\geq 1}}m_am_bm_c. \tag{5.28}$$

By the definition of m_k that $m_0=1$, we have

$$\begin{aligned}\sum_{\substack{a+b+c=k-1\\a,b,c\geq 1}}m_am_bm_c&=\sum_{\substack{a+b+c=k-1\\a,b,c\geq 0}}m_am_bm_c-3\sum_{\substack{a+b=k-1\\a,b\geq 1}}m_am_b-3m_{k-1},\\ \sum_{\substack{a+b=k-1\\a,b\geq 1}}m_am_b&=\sum_{\substack{a+b=k-1\\a,b\geq 0}}m_am_b-2m_{k-1}.\end{aligned} \tag{5.29}$$

Bringing these two equations in (5.29) into (5.28), we get:

$$y^2 \sum_{\substack{a+b+c=k-1\\ a,b,c\geq 0}} m_a m_b m_c + (y-y^2) \sum_{\substack{a+b=k-1\\ a,b\geq 0}} m_a m_b = m_k. \tag{5.30}$$

Then let $h(x)$ be the moment generating function: $h(x)=\sum_{k=0}^{\infty} m_k x^k$, we multiply x^k on both sides of (5.30) and do summation from $k=1$ to ∞ and combine with the fact that

$$h(x) = 1 + \sum_{k=1}^{\infty} m_k x^k \tag{5.31}$$

leading to:

$$xy^2 h^3(x) + x(y-y^2)h^2(x) - h(x) + 1 = 0. \tag{5.32}$$

Based on the theory of Bürmann–Lagrange series, see page 145 of [12], and let $z = h(x)-1$ and $\varphi = y^2(z+1)^3 + (y-y^2)(z+1)^2$, we may invert (5.32) to obtain that

$$z = \sum_{n=1}^{\infty} \frac{w^n}{n!} \left[\frac{d^{n-1}[y^n(z+1)^{2n}(yz+1)^n]}{dz^{n-1}} \right] \bigg|_{z=0},$$

where $w = z/\varphi = x$. Then based on the Leibniz's rule in differential calculus, we have

$$\frac{d^{n-1}[(z+1)^{2n}(yz+1)^n]}{dz^{n-1}} = \sum_{i=0}^{n-1} \binom{n-1}{i} \left[\frac{d^i[(z+1)^{2n}]}{dz^i} \cdot \frac{d^{n-1-i}[(yz+1)^n]}{dz^{n-1-i}} \right],$$

which leads to the fact that

$$h(x) = 1 + z = 1 + \sum_{n=1}^{\infty} \left[\sum_{i=0}^{n-1} \frac{1}{n} \binom{2n}{i} \binom{n}{i+1} y^{2n-1-i} \right] \cdot x^n,$$

and this is equivalent to

$$m_k = \sum_{i=0}^{k-1} \frac{1}{k} \binom{2k}{i} \binom{k}{i+1} y^{2k-1-i}. \tag{5.33}$$

Remark 5.1. *Since*

$$m_k = y^{2k-1} F_k\left(\frac{1}{y}\right) = y^{2k-1} \sum_{m=0}^{k} f_m(k) \frac{1}{y^m} = \sum_{m=0}^{k-1} f_m(k) y^{2k-1-m},$$

(5.33) *reduces to the fact that*

$$f_m(k) = \frac{1}{k} \binom{2k}{m} \binom{k}{m+1}. \tag{5.34}$$

Remark 5.2. *The recursion* (5.30) *has a remarkable nature. Notice that the recursion*

$$c_k = \sum_{\substack{a+b=k-1\\ a,b\geq 0}} c_a c_b$$

and

$$d_k = \sum_{\substack{a+b+c=k-1 \\ a,b,c \geq 0}} d_a d_b d_c$$

define the (*standard*) *Catalan numbers and the generalized Catalan numbers of order three*, *respectively* (*see* [5]). *The moment sequence* (m_k) *of the LSD of this paper can be thought as a complex combination of these two families of Catalan numbers*.

Acknowledgements

The authors thank funding from Research Grants Council, Hong Kong SAR, China [GRF 705413, GRF 17305814], NSFC Grant No. 11371317 and the Fundamental Research Funds for the Central Universities.

References

[1] Z. D. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*, 2nd edition. Springer, New York, 2010. MR2567175
[2] Z. D. Bai and Y. Q. Yin. Necessary and sufficient conditions for the almost sure convergence of the largest eigenvalue of Wigner matrices. *Ann. Probab.* **16** (4) (1988) 1729–1741. MR0958213
[3] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.* **8** (1980) 252–261. MR0566592
[4] V. L. Girko. Circle law. *Theory Probab. Appl.* **4** (1984) 694–706.
[5] P. Hilton and J. Pedersen. Catalan numbers, their generalization, and their uses. *Math. Intelligencer* **13** (1991) 64–75. MR1098222
[6] B. S. Jin, C. Wang, Z. D. Bai, K. K. Nair and M. C. Harding. Limiting spectral distribution of a symmetrized auto-cross covariance matrix. *Ann. Appl. Probab.* **24** (2014) 1199–1225. MR3199984
[7] C. Lam and Q. W. Yao. Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** (2012) 694–726. MR2933663
[8] Z. Li, G. M. Pan and J. F. Yao. On singular value distribution of large-dimensional autocovariance matrices. *J. Multivariate Anal.* **137** (2015) 119–140. MR3332802
[9] H. Y. Liu, A. Aue and D. Paul. On the Marčenko-Pastur law for linear time series. *Ann. Statist.* **43** (2015) 675–712. MR3319140
[10] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR. Sbornik* **1** (1967) 457–483.
[11] S. Péché and A. Soshnikov. On the lower bound of the spectral norm of symmetric random matrices with independent entries. *Electron. Commun. Probab.* **13** (2008) 280–290. MR2415136
[12] G. Pólya and G. Szegö. *Problems and Theorems in Analysis (Volume I)*. Springer, New York, 1976. MR0396134
[13] T. Tao. *Topics in Random Matrix Theory*. Amer. Math. Soc., Providence, RI, 2012. MR2906465
[14] R. Vershynin. Spectral norm of products of random and deterministic matrices. *Probab. Theory Related Fields* **150** (2015) 471–509. MR2824864
[15] V. Vu. Spectral norm of random matrices. *Combinatorica* **27** (6) (2007) 721–736. MR2384414
[16] C. Wang, B. S. Jin, Z. D. Bai, K. K. Nair and M. C. Harding. Strong limit of the extreme eigenvalues of a symmetrized auto-cross covariance matrix. Preprint, 2013. Available at arXiv:1312.2277.
[17] Q. W. Wang and J. F. Yao. On singular values distribution of a large auto-covariance matrix in the ultra-dimensional regime. Preprint, 2015. Available at arXiv:1501.06641.
[18] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)* **67** (1958) 325–327. MR0095527
[19] Y. Q. Yin, Z. D. Bai and P. R. Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** (1988) 509–521. MR0950344