ON THE LIMITING PROBABILITY DISTRIBUTION OF A TRANSITION PROBABILITY TENSOR

WEN LI[*] AND MICHAEL K. NG[†]

Abstract. In this paper we propose and develop an iterative method to calculate a limiting probability distribution vector of a transition probability tensor $\mathcal{P}$ arising from a higher-order Markov chain. In the model, the computation of such limiting probability distribution vector $\mathbf{x}$ can be formulated as a Z-eigenvalue problem $\mathcal{P}\mathbf{x}^{m-1} = \mathbf{x}$ associated with the eigenvalue 1 of $\mathcal{P}$ where all the entries of $\mathbf{x}$ are required to be non-negative and its summation must be equal to one. This is an analog of the matrix case for a limiting probability vector of a transition probability matrix arising from the first order Markov chain. We show that if $\mathcal{P}$ is a transition probability tensor, then solutions of this Z-eigenvalue problem exist. When $\mathcal{P}$ is irreducible, all the entries of solutions are positive. With some suitable conditions of $\mathcal{P}$, the limiting probability distribution vector is even unique. Under the same uniqueness assumption, the linear convergence of the iterative method can be established. Numerical examples are presented to illustrate the theoretical results of the proposed model and the iterative method.

Key words. limiting probability distribution vector, transition probability tensor, non-negative tensor, Z-eigenvalue, iterative method, higher-order Markov chains.

1. Introduction.

1.1. Markov Chains. The first order Markov chain concerns about a sequence of random variables, which correspond to the states of a certain system, in such a way that the state at one time epoch depends only on the state in the previous time epoch. We consider a stochastic process $\{X_t, t = 0, 1, 2, \ldots\}$ that takes on a finite set $\{1, 2, \cdots, n\} \equiv \langle n \rangle$. An element in $\langle n \rangle$ refers to a state of the system.

DEFINITION 1.1. *Suppose there is a fixed probability $p_{i,j}$ independent of time such that*

$$\mathrm{Prob}(X_{t+1} = i | X_t = j, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) = \mathrm{Prob}(X_{t+1} = i | X_t = j) = p_{i,j}$$

where $i, j, i_0, i_1, \ldots, i_{t-1} \in \langle n \rangle$. Then this $\{X_t\}$ is called a Markov chain process.

The probability $p_{i,j}$ represents the probability that the process will make a transition to the state i given that currently the process is in the state j. Clearly one has

$$p_{i,j} \geq 0, \quad \sum_{i=1}^{n} p_{i,j} = 1, \quad j = 1, \ldots, n.$$

The matrix $P = (p_{i,j})$ is called the one-step transition probability matrix of the process[1]. Let $\mathbf{x}_t = (x_1(t), x_2(t), \cdots, x_n(t))^T$ be the probability distribution vector

[1]We note that the sum of each column of the transition probability matrix is equal to one. It may be different from the probability context where the sum of each row is set to be one in the notation.

of the states in a Markov chain process at the tth transition. Here $(\cdot)^T$ denotes the transpose of a vector.

It is easy to check that $\mathbf{x}_{t+1} = P\mathbf{x}_t$ and $\mathbf{x}_{t+1} = P^{t+1}\mathbf{x}_0$, where $\mathbf{x}_0$ is the initial probability distribution vector. If we let

$$\lim_{t \to \infty} \mathbf{x}_t = \bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n)^T \quad \text{or} \quad \lim_{t \to \infty} x_i(t) = \bar{x}_i,$$

then

$$\bar{\mathbf{x}} = \lim_{t \to \infty} \mathbf{x}_t = \lim_{t \to \infty} P\mathbf{x}_{t-1} = P\bar{\mathbf{x}}. \tag{1.1}$$

A vector $\bar{\mathbf{x}}$ is said to be a limiting or stationary probability distribution of a finite Markov chain having n states with $\bar{x}_i \geq 0$ for all i, $\sum_{i=1}^n \bar{x}_i = 1$ and $P\bar{\mathbf{x}} = \bar{\mathbf{x}}$. The vector $\bar{\mathbf{x}}$ means that in the long run, the probability that the process in the state i is given by $\bar{x}_i$. We remark that the limiting/stationary probability distribution vector $\bar{\mathbf{x}}$ can be given by the normalized eigenvector associated with the largest eigenvalue of P being equal to 1. For ease of presentation, we refer that a positive (or nonnegative) vector means all its entries are positive (or non-negative). It is denoted by $\mathbf{x} > 0$ (or $\mathbf{x} \geq 0$).

THEOREM 1.2. *Suppose $P = (p_{i,j})$ is a transition probability matrix, then there exists a non-negative vector $\bar{\mathbf{x}}$ such that $P\bar{\mathbf{x}} = \bar{\mathbf{x}}$. In particular, if P is irreducible, then $\bar{\mathbf{x}}$ must be positive and unique. When P is primitive (it must be irreducible), (1.1) is satisfied for any initial distribution vector $\mathbf{x}_0$.*

Interested readers can consult the book by Ross [21] for more detailed information.

1.2. Higher-order Markov Chains. There are many situations that one would like to employ higher-order Markov chain models as a mathematical tool to analyze data sequences in different application areas, see the examples in [5]. The $(m-1)^{\text{th}}$ order Markov chain model is used to fit the observed data through the calculation of the $(m-1)^{th}$ order transition probabilities:

$$0 \leq p_{i_1, i_2, \cdots, i_m} = \text{Prob}(X_{t+1} = i_1 | X_t = i_2, \ldots, X_{t-m+2} = i_m) \leq 1 \tag{1.2}$$

where $i_1, i_2, \ldots, i_m \in \langle n \rangle$, and

$$\sum_{i_1=1}^n p_{i_1, i_2, \cdots, i_m} = 1. \tag{1.3}$$

The probability $p_{i_1, i_2, \cdots, i_m}$ represents the probability that the process will make a transition to the state i_1 given that currently the process is in the state i_2 and previously the process are in the states $i_3, \cdots, i_m$. It is clear when $m = 2$, the situation reduces to the first order Markov chain in Definition 1.1.

A number of applications can be found in the literature, see for instance [5, 13, 19, 20, 22]. For example, a higher-order Markov chain model has been used in fitting observed data and apply to the wind turbine design [19]. Alignment of sequences is an important topic in DNA sequence analysis. It involves searching of patterns in a DNA sequence of huge size [5, 19, 20]. A higher-order Markov model can be employed to study the growth of some polymer chains due to steric effect [8]. A higher-order Markov chain model can be built for a web server log file to be preprocessed into a collection of user sessions [23]. By using citations among researchers, we can construct higher-order Markov chains with respect their order of citations, and employ

limiting probability distribution vectors to study their citation ranking [10]. In [15, 11], co-ranking schemes for objects and relations in multi-relational or tensor data are developed. Their idea [15, 11] is similar to the PageRank algorithm [17] where the limiting probability vector is computed based on hyperlinks among webpages.

In these applications and many others, one would like to characterize data sequences for the purpose of comparison and classification; or to model data sequences and hence to make predictions in the control and planning process. It has been shown that higher-order Markov chain models can be a promising approach for different applications, and their limiting probability distributions can play an important role in the analysis, see for instance in [5] and therein references.

1.3. The Problem. In this paper, we are interested to study a limiting probability distribution of a higher-order Markov chain similar to the results in Theorem 1.1. In the literature, limiting probability distributions of some specific higher-order Markov chains have been studied in [19, 1, 5]. Their idea is to approximate a higher-order Markov chain by a linear combination of $p_{i_j, i_{j'}}$ for some j and j'. The resulting transition probability matrix for this approximated higher-order Markov chain is given by a linear combination of transition matrices $(p_{i_j, i_{j'}})$. In this setting, the limiting probability distribution can be obtained by solving the normalized eigenvector associated with the largest eigenvalue 1 of this approximated transition probability matrix.

1.3.1. An Example. Let us consider an example of a second order Markov chain ($m = 3$ and $n = 3$) to illustrate the problem and the motivation. By using (1.2) and (1.3), we can calculate the following probabilities based on the conditional probabilities p_{i_1, i_2, i_3}:

$$\mathbf{x}_{t+1} = P\mathbf{x}_{t,t-1} \tag{1.4}$$

where

$$\mathbf{x}_{t+1} = \begin{pmatrix} \text{Prob}(X_{t+1} = 1) \\ \text{Prob}(X_{t+1} = 2) \\ \text{Prob}(X_{t+1} = 3) \end{pmatrix}, \quad \mathbf{x}_{t,t-1} = \begin{pmatrix} \text{Prob}(X_t = 1, X_{t-1} = 1) \\ \text{Prob}(X_t = 1, X_{t-1} = 2) \\ \text{Prob}(X_t = 1, X_{t-1} = 3) \\ \text{Prob}(X_t = 2, X_{t-1} = 1) \\ \text{Prob}(X_t = 2, X_{t-1} = 2) \\ \text{Prob}(X_t = 2, X_{t-1} = 3) \\ \text{Prob}(X_t = 3, X_{t-1} = 1) \\ \text{Prob}(X_t = 3, X_{t-1} = 2) \\ \text{Prob}(X_t = 3, X_{t-1} = 3) \end{pmatrix}$$

and

$$P = \begin{pmatrix} p_{1,1,1} & p_{1,1,2} & p_{1,1,3} & p_{1,2,1} & p_{1,2,2} & p_{1,2,3} & p_{1,3,1} & p_{1,3,2} & p_{1,3,3} \\ p_{2,1,1} & p_{2,1,2} & p_{2,1,3} & p_{2,2,1} & p_{2,2,2} & p_{2,2,3} & p_{2,3,1} & p_{2,3,2} & p_{2,3,3} \\ p_{3,1,1} & p_{3,1,2} & p_{3,1,3} & p_{3,2,1} & p_{3,2,2} & p_{3,2,3} & p_{3,3,1} & p_{3,3,2} & p_{3,3,3} \end{pmatrix}.$$

Here P is not a square matrix, but its column sum is equal to one as required in (1.3). In the next subsection, we will represent P as a tensor.

In this example, we are interested to determine the probabilities in the limiting situation (in the long run):

$$\bar{\mathbf{x}} = \lim_{t \to \infty} \mathbf{x}_t = P \lim_{t \to \infty} \mathbf{x}_{t-1, t-2} \quad \text{with} \quad \bar{x}_i = \lim_{t \to \infty} \text{Prob}(X_t = i). \tag{1.5}$$

3

According to (1.5), it is necessary to know a joint probability distribution in the limiting situation $\lim_{t\to\infty} \mathbf{x}_{t-1,t-2}$ of the second order Markov chain in order to compute $\lim_{t\to\infty} \mathbf{x}_t$. One solution is to assume that a limiting joint probability distribution of the second order Markov chain is in the Kronecker product form of its limiting probability distribution:

$$\lim_{t\to\infty} \mathbf{x}_{t-1,t-2} = \lim_{t\to\infty} \mathbf{x}_{t-1} \otimes \lim_{t\to\infty} \mathbf{x}_{t-2} = \bar{\mathbf{x}} \otimes \bar{\mathbf{x}}. \tag{1.6}$$

By putting (1.6) into (1.5), we require to solve for the following multivariate polynomial equations: $\bar{\mathbf{x}} = P(\bar{\mathbf{x}} \otimes \bar{\mathbf{x}})$. The Kronecker product form solution has been explored and used in higher singular value decomposition [6]. The main aim of this paper is to study solutions (existence and uniqueness) of this kind of multivariate polynomial equations.

1.4. The Outline. Let us first introduce a tensor to represent a higher-order Markov chain. We consider an mth order n-dimensional tensor $\mathcal{A}$ consisting of n^m entries in $\mathbb{R}$:

$$\mathcal{A} = (a_{i_1,i_2,\cdots,i_m}), \quad a_{i_1,i_2,\cdots,i_m} \in \mathbb{R}, \quad 1 \le i_1, i_2, \cdots, i_m \le n.$$

Here $\mathbb{R}$ is the set of real numbers. $\mathcal{A}$ is called non-negative (or, respectively, positive) if $a_{i_1,i_2,\cdots,i_m} \ge 0$ (or, respectively, $a_{i_1,i_2,\cdots,i_m} > 0$). For an $(m-1)^{th}$ order Markov chain in (1.2) and (1.3), we can consider the conditional probabilities $p_{i_1,i_2,\cdots,i_m}$ as an mth order n-dimensional tensor $\mathcal{P}$ consisting of n^m entries in between 0 and 1. We call $\mathcal{P}$ to be a transition probability tensor arising from a higher-order Markov chain. It is clear that $\mathcal{P}$ is non-negative.

To an n-dimensional column vector $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n$, we can define a tensor multiplication with $\mathbf{x}$:

$$\mathcal{A}\mathbf{x}^{m-1} := \left(\sum_{i_2,\cdots,i_m=1}^{n} a_{i_1,i_2,\cdots,i_m} x_{i_2} \cdots x_{i_m} \right)_{1\le i_1 \le n} \tag{1.7}$$

is an n-dimensional column vector.

By assuming a limiting joint probability distribution of an $(m-1)^{th}$ order Markov chain is in the Kronecker product form of its limiting probability distribution, i.e.,

$$\lim_{t\to\infty} \text{Prob}(X_{t-1}=i_2, X_{t-2}=i_3, \cdots, X_{t-m+1}=i_m) = \prod_{j=2}^{m} \lim_{t\to\infty} \text{Prob}(X_t=i_j), \tag{1.8}$$

We note that

$$
\begin{aligned}
\bar{\mathbf{x}}_{i_1} &= \lim_{t\to\infty} \text{Prob}(X_t = i_1) \\
&= \sum_{i_2,\cdots,i_m=1}^{n} p_{i_1,i_2,\cdots,i_m} \lim_{t\to\infty} \text{Prob}(X_{t-1}=i_2, X_{t-2}=i_3, \cdots, X_{t-m+1}=i_m) \\
&= \sum_{i_2,\cdots,i_m=1}^{n} p_{i_1,i_2,\cdots,i_m} \prod_{j=2}^{m} \lim_{t\to\infty} \text{Prob}(X_t=i_j) \\
&= \sum_{i_2,\cdots,i_m=1}^{n} p_{i_1,i_2,\cdots,i_m} \bar{x}_{i_2} \cdots \bar{x}_{i_m} = (\mathcal{P}\bar{\mathbf{x}}^{m-1})_{i_1},
\end{aligned}
$$

4

where $(\mathcal{P}\mathbf{x}^{m-1})_{i_1}$ is the i_1-th component of $\mathcal{P}\mathbf{x}^{m-1}$. Under a tensor multiplication defined in (1.7), we would like to determine a limiting probability distribution vector of a $(m-1)^{th}$ order Markov chain by solving the following tensor equations:

$$\bar{\mathbf{x}} = \mathcal{P}\bar{\mathbf{x}}^{m-1} \qquad (1.9)$$

By comparing (1.9) with (1.1) and the results in Theorem 1.1., the computation of such limiting probability distribution vector $\bar{\mathbf{x}}$ can be formulated as a Z-eigenvalue problem $\mathcal{P}\mathbf{x}^{m-1} = \mathbf{x}$ associated with the eigenvalue 1 of $\mathcal{P}$ with $\mathbf{x} \geq 0$ and $\|\mathbf{x}\|_1 = 1$. Here $\|\mathbf{x}\|_1$ is the 1-norm of $\mathbf{x}$.

In this paper, we show that if $\mathcal{P}$ is a transition probability tensor, then there exists a nonzero non-negative vector $\bar{\mathbf{x}}$ such that $\mathcal{P}\bar{\mathbf{x}}^{m-1} = \bar{\mathbf{x}}$. When $\mathcal{P}$ is irreducible, the limiting probability distribution vector $\bar{\mathbf{x}}$ must be positive. With some suitable conditions of $\mathcal{P}$, $\bar{\mathbf{x}}$ is even unique. Based on these results, we also develop an iterative method for solving such limiting probability distribution vector. Under the same uniqueness assumption, the linear convergence of the iterative method can be established. Numerical examples are presented to illustrate the theoretical results of the proposed model and the iterative method.

The outline of this paper is given as follows. In Section 2, we analyze existence and uniqueness of limiting probability distribution for a higher-order Markov chain. In Section 3, we present an iterative method and study the convergence of the method. In Section 4, we demonstrate the theoretical results by numerical examples. In Section 5, we give some concluding remarks and mention some future work.

2. The Main Results.

2.1. Irreducible Tensors. We first introduce the concept of irreducible tensors. In [12], Lim introduced the following definition of irreducible tensors.

DEFINITION 2.1. *An mth order n-dimensional tensor $\mathcal{A}$ is called reducible if there exists a nonempty proper index subset $I \subset \{1, 2, \cdots, n\}$ such that*

$$a_{i_1, i_2, \cdots, i_m} = 0, \quad \forall i_1 \in I, \quad \forall i_2, \cdots, i_m \notin I.$$

If $\mathcal{A}$ is not reducible, then we call $\mathcal{A}$ irreducible.

By using the Matlab multi-dimensional array notation, we give examples of irreducible and reducible transition probability tensors.

Example 1: (a 3th order 3-dimensional irreducible transition probability tensor)

$$P(:,:,1) = \begin{pmatrix} 1/5 & 1/3 & 3/10 \\ 2/5 & 1/3 & 2/5 \\ 2/5 & 1/3 & 2/5 \end{pmatrix}, \quad P(:,:,2) = \begin{pmatrix} 1/5 & 1/3 & 2/5 \\ 2/5 & 1/3 & 2/5 \\ 2/5 & 1/3 & 1/5 \end{pmatrix},$$

$$P(:,:,3) = \begin{pmatrix} 1/3 & 1/3 & 1/5 \\ 1/3 & 1/3 & 2/5 \\ 1/3 & 1/3 & 2/5 \end{pmatrix}.$$

As all the entries $a_{i_1, i_2, i_3} > 0$, the transition probability tensor must be irreducible.

Example 2: (a 3th order 3-dimensional reducible transition probability tensor)

$$P(:,:,1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad P(:,:,2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$P(:,:,3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We set $I = \{1\}$, and find that $a_{1,2,3} = 0$, $a_{1,3,2} = 0$, $a_{1,2,2} = 0$ and $a_{1,3,3} = 0$. The transition probability tensor is reducible. For example, when the process is at $X_t = 1$ and $X_{t-1} = 1$, the process is always at $X_{t+1} = 1$, i.e., the states 2 and 3 cannot be reached. In this example, it is easy to check that there exist nonnegative vectors $\bar{\mathbf{x}} = (1,0,0)^T$, $(0,1,0)^T$, and $(0,0,1)^T$ with the sum of entries being equal to 1, such that $\mathcal{P}\bar{\mathbf{x}}^2 = \bar{\mathbf{x}}$. It is clear that solutions of $\mathcal{P}\mathbf{x}^2 = \mathbf{x}$ are not unique and positive.

2.2. Theoretical Analysis. It is straightforward to show that when $x_i \geq 0$ and $\sum_i x_i = 1$, $(\mathcal{P}\mathbf{x}^{m-1})_i$ must be nonnegative, and $\sum_i (\mathcal{P}\mathbf{x}^{m-1})_i = 1$. We can interpret that the probability vector is preserved after the transition probability calculation via $\mathcal{P}$. This fact is the same as that in the first order Markov chain. Next we state the main results of this paper.

THEOREM 2.2. *If $\mathcal{P}$ is a non-negative tensor of order m and dimension n with (1.2) and (1.3), then there exists a nonzero non-negative vector $\bar{\mathbf{x}}$ such that $\mathcal{P}\bar{\mathbf{x}}^{m-1} = \bar{\mathbf{x}}$ and $\sum_{i=1}^{n} \bar{x}_i = 1$. In particular, if $\mathcal{P}$ is irreducible, then $\bar{\mathbf{x}}$ must be positive.*

Proof. The problem can be reduced to a fixed point problem as follows. Let

$$\Omega = \{(x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n \mid x_i \geq 0, \ 1 \leq i \leq n, \ \sum_{i=1}^{n} x_i = 1\}.$$

It is clear that Ω is a closed and convex set. We define the following map $\Phi : \Omega \to \Omega$:

$$(\Phi(\mathbf{x}))_i = (\mathcal{P}\mathbf{x}^{m-1})_i, \quad 1 \leq i \leq n. \tag{2.1}$$

It is clear that Φ is well-defined and continuous. According to the Brouwer Fixed Point Theorem, there exists $\bar{\mathbf{x}} \in \Omega$ such that $\Phi(\bar{\mathbf{x}}) = \bar{\mathbf{x}}$.

Next we would like to show that $\bar{\mathbf{x}}$ is positive when $\mathcal{P}$ is irreducible. Assume that $\bar{\mathbf{x}}$ is not positive, i.e., there exist some entries of $\bar{\mathbf{x}}$ are zero. Let $I = \{i | \bar{x}_i = 0\}$. It is obvious that I is a proper subset of $\{1, 2, \cdots, n\}$. Let $\delta = \min\{\bar{x}_j | j \notin I\}$. We must have $\delta > 0$. Since $\bar{\mathbf{x}}$ satisfies $\mathcal{P}\bar{\mathbf{x}}^{m-1} = \bar{\mathbf{x}}$, we have

$$\sum_{i_2, \cdots, i_m = 1}^{n} p_{i,i_2,\cdots,i_m} \bar{x}_{i_2} \cdots \bar{x}_{i_m} = \bar{x}_i = 0, \quad \forall i \in I.$$

It follows that

$$\delta^{m-1} \sum_{i_2,\cdots,i_m \notin I} p_{i,i_2,\cdots,i_m} \leq \sum_{i_2,\cdots,i_m \notin I} p_{i,i_2,\cdots,i_m} \bar{x}_{i_2} \cdots \bar{x}_{i_m} \leq 0, \quad \forall i \in I.$$

Hence we have $p_{i,i_2,\cdots,i_m} = 0$ for all $i \in I$ and for all $i_2, \cdots i_m \notin I$, i.e., $\mathcal{P}$ is reducible. This is a contradiction, and the results follow. □

Next we show that with some suitable conditions of $\mathcal{P}$, $\bar{\mathbf{x}}$ is unique in Theorem 2.2. We first consider the case for tensors of order 3. Let S be a proper subset of $\langle n \rangle$ and S' be its complementary set in $\langle n \rangle$, i.e., $S' = \langle n \rangle \backslash S$. For $\mathcal{P} = (p_{i_1,i_2,i_3})$, let

$$\gamma := \min_{S \subset \langle n \rangle} \left\{ \min_{i_3 \in \langle n \rangle} \left(\min_{i_2 \in S} \sum_{i \in S'} p_{i,i_2,i_3} + \min_{i_2 \in S'} \sum_{i \in S} p_{i,i_2,i_3} \right) + \right.$$

6

$$\min_{i_2\in\langle n\rangle}\left(\min_{i_3\in S}\sum_{i\in S'}p_{i,i_2,i_3}+\min_{i_3\in S'}\sum_{i\in S}p_{i,i_2,i_3}\right)\right\}. \tag{2.2}$$

According to the above definition, γ is bounded above by $\sum_{i\in S'}p_{i,i_2,i_3}+\sum_{i\in S}p_{i,i_2,i_3}+\sum_{i\in S'}p_{i,i_2,i_3}+\sum_{i\in S}p_{i,i_2,i_3}$, i.e., γ is always less than or equal to 2.

THEOREM 2.3. *Suppose $\mathcal{P}$ is a non-negative tensor of order 3 and dimension n with (1.2) and (1.3). If $\gamma > 1$, then the nonzero non-negative vector $\bar{\mathbf{x}}$ in Theorem 2.2 is unique.*

Proof. Assume that there exist two positive vectors $\mathbf{x} = [x_1, x_2, \cdots, x_n]$ and $\mathbf{y} = [y_1, y_2, \cdots, y_n]$ with $\mathbf{x} \neq \mathbf{y}$ such that $\mathbf{x} = \mathcal{P}\mathbf{x}^2$ and $\mathbf{y} = \mathcal{P}\mathbf{y}^2$. Let $V = \{i : x_i > y_i\}$ and $V' = \{i : x_i \leq y_i\}$. It is clear that V and V' are nonempty and are not equal to $\langle n\rangle$, and $V \cup V' = \langle n\rangle$.

We first note that

$$x_i - y_i = \sum_{i_2,i_3}p_{i,i_2,i_3}(x_{i_2}x_{i_3}-y_{i_2}y_{i_3}) = \sum_{i_2,i_3}p_{i,i_2,i_3}[(x_{i_2}-y_{i_2})x_{i_3}+(x_{i_3}-y_{i_3})y_{i_2}] \tag{2.3}$$

By (2.3), we have

$$\sum_{i\in V}(x_i-y_i) = \sum_{i_2,i_3}\sum_{i\in V}p_{i,i_2,i_3}(x_{i_2}-y_{i_2})x_{i_3} + \sum_{i_2,i_3}\sum_{i\in V}p_{i,i_2,i_3}(x_{i_3}-y_{i_3})y_{i_2} \tag{2.4}$$

Notice that $\sum_{i=1}^{n}(x_i-y_i) = 0$, and therefore we have

$$\sum_{i\in V}(x_i-y_i) = \sum_{i\in V'}(y_i-x_i). \tag{2.5}$$

We note that

$$\sum_{i_2,i_3}\sum_{i\in V}p_{i,i_2,i_3}(x_{i_2}-y_{i_2})x_{i_3}$$

$$=\sum_{i_2\in V}\sum_{i_3}\sum_{i\in V}p_{i,i_2,i_3}(x_{i_2}-y_{i_2})x_{i_3} + \sum_{i_2\in V'}\sum_{i_3}\sum_{i\in V}p_{i,i_2,i_3}(x_{i_2}-y_{i_2})x_{i_3}$$

$$\leq \sum_{i_2\in V}\sum_{i_3}\left(\max_{i_2\in V}\sum_{i\in V}p_{i,i_2,i_3}\right)(x_{i_2}-y_{i_2})x_{i_3} - \sum_{i_2\in V'}\sum_{i_3}\sum_{i\in V}p_{i,i_2,i_3}(y_{i_2}-x_{i_2})x_{i_3}$$

$$= \sum_{i_2\in V'}\sum_{i_3}\left(\max_{i_2\in V}\sum_{i\in V}p_{i,i_2,i_3}\right)(y_{i_2}-x_{i_2})x_{i_3} - \sum_{i_2\in V'}\sum_{i_3}\sum_{i\in V}p_{i,i_2,i_3}(y_{i_2}-x_{i_2})x_{i_3}$$

$$= \sum_{i_2\in V'}\sum_{i_3}\left(\max_{i_2\in V}\sum_{i\in V}p_{i,i_2,i_3}-\sum_{i\in V}p_{i,i_2,i_3}\right)(y_{i_2}-x_{i_2})x_{i_3}$$

$$\leq \max_{i_2\in V',i_3\in\langle n\rangle}\left(\max_{i_2\in V}\sum_{i\in V}p_{i,i_2,i_3}-\sum_{i\in V}p_{i,i_2,i_3}\right)\sum_{i_2\in V'}\sum_{i_3}(y_{i_2}-x_{i_2})x_{i_3}.$$

By using (2.5), we obtain the following inequality:

$$\sum_{i_2,i_3}\sum_{i\in V}p_{i,i_2,i_3}(x_{i_2}-y_{i_2})x_{i_3} \leq \max_{i_3\in\langle n\rangle}\left(\max_{i_2\in V}\sum_{i\in V}p_{i,i_2,i_3}-\min_{i_2\in V'}\sum_{i\in V}p_{i,i_2,i_3}\right)\sum_{i_2\in V}(x_{i_2}-y_{i_2}).$$

$$\tag{2.6}$$

Similarly, we have

$$\sum_{i_2,i_3}\sum_{i\in V} p_{i,i_2,i_3}(x_{i_3}-y_{i_3})y_{i_2} \le \max_{i_2\in\langle n\rangle}\left(\max_{i_3\in V}\sum_{i\in V}p_{i,i_2,i_3}-\min_{i_3\in V'}\sum_{i\in V}p_{i,i_2,i_3}\right)\sum_{i_3\in V}(x_{i_3}-y_{i_3}),$$

which together with (2.4) and (2.6) gives

$$1 \le \max_{i_3\in\langle n\rangle}\left(\max_{i_2\in V}\sum_{i\in V}p_{i,i_2,i_3}-\min_{i_2\in V'}\sum_{i\in V}p_{i,i_2,i_3}\right)+\max_{i_2\in\langle n\rangle}\left(\max_{i_3\in V}\sum_{i\in V}p_{i,i_2,i_3}-\min_{i_3\in V'}\sum_{i\in V}p_{i,i_2,i_3}\right)$$

$$=\max_{i_3\in\langle n\rangle}\left(\max_{i_2\in V}\left(1-\sum_{i\in V'}p_{i,i_2,i_3}\right)-\min_{i_2\in V'}\sum_{i\in V}p_{i,i_2,i_3}\right)+$$

$$\max_{i_2\in\langle n\rangle}\left(\max_{i_3\in V}\left(1-\sum_{i\in V'}p_{i,i_2i_3}\right)-\min_{i_3\in V'}\sum_{i\in V}p_{i,i_2,i_3}\right)$$

$$=\max_{i_3\in\langle n\rangle}\left(1-\min_{i_2\in V}\sum_{i\in V'}p_{i,i_2,i_3}-\min_{i_2\in V'}\sum_{i\in V}p_{i,i_2,i_3}\right)+$$

$$\max_{i_2\in\langle n\rangle}\left(1-\min_{i_3\in V}\sum_{i\in V'}p_{i,i_2i_3}-\min_{i_3\in V'}\sum_{i\in V}p_{i,i_2,i_3}\right)$$

$$=2-\min_{i_3\in\langle n\rangle}\left(\min_{i_2\in V}\sum_{i\in V'}p_{i,i_2,i_3}+\min_{i_2\in V'}\sum_{i\in V}p_{i,i_2,i_3}\right)-$$

$$\min_{i_2\in\langle n\rangle}\left(\min_{i_3\in V}\sum_{i\in V'}p_{i,i_2,i_3}+\min_{i_3\in V'}\sum_{i\in V}p_{i,i_2,i_3}\right),$$

i.e.,

$$1 \ge \min_{i_3\in\langle n\rangle}\left(\min_{i_2\in V}\sum_{i\in V'}p_{i,i_2,i_3}+\min_{i_2\in V'}\sum_{i\in V}p_{i,i_2,i_3}\right)+\min_{i_2\in\langle n\rangle}\left(\min_{i_3\in V}\sum_{i\in V'}p_{i,i_2,i_3}+\min_{i_3\in V'}\sum_{i\in V}p_{i,i_2,i_3}\right).$$

It implies that $\gamma \le 1$, but this contradicts the assumption, the result follows. $\square$

Remark 1: The condition that $\gamma > 1$ cannot be omitted. By considering Example 2 in Section 2.1, $\gamma \le 1$ for this tensor. We have shown that there are three non-negative vectors $\mathbf{x} = (1,0,0)^T$, $(0,1,0)^T$, or $(0,0,1)^T$ such that $\mathcal{P}\mathbf{x}^2 = \mathbf{x}$.

Remark 2: We note that the condition in Theorem 2.3 can be achieved by using the following condition: $\delta_3 > \frac{1}{2}$ where

$$\delta_3 := \min_{S\subset\langle n\rangle}\left\{\min_{i_2,i_3\in\langle n\rangle}\sum_{i\in S'}p_{i,i_2,i_3}+\min_{i_2,i_3\in\langle n\rangle}\sum_{i\in S}p_{i,i_2,i_3}\right\}. \tag{2.7}$$

It is easy to check that δ_3 is always less than or equal to 1, and $\gamma \ge 2\delta_3$. Thus when $\delta_3 > 1/2$, we obtain $\gamma > 1$.

Remark 3: Here we further give a special kind of tensors satisfying the required condition in Theorem 2.3:

$$|p_{i,i_2,i_3}-p_{i,j_2,j_3}| < \frac{1}{n}, \quad \forall i, i_2, i_3, j_2, j_3 \in \langle n\rangle. \tag{2.8}$$

Without loss of generality, we assume $|S| \leq \frac{n}{2}$. Let

$$\min_{i_2,i_3 \in \langle n \rangle} \sum_{i \in S'} p_{i,i_2,i_3} = \sum_{i \in S'} p_{i,k_2,k_3}.$$

By (2.8), for such k_2 and k_3, we have

$$p_{i,i_2,i_3} > p_{i,k_2,k_3} - \frac{1}{n}, \quad \forall i, i_2, i_3,$$

this implies that

$$\min_{i_2,i_3 \in \langle n \rangle} \sum_{i \in S'} p_{i,i_2,i_3} + \min_{i_2,i_3 \in \langle n \rangle} \sum_{i \in S} p_{i,i_2,i_3} > \sum_{i \in S'} p_{i,k_2,k_3} + \sum_{i \in S} (p_{i,k_2,k_3} - \frac{1}{n}) \geq \frac{1}{2}.$$

We further note that when n is odd, the condition in (2.8) can be changed to

$$|p_{i,i_2,i_3} - p_{i,j_2,j_3}| \leq \frac{1}{n}, \quad \forall i, i_2, i_3, j_2, j_3 \in \langle n \rangle.$$

Because when n is odd, we find that either $|S|$ or $|S'|$ is smaller than $n/2$. Without loss of generality, we let $|S| < n/2$. It follows that

$$\min_{i_2,i_3 \in \langle n \rangle} \sum_{i \in S'} p_{i,i_2,i_3} + \min_{i_2,i_3 \in \langle n \rangle} \sum_{i \in S} p_{i,i_2,i_3} \geq \sum_{i \in S'} p_{i,k_2,k_3} + \sum_{i \in S} (p_{i,k_2,k_3} - \frac{1}{n}) > \frac{1}{2}.$$

According to Remark 2, we know that for the above transition probability tensor, the unique probability vector can be determined.

Let us consider Example 1 in Section 2.1, this irreducible transition probability tensor satisfies the condition (2.8). The unique probability vector is given by $\bar{\mathbf{x}} = (0.2970, 0.3688, 0.3342)^T$. By observing this example, we can interpret the condition $|p_{i,i_2,i_3} - p_{i,j_2,j_3}| \leq \frac{1}{3}$ requires that the difference between the transition probabilities from $(X_t = i_2, X_{t-1} = i_3)$ and $(X_t = j_2, X_{t-1} = j_3)$ should be small enough. This allows the uniqueness of a fixed point in the dynamics of limiting probabilities calculation among different possible states.

Next we consider tensors with orders being greater than 3. A similar condition in (2.7) can be used. For $\mathcal{P} = (p_{i_1,i_2,\cdots,i_m})$, let

$$\delta_m := \min_S \left\{ \min_{i_2,\cdots,i_m \in \langle n \rangle} \sum_{i \in S'} p_{i,i_2,\cdots,i_m} + \min_{i_2,\cdots,i_m \in \langle n \rangle} \sum_{i \in S} p_{i,i_2,\cdots,i_m} \right\}. \tag{2.9}$$

It is easy to check that δ_m is always less than or equal to 1. Similar arguments in the proof of Theorem 2.3 can be employed to show the following theorem.

THEOREM 2.4. *Suppose $\mathcal{P}$ is a non-negative tensor of order m and dimension n with (1.2) and (1.3). If $\delta_m > \frac{m-2}{m-1}$, then the nonzero non-negative vector $\bar{\mathbf{x}}$ in Theorem 2.2 is unique.*

Proof. Assume that there exist two positive vectors $\mathbf{x} = [x_1, x_2, \cdots, x_n]$ and $\mathbf{y} = [y_1, y_2, \cdots, y_n]$ with $\mathbf{x} \neq \mathbf{y}$ such that $\mathbf{x} = \mathcal{P}\mathbf{x}^{m-1}$ and $\mathbf{y} = \mathcal{P}\mathbf{y}^{m-1}$. Let $V = \{i : x_i > y_i\}$ and $V' = \{i : x_i \leq y_i\}$. By the assumption we have $V_i \neq \emptyset$ for $i = 1, 2$ and $V \cup V' = \langle n \rangle$. We note that

$$(x_i - y_i)$$

9

$$= \sum_{i_2,\cdots,i_m} p_{i,i_2,\cdots,i_m}(x_{i_2}\cdots x_{i_m} - y_{i_2}\cdots y_{i_m})$$

$$= \sum_{i_2,\cdots,i_m} p_{i,i_2,\cdots,i_m}[(x_{i_2}-y_{i_2})x_{i_3}\cdots x_{i_m} + \cdots + (x_{i_m}-y_{i_m})y_{i_2}\cdots y_{i_{m-1}}].$$

$$(2.10)$$

It implies that

$$\sum_{i\in V}(x_i - y_i)$$
$$= \sum_{i_2,\cdots,i_m}\sum_{i\in V} p_{i,i_2,\cdots,i_m}[(x_{i_2}-y_{i_2})x_{i_3}\cdots x_{i_m} + \cdots + (x_{i_m}-y_{i_m})y_{i_2}\cdots y_{i_{m-1}}]$$

$$(2.11)$$

Notice that $\sum_{i=1}^{n}(x_i - y_i) = 0$, therefore again we have (2.5). By (2.5) we have

$$\sum_{i_2,\cdots,i_m}\sum_{i\in V} p_{i,i_2,\cdots,i_m}(x_{i_2}-y_{i_2})x_{i_3}\cdots x_{i_m}$$

$$= \sum_{i_2\in V}\sum_{i_3,\cdots,i_m}\sum_{i\in V} p_{i,i_2,\cdots,i_m}(x_{i_2}-y_{i_2})x_{i_3}\cdots x_{i_m} +$$
$$\sum_{i_2\in V'}\sum_{i_3,\cdots,i_m}\sum_{i\in V} p_{i,i_2,\cdots,i_m}(x_{i_2}-y_{i_2})x_{i_3}\cdots x_{i_m}$$

$$\leq \sum_{i_2\in V}\sum_{i_3,..,i_m}\left(\max_{i_2\in V,i_3,\cdots,i_m\in\langle n\rangle}\sum_{i\in V} p_{i,i_2,\cdots,i_m}\right)(x_{i_2}-y_{i_2})x_{i_3}\cdots x_{i_m} -$$
$$\sum_{i_2\in V'}\sum_{i_3,\cdots,i_m}\sum_{i\in V} p_{i,i_2,\cdots,i_m}(y_{i_2}-x_{i_2})x_{i_3}\cdots x_{i_m}$$

$$= \sum_{i_2\in V'}\sum_{i_3,..,i_m}\left(\max_{i_2\in V,i_3,\cdots,i_m\in\langle n\rangle}\sum_{i\in V} p_{i,i_2,\cdots,i_m}\right)(y_{i_2}-x_{i_2})x_{i_3}\cdots x_{i_m} -$$
$$\sum_{i_2\in V'}\sum_{i_3,\cdots,i_m}\sum_{i\in V} p_{i,i_2,\cdots,i_m}(y_{i_2}-x_{i_2})x_{i_3}\cdots x_{i_m}$$

$$= \sum_{i_2\in V'}\sum_{i_3,..,i_m}\left(\max_{i_2\in V,i_3,\cdots,i_m\in\langle n\rangle}\sum_{i\in V} p_{i,i_2,\cdots,i_m} - \sum_{i\in V} p_{i,i_2,\cdots,i_m}\right)(y_{i_2}-x_{i_2})x_{i_3}\cdots x_{i_m}$$

$$\leq \left(\max_{i_2\in V,i_3,\cdots,i_m\in\langle n\rangle}\sum_{i\in V} p_{i,i_2,\cdots,i_m} - \min_{i_2\in V',i_3,\cdots,i_m\in\langle n\rangle}\sum_{i\in V} p_{i,i_2,\cdots,i_m}\right) \times$$
$$\sum_{i_2\in V'}\sum_{i_3,\cdots,i_m}(y_{i_2}-x_{i_2})x_{i_3}\cdots x_{i_m}$$

$$\leq \left(\max_{i_2,\cdots,i_m\in\langle n\rangle}\sum_{i\in V} p_{i,i_2,\cdots,i_m} - \min_{i_2,\cdots,i_m\in\langle n\rangle}\sum_{i\in V} p_{i,i_2,\cdots,i_m}\right)\sum_{i_2\in V'}(y_{i_2}-x_{i_2})$$

$$= \left(\max_{i_2,\cdots,i_m\in\langle n\rangle}\sum_{i\in V} p_{i,i_2,\cdots,i_m} - \min_{i_2,\cdots,i_m\in\langle n\rangle}\sum_{i\in V} p_{i,i_2,\cdots,i_m}\right)\sum_{i\in V}(x_i - y_i). \qquad (2.12)$$

Similarly, for any i_k, we can derive

$$\sum_{i_2,\cdots,i_m} \sum_{i \in V} p_{i,i_2,\cdots,i_m}(x_{i_k} - y_{i_k})y_{i_2} \cdots y_{i_{k-1}}x_{i_{k+1}} \cdots x_{i_m}$$

$$\leq \left(\max_{i_2,\cdots,i_m \in \langle n \rangle} \sum_{i \in V} p_{i,i_2,\cdots,i_m} - \min_{i_2,\cdots,i_m \in \langle n \rangle} \sum_{i \in V} p_{i,i_2,\cdots,i_m} \right) \sum_{i \in V}(x_i - y_i). \quad (2.13)$$

We note that there are $(m-1)$ terms in (2.11) and their bounds are given by (2.12) and (2.13). Therefore, we obtain the following inequality:

$$\frac{1}{m-1} \leq \max_{i_2,\cdots,i_m \in \langle n \rangle} \sum_{i \in V} p_{i,i_2,\cdots,i_m} - \min_{i_2,\cdots,i_m \in \langle n \rangle} \sum_{i \in V} p_{i,i_2,\cdots,i_m}. \quad (2.14)$$

Since

$$\max_{i_2,\cdots,i_m \in \langle n \rangle} \sum_{i \in V} p_{i,i_2,\cdots,i_m} = 1 - \min_{i_2,\cdots,i_m \in \langle n \rangle} \sum_{i \in V'} p_{i,i_2,\cdots,i_m},$$

it follows from (2.14) that

$$\min_{i_2,\cdots,i_m \in \langle n \rangle} \sum_{i \in V} p_{i,i_2,\cdots,i_m} + \min_{i_2,\cdots,i_m \in \langle n \rangle} \sum_{i \in V'} p_{i_i,i_2,\cdots,i_m} \leq \frac{m-2}{m-1},$$

i.e., $\delta_m \leq \frac{m-2}{m-1}$ which contradicts the assumption. This proves the theorem. $\square$

Remark 4: It may be difficult to estimate quantities γ_m and δ_m. However, similar to (2.8), we provide a special kind of tensors of order m and dimension n satisfying the required condition in Theorem 2.4:

$$|p_{i,i_2,\dots,i_m} - p_{i,j_2,\dots,j_m}| < \frac{2}{n(m-1)}, \quad \forall i, i_2, \dots, i_m, j_2, \dots, j_m \in \langle n \rangle.$$

It is noted that the irreducibility can guarantee the existence of a positive limiting probability distribution (Theorem 2.2). There is no proof for the uniqueness. However, the condition of Theorem 2.3 or the condition of Theorem 2.4 can ensure the uniqueness of the limiting probability distribution. Here we give an example which is reducible but satisfies the condition of Theorem 2.3.

Example 3: (a 3th order 2-dimensional reducible transition probability tensor)

$$P(:,:,1) = \begin{pmatrix} a & c \\ 1-a & 1-c \end{pmatrix} \quad \text{and} \quad P(:,:,2) = \begin{pmatrix} b & 0 \\ 1-b & 1 \end{pmatrix},$$

where

$$a < b + c < 1.$$

Then it is easy to see that

$$\gamma = 2 - (b+c) > 1.$$

We find that the condition in Theorem 2.3 is satisfied. It is easy to check that $\mathcal{P}\mathbf{x}^2 = \mathbf{x}$ has the unique solution $(0,1)^T$.

Remark 5: The condition that $\gamma > 1$ (in Theorem 2.3) or $\delta_m > \frac{m-2}{m-1}$ (in Theorem 2.4) is not a necessary condition. Here we give an example to demonstrate this issue.

Example 4: (a 3th order 2-dimensional irreducible transition probability tensor)

$$P(:,:,1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad P(:,:,2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This transition probability tensor is irreducible. We note that when $(X_t = 1, X_{t-1} = 1)$, X_{t+1} will be equal to 2 (i.e., $p_{2,1,1} = 1$); when $(X_t = 1, X_{t-1} = 2)$, X_{t+1} will be equal to 2 (i.e., $p_{2,1,2} = 1$); when $(X_t = 2, X_{t-1} = 1)$, X_{t+1} will be equal to 1 (i.e., $p_{1,2,1} = 1$); when $(X_t = 2, X_{t-1} = 2)$, X_{t+1} will be equal to 1 (i.e., $p_{1,2,2} = 1$). It is interesting to note that $\gamma = 1$ for this tensor, but we still find that the unique probability vector $\bar{\mathbf{x}}$ satisfying $\bar{\mathbf{x}} = \mathcal{P}\bar{\mathbf{x}}^2$ is given by $(1/2, 1/2)^T$.

In the section of numerical examples (Section 4), we give a 4th order transition probability tensor (vii) that $\delta_4 \leq 2/3$ such that there is a unique positive vector such that (1.9) holds.

2.3. Z-eigenvalue Problem of Transition Probability Tensors. In a Markov chain, the limiting probability distribution can be given by the normalized eigenvector associated with the largest eigenvalue of P being equal to 1 (see Theorem 1.1). In this subsection, we discuss the relation between solutions of $\mathcal{P}\mathbf{x}^{m-1} = \mathbf{x}$ and the eigenvalue problem of transition probability tensors.

The study of eigenpair of a tensor can be found in [18, 12, 4].

DEFINITION 2.5. *Let $\mathcal{A}$ be an mth order n-dimensional tensor and $\mathbb{C}$ be the set of all complex numbers. Assume that $\mathcal{A}\mathbf{x}^{m-1}$ is not identical to zero. We say $(\lambda, \mathbf{x}) \in \mathbb{C} \times (\mathbb{C}^n \backslash \{0\})$ is an H-eigenpair of $\mathcal{A}$ if*

$$\mathcal{A}\mathbf{x}^{m-1} = \lambda \mathbf{x}^{[m-1]}. \tag{2.15}$$

Here, $\mathbf{x}^{[\alpha]} = [x_1^\alpha, x_2^\alpha, \cdots, x_n^\alpha]^T$. On the other hand, we say $(\lambda, \mathbf{x}) \in \mathbb{C} \times (\mathbb{C}^n \backslash \{0\})$ is an Z-eigenpair of $\mathcal{A}$ if

$$\mathcal{A}\mathbf{x}^{m-1} = \lambda \mathbf{x} \quad \text{with} \quad \|\mathbf{x}\|_2 = 1. \tag{2.16}$$

where $\|\mathbf{x}\|_2$ is the Euclidean norm of $\mathbf{x}$.

This definition was introduced by Qi [18] when m is even and $\mathcal{A}$ is symmetric. Independently, Lim [12] gave such a definition but restricted $\mathbf{x}$ to be a real vector and λ to be a real number. For the largest H-eigenvalue of a non-negative tensor, the following Perron-Frobenius theorem was proved by Chang et al [4], Friedland et al. [7] and Lim [12].

THEOREM 2.6. *[12, 4, 7] If $\mathcal{A}$ is an irreducible non-negative tensor of order m and dimension n, then there exist $\lambda_0 > 0$ and $\bar{\mathbf{x}} > 0$ with $\bar{\mathbf{x}} \in \mathbb{R}^n$ such that*

$$\mathcal{A}\bar{\mathbf{x}}^{m-1} = \lambda_0 \bar{\mathbf{x}}^{[m-1]}. \tag{2.17}$$

Moreover, if λ is an eigenvalue with a non-negative eigenvector, then $\lambda = \lambda_0$. If λ is an eigenvalue of $\mathcal{A}$, then $|\lambda| \leq \lambda_0$.

In Theorem 2.6, there is no guarantee that λ_0 is equal to 1 for transition probability tensors. Numerical results in Table 4 given in [16] have shown that λ_0 is not equal to 1. Therefore, the results in Theorem 2.6 are different from a second-order irreducible non-negative tensor (i.e., the first order Markov chain matrix P) that the

12

largest eigenvalue of the first order Markov chain matrix is always equal to one [3], and the corresponding eigenvector $\mathbf{x} > 0$ and is normalized with $P\mathbf{x} = \mathbf{x}$ as stated in Theorem 1.1.

In our setting, the summation of all the entries of $\mathcal{P}\bar{\mathbf{x}}^{m-1}$ and $\bar{\mathbf{x}}$ are both equal to one. It is clear that 1 must be the largest Z-eigenvalue of $\mathcal{P}$ with a non-negative eigenvector $\bar{\mathbf{x}}$ and $\sum_{i=1}^n \bar{x}_i = 1$ as stated in Theorem 2.2. Thus our problem can be formulated as the Z-eigenvalue problem of $\mathcal{P}$:

$$\mathcal{P}\mathbf{x}^{m-1} = \mathbf{x} \quad \text{with } \mathbf{x} \geq 0, \text{ and } \|\mathbf{x}\|_1 = 1. \tag{2.18}$$

This eigenvalue problem is different from the eigenvalue problems in (2.15) and (2.16).

3. The Iterative Method. According to Theorems 2.2-2.4, we propose and study a simple iterative algorithm for computing the limiting probability distribution vector of a transition probability tensor arising from a higher order Markov chain.

    1. The Iterative Algorithm:
    (i) Input $\mathbf{x}_0$ is any n-vector with $\sum_{i=1}^n [\mathbf{x}_0]_i = 1$;
    (ii) Set $t = 1$;
        (a) Compute $\mathbf{x}_t = \mathcal{P}\mathbf{x}_{t-1}^{m-1}$;
        (b) Set $t = t + 1$;
        (c) If $\mathbf{x}_t = \mathbf{x}_{t-1}$, then stop, otherwise goto Step ii(a).

In the algorithm, the subscript index t refers to the iteration number of the iterate $\mathbf{x}_t$. We note that the computation requires several iterations, through the collection to adjust approximate probabilities to more closely reflect their theoretical true values (underlying limiting probability distribution). The iterative method is similar to the power method for computing the eigenvector corresponding to the largest eigenvalue of a matrix [17]. The main computational cost of the algorithm depends on the cost of performing tensor operation. Assume that there are $O(N)$ nonzero entries (sparse data) in $\mathcal{P}$, the cost of this tensor calculation are of $O(N)$ arithmetic operations.

We note that when $x_i \geq 0$ and $\sum_i x_i = 1$, we have the i-th entry of $\mathcal{P}\mathbf{x}^{m-1}$ is greater than or equal to 0 for $1 \leq i \leq n$ and the summation of all entries is equal to 1. Therefore it is not necessary to perform the normalization in the exact arithmetic. However, under the inexact arithmetic computing environment, we can perform the normalization of $\mathbf{x}_t$ after Step ii(a).

Here we use Example 3 in Section 2.2 to demonstrate the method does not converge even for an irreducible transition probability tensor. When the algorithm starts with an initial probability vector $(1,0)^T$, the next iterate of the algorithm is given by $(0,1)^T$, and then the next iterate of the algorithm returns to the same initial probability vector. Thus the algorithm does not converge.

3.1. The Convergence Analysis. The main contribution of this subsection is to show a linear convergence of the iterative algorithm under the same assumption in Theorems 2.3 and 2.4. Let us first consider for tensors of order 3.

THEOREM 3.1. *Suppose $\mathcal{P}$ is a non-negative tensor of order 3 and dimension n with (1.2) and (1.3). Then $\{\mathbf{x}_t\}$ generated by the iterative method, satisfies*

$$\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1 \leq (2 - \gamma)\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1, \quad \forall t = 0, 1, \cdots,$$

where γ is defined in (2.2). If $\gamma > 1$, then $\{\mathbf{x}_t\}$ converges linearly to the unique solution $\bar{\mathbf{x}}$ (which is guaranteed to exist by Theorem 2.3), for any initial distribution vector $\mathbf{x}_0$.

Proof. We consider the iterates at the t-th and the $(t-1)$-th iterations of the iterative method: $\mathbf{x}_{t+1} = [x_{t+1,1}, x_{t+1,2}, \cdots, x_{t+1,n}]^T \neq \mathbf{x}_t = [x_{t,1}, x_{t,2}, \cdots, x_{t,n}]^T$ where $\mathbf{x}_{t+1} = \mathcal{P}\mathbf{x}_t^2$ and $\mathbf{x}_t = \mathcal{P}\mathbf{x}_{t-1}^2$.

Let $V(t) = \{i : x_{t+1,i} > x_{t,i}\}$ and $V'(t) = \{i : x_{t+1,i} \leq x_{t,i}\}$. As both $\mathbf{x}_{t+1}$ and $\mathbf{x}_t$ are positive, and $\sum_{i=1} x_{t+1,i} = \sum_{i=1} x_{t,i} = 1$, we must have $V(t), V'(t) \neq \emptyset$ and $V(t) \cup V'(t) = \langle n \rangle$, otherwise, $\mathbf{x}_{t+1} = \mathbf{x}_t$, and the result follows. It is clear that $V'(t)$ is complementary set of $V(t)$ in $\langle n \rangle$.

By using the similar argument in Theorem 2.3, we have

$$x_{t+1,i} - x_{t,i} = \sum_{i_2,i_3} p_{i,i_2,i_3}((x_{t,i_2} - x_{t-1,i_2})x_{t,i_3} + (x_{t,i_3} - x_{t-1,i_3})x_{t-1,i_2}) \quad (3.1)$$

Therefore, we have

$$\sum_{i \in V(t)} (x_{t+1,i} - x_{t,i}) = \sum_{i_2,i_3} \sum_{i \in V(t)} p_{i,i_2,i_3}(x_{t,i_2} - x_{t-1,i_2})x_{t,i_3} +$$
$$\sum_{i_2,i_3} \sum_{i \in V(t)} p_{i,i_2,i_3}(x_{t,i_3} - x_{t-1,i_3})x_{t-1,i_2} \quad (3.2)$$

$$\sum_{i \in V(t)} (x_{t+1,i} - x_{t,i}) = \sum_{i \in V'(t)} (x_{t,i} - x_{t+1,i}) \quad (3.3)$$

$$\sum_{i \in V(t-1)} (x_{t,i} - x_{t-1,i}) = \sum_{i \in V'(t-1)} (x_{t-1,i} - x_{t,i}) \quad (3.4)$$

By using (3.4), we obtain

$$\sum_{i_2,i_3} \sum_{i \in V(t)} p_{i,i_2,i_3}(x_{t,i_2} - x_{t-1,i_2})x_{t,i_3}$$

$$\leq \left(\max_{i_2,i_3 \in \langle n \rangle} \sum_{i \in V(t)} p_{i,i_2,i_3} - \min_{i_2,i_3 \in \langle n \rangle} \sum_{i \in V(t)} p_{i,i_2,i_3} \right) \sum_{i_2 \in V(t-1)} (x_{t,i_2} - x_{t-1,i_2}), (3.5)$$

and

$$\sum_{i_2,i_3} \sum_{i \in V(t)} p_{i,i_2,i_3}(x_{t,i_3} - x_{t-1,i_3})x_{t-1,i_2}$$

$$\leq \left(\max_{i_2,i_3 \in \langle n \rangle} \sum_{i \in V(t)} p_{i,i_2,i_3} - \min_{i_2,i_3 \in \langle n \rangle} \sum_{i \in V(t)} p_{i,i_2,i_3} \right) \sum_{i_3 \in V(t-1)} (x_{t,i_3} - x_{t-1,i_3}),$$

which together with (3.2) and (3.5) gives

$$\sum_{i \in V(t)} (x_{t+1,i} - x_{t,i})$$

$$\leq \left[2 - \left(\min_{i_2,i_3 \in \langle n \rangle} \sum_{i \in V'(t)} p_{i,i_2,i_3} + \min_{i_2,i_3 \in \langle n \rangle} \sum_{i \in V(t)} p_{i,i_2,i_3} \right) - \right.$$
$$\left. \left(\min_{i_2,i_3 \in \langle n \rangle} \sum_{i \in V'(t)} p_{i,i_2,i_3} + \min_{i_2,i_3 \in \langle n \rangle} \sum_{i \in V(t)} p_{i,i_2,i_3} \right) \right] \sum_{i \in V(t-1)} (x_{t,i} - x_{t-1,i})$$

14

$$\leq (2-\gamma) \sum_{i \in V(t-1)} (x_{t,i} - x_{t-1,i}).$$

By using (3.3) and (3.4), we have

$$\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1 = \sum_{i \in V(t)} (x_{t+1,i} - x_{t,i}) + \sum_{i \in V'(t)} (x_{t,i} - x_{t+1,i})$$

$$= 2 \sum_{i \in V(t)} (x_{t+1,i} - x_{t,i})$$

$$\leq 2(2-\gamma) \sum_{i \in V(t-1)} (x_{t,i} - x_{t-1,i})$$

$$= (2-\gamma) \left(\sum_{i \in V(t-1)} (x_{t,i} - x_{t-1,i}) + \sum_{i \in V'(t-1)} (x_{t-1,i} - x_{t,i}) \right)$$

$$= (2-\gamma)\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1.$$

If the iteration does not terminate in a finite number of iterations, i.e., $\mathbf{x}_{t+1} \neq \mathbf{x}_t$ ($t = 0, 1, \cdots,$), then we have

$$\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1 \leq (2-\gamma)^t \|\mathbf{x}_1 - \mathbf{x}_0\|_1.$$

When $\gamma > 1$, it implies that $\{\mathbf{x}_t\}$ converges as $\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1$ converges to zero when t tends to infinity, i.e., the iterative method converges.

By using the above argument again, we can also show that

$$\|\mathbf{x}_{t+1} - \bar{\mathbf{x}}\|_1 \leq (2-\gamma)^{t+1} \|\mathbf{x}_0 - \bar{\mathbf{x}}\|_1,$$

It implies that $\{\mathbf{x}_t\}$ converges to $\bar{\mathbf{x}}$ for any given initial $\mathbf{x}_0$ with a linear convergence rate. $\square$

Here we can make use of the similar argument in Theorem 3.1, we can obtain the convergence theorem for tensors of order m.

THEOREM 3.2. *Suppose $\mathcal{P}$ is a non-negative tensor of order m and dimension n with (1.2) and (1.3). Then $\{\mathbf{x}_t\}$ generated by the iterative method, satisfies*

$$\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1 \leq (m-1)(1 - \delta_m)\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1, \quad \forall t = 0, 1, \cdots,$$

If $\delta_m > \frac{m-2}{m-1}$, then $\{\mathbf{x}_t\}$ converges linearly to the unique solution $\bar{\mathbf{x}}$ (which is guaranteed to exist by Theorem 2.4), for any initial distribution vector $\mathbf{x}_0$.

Proof. The proof is similar to that in Theorem 3.1. The main equality and inequality are given as follows:

$$([x_{t+1}]_i - [x_t]_i) = \sum_{i_2, \cdots, i_m} p_{i,i_2,\cdots,i_m} (x_{t,i_2} \cdots x_{t,i_m} - x_{t-1,i_2} \cdots x_{t-1,i_m})$$

$$= \sum_{i_2,\cdots,i_m} p_{i,i_2,\cdots,i_m} \left\{ (x_{t,i_2} - x_{t-1,i_2}) x_{t,i_3} \cdots x_{t,i_m} + \cdots + \right.$$

$$\left. (x_{t,i_m} - x_{t-1,i_m}) x_{t-1,i_2} \cdots x_{t-1,i_{m-1}} \right\}.$$

and

$$\sum_{i_2,\cdots,i_m} \sum_{i \in V(t)} p_{i,i_2,\cdots,i_m} (x_{t,i_2} - x_{t-1,i_2}) x_{t,i_3} \cdots x_{t,i_m}$$

15

$$\leq \left(\max_{i_2,\cdots,i_m\in\langle n\rangle} \sum_{i\in V(t)} p_{i,i_2,\cdots,i_m} - \min_{i_2,\cdots,i_m\in\langle n\rangle} \sum_{i\in V(t)} p_{i,i_2,\cdots,i_m} \right) \sum_{i\in V(t-1)} (x_{t,i} - x_{t-1,i}).$$

□

Remark 6: Under the assumption of Theorem 3.1 or Theorem 3.2, the iterative method converges linearly to the unique solution of (2.18). In addition, if the transition probability tensor $\mathcal{P}$ is further assumed to be irreducible, then by Theorem 2.2 the iterative method will converge linearly to the positive solution of (2.18).

4. Numerical Examples. In this section, we present numerical examples to demonstrate the theoretical results in Sections 2 and 3.

4.1. Demonstration I. The first two examples comes from DNA sequence data in Tables 6 and 10 of [20]. In these two transition probability tensors, their orders m are 3 and their numbers of states n are 3 by considering three categories ($\{A/G, C, T\}$). By using the Matlab multi-dimensional array notation, the transition probability tensors are given by

$$(i) \quad P(:,:,1) = \begin{pmatrix} 0.6000 & 0.4083 & 0.4935 \\ 0.2000 & 0.2568 & 0.2426 \\ 0.2000 & 0.3349 & 0.2639 \end{pmatrix}, \quad P(:,:,2) = \begin{pmatrix} 0.5217 & 0.3300 & 0.4152 \\ 0.2232 & 0.2800 & 0.2658 \\ 0.2551 & 0.3900 & 0.3190 \end{pmatrix},$$

$$P(:,:,3) = \begin{pmatrix} 0.5565 & 0.3648 & 0.4500 \\ 0.2174 & 0.2742 & 0.2600 \\ 0.2261 & 0.3610 & 0.2900 \end{pmatrix},$$

and

$$(ii) \quad P(:,:,1) = \begin{pmatrix} 0.5200 & 0.2986 & 0.4462 \\ 0.2700 & 0.3930 & 0.3192 \\ 0.2100 & 0.3084 & 0.2346 \end{pmatrix}, \quad P(:,:,2) = \begin{pmatrix} 0.6514 & 0.4300 & 0.5776 \\ 0.1970 & 0.3200 & 0.2462 \\ 0.1516 & 0.2500 & 0.1762 \end{pmatrix},$$

$$P(:,:,3) = \begin{pmatrix} 0.5638 & 0.3424 & 0.4900 \\ 0.2408 & 0.3638 & 0.2900 \\ 0.1954 & 0.2938 & 0.2200 \end{pmatrix}$$

respectively.

By considering three categories ($\{A, C/T, G\}$), we construct a transition probability tensor of order 4 and dimension 3 for the DNA sequence in [14]:

$$(iii) \quad P(:,:,1,1) = \begin{pmatrix} 0.3721 & 0.2600 & 0.4157 \\ 0.4477 & 0.5000 & 0.4270 \\ 0.1802 & 0.2400 & 0.1573 \end{pmatrix}, \quad P(:,:,2,1) = \begin{pmatrix} 0.3692 & 0.2673 & 0.3175 \\ 0.4667 & 0.5594 & 0.5079 \\ 0.1641 & 0.1733 & 0.1746 \end{pmatrix},$$

$$P(:,:,3,1) = \begin{pmatrix} 0.4227 & 0.2958 & 0.2353 \\ 0.4124 & 0.5563 & 0.5588 \\ 0.1649 & 0.1479 & 0.2059 \end{pmatrix}, \quad P(:,:,1,2) = \begin{pmatrix} 0.3178 & 0.2632 & 0.3194 \\ 0.5212 & 0.6228 & 0.5833 \\ 0.1610 & 0.1140 & 0.0972 \end{pmatrix},$$

$$P(:,:,2,2) = \begin{pmatrix} 0.2836 & 0.2636 & 0.3042 \\ 0.5012 & 0.6000 & 0.5250 \\ 0.2152 & 0.1364 & 0.1708 \end{pmatrix}, \quad P(:,:,3,2) = \begin{pmatrix} 0.3382 & 0.2396 & 0.3766 \\ 0.5147 & 0.6406 & 0.4935 \\ 0.1471 & 0.1198 & 0.1299 \end{pmatrix},$$

$$P(:,:,1,3) = \begin{pmatrix} 0.3204 & 0.2985 & 0.3500 \\ 0.4854 & 0.5000 & 0.5000 \\ 0.1942 & 0.2015 & 0.1500 \end{pmatrix}, \quad P(:,:,2,3) = \begin{pmatrix} 0.4068 & 0.2816 & 0.3594 \\ 0.3898 & 0.5143 & 0.4219 \\ 0.2034 & 0.2041 & 0.2188 \end{pmatrix},$$

$$P(:,:,3,3) = \begin{pmatrix} 0.3721 & 0.3529 & 0.3000 \\ 0.5349 & 0.3971 & 0.5500 \\ 0.0930 & 0.2500 & 0.1500 \end{pmatrix}.$$

In [2], it is found that a better model for DNA sequence is obtained by considering both sides, i.e., the left to the right and the right to the left in the DNA sequence together. An example is given in [2] is a transition probability tensor of order 3 and dimension 4 by considering four bases ($\{A, C, G, T\}$):

$$(iv) \quad P(:,:,1) = \begin{pmatrix} 0.2091 & 0.2834 & 0.2194 & 0.1830 \\ 0.3371 & 0.3997 & 0.3219 & 0.3377 \\ 0.3265 & 0.0560 & 0.3119 & 0.2961 \\ 0.1273 & 0.2608 & 0.1468 & 0.1832 \end{pmatrix},$$

$$P(:,:,2) = \begin{pmatrix} 0.1952 & 0.2695 & 0.2055 & 0.1690 \\ 0.3336 & 0.3962 & 0.3184 & 0.3342 \\ 0.2954 & 0.0249 & 0.2808 & 0.2650 \\ 0.1758 & 0.3094 & 0.1953 & 0.2318 \end{pmatrix},$$

$$P(:,:,3) = \begin{pmatrix} 0.3145 & 0.3887 & 0.3248 & 0.2883 \\ 0.0603 & 0.1230 & 0.0451 & 0.0609 \\ 0.3960 & 0.1255 & 0.3814 & 0.3656 \\ 0.2293 & 0.3628 & 0.2487 & 0.2852 \end{pmatrix},$$

$$P(:,:,4) = \begin{pmatrix} 0.1686 & 0.2429 & 0.1789 & 0.1425 \\ 0.3553 & 0.4180 & 0.3402 & 0.3559 \\ 0.3189 & 0.0484 & 0.3043 & 0.2885 \\ 0.1571 & 0.2907 & 0.1766 & 0.2131 \end{pmatrix}.$$

In this case, the higher-order Markov chain is actually a fitted low-parameter version of the original data.

For the four transition probability tensors (i)–(iv), they are irreducible. Also the condition in Theorem 2.4 is satisfied. Each has unique and positive limiting probability distribution vector. In Table 4.1, we list δ_m for these four tensors. It is clear that for (i), (ii) and (iv), the quantity $\delta_3 > 1/2$; and for (iii), the quantity $\delta_4 > 2/3$. According to Theorem 3.2, we know that the iterative method converges linearly to the unique and positive probability vector. Their corresponding theoretical convergence rates $(m-1)(1-\delta_m)$ are also shown in Table 4.1.

In the numerical experiment, we set the stopping criterion of the iterative method to be $\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1 < 10^{-10}$. We plot in Figure 4.1(a) the successive differences $\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1$ (the convergence history) with respect to iteration numbers for the transition probability tensor (ii). We see from the figure that the method converges linearly. Similar observations can be found for the transition probability tensors (i), (iii) and (iv). In Table 4.1, we also show the maximum value of $\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1 / \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1$ over the iterations of the proposed method. We find that the maximum value is always bounded below by $(m-1)(1-\delta_m)$. However, the maximum value is significantly smaller than $(m-1)(1-\delta_m)$, and the iterative method can converge within 20 iterations.

Example	δ_m	$(m-1)(1-\delta_m)$	$\max \dfrac{\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1}{\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1}$	number of iterations
(i)	0.7300	0.5400	0.2321	14
(ii)	0.6472	0.7056	0.0930	9
(iii)	0.7492	0.7524	0.1628	12
(iv)	0.5742	0.8516	0.3599	20

Example	limiting probability distribution vector
(i)	$[0.4963, 0.2349, 0.2688]$
(ii)	$[0.4873, 0.2889, 0.2238]$
(iii)	$[0.3503, 0.4724, 0.1773]$
(iv)	$[0.2395, 0.2869, 0.2447, 0.2289]$

TABLE 4.1

The computed quantities of the examples of tensors (i)-(iv).
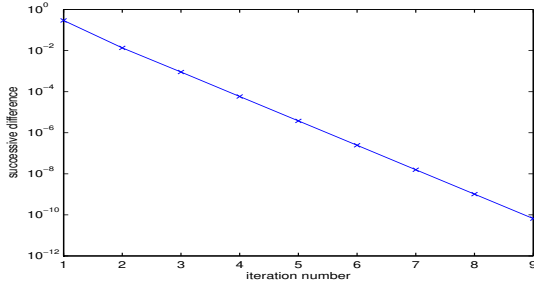
(a)
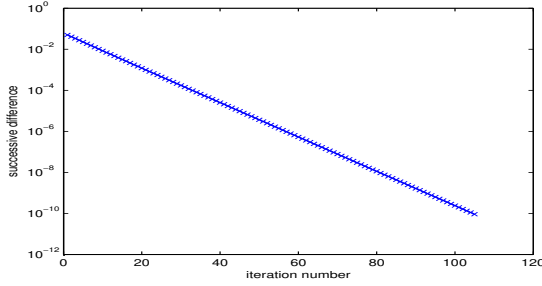
(b)

FIG. 4.1. *The convergence history of the iterative method for (a) the tensor in (ii) and (b) the tensor in (vi).*

4.2. Demonstration II. In this subsection, we show three more examples of transition probability tensors studied in [19, 20].

The first one comes from inter-personal relationships data in [19]. The order m is 3 and the number n of states is 3. By using the Matlab multi-dimensional array notation, the transition probability tensor is given by

$$(v) \quad P(:,:,1) = \begin{pmatrix} 0.5810 & 0.2432 & 0.1429 \\ 0 & 0.4109 & 0.0701 \\ 0.4190 & 0.3459 & 0.7870 \end{pmatrix}, \quad P(:,:,2) = \begin{pmatrix} 0.4708 & 0.1330 & 0.0327 \\ 0.1341 & 0.5450 & 0.2042 \\ 0.3951 & 0.3220 & 0.7631 \end{pmatrix},$$

18

$$P(:,:,3) = \begin{pmatrix} 0.4381 & 0.1003 & 0 \\ 0.0229 & 0.4338 & 0.0930 \\ 0.5390 & 0.4659 & 0.9070 \end{pmatrix}.$$

The second one comes from occupational mobility of physicists data in [19]. The order m is 3 and the number n of states is 3. The transition probability tensor is given by

$$(vi) \quad P(:,:,1) = \begin{pmatrix} 0.9000 & 0.3340 & 0.3106 \\ 0.0690 & 0.6108 & 0.0754 \\ 0.0310 & 0.0552 & 0.6140 \end{pmatrix}, \quad P(:,:,2) = \begin{pmatrix} 0.6700 & 0.1040 & 0.0805 \\ 0.2892 & 0.8310 & 0.2956 \\ 0.0408 & 0.0650 & 0.6239 \end{pmatrix},$$

$$P(:,:,3) = \begin{pmatrix} 0.6604 & 0.0945 & 0.0710 \\ 0.0716 & 0.6133 & 0.0780 \\ 0.2680 & 0.2922 & 0.8510 \end{pmatrix}.$$

The third one comes from wind power data in [19]. The order m is 4 and the number n of states is 4.

$$(vii) \quad P(:,:,1,1) = \begin{pmatrix} 0.8370 & 0.3470 & 0.3105 & 0.3105 \\ 0.1630 & 0.5976 & 0.1316 & 0.0605 \\ 0 & 0.0544 & 0.5328 & 0.0730 \\ 0 & 0 & 0.0252 & 0.5560 \end{pmatrix},$$

$$P(:,:,2,1) = \begin{pmatrix} 0.7984 & 0.1865 & 0.1501 & 0.1835 \\ 0.1802 & 0.7400 & 0.2739 & 0.0254 \\ 0.0214 & 0.0735 & 0.5509 & 0.1114 \\ 0 & 0 & 0.0252 & 0.6797 \end{pmatrix},$$

$$P(:,:,3,1) = \begin{pmatrix} 0.6646 & 0.1746 & 0.1381 & 0.1381 \\ 0.1527 & 0.5873 & 0.1212 & 0.0502 \\ 0.1745 & 0.2298 & 0.7072 & 0.2474 \\ 0.0082 & 0.0082 & 0.0334 & 0.5643 \end{pmatrix},$$

$$P(:,:,4,1) = \begin{pmatrix} 0.6646 & 0.1746 & 0.1381 & 0.1381 \\ 0.1294 & 0.5641 & 0.0980 & 0.0269 \\ 0.0239 & 0.0792 & 0.5567 & 0.0969 \\ 0.1821 & 0.1821 & 0.2073 & 0.7381 \end{pmatrix},$$

$$P(:,:,1,2) = \begin{pmatrix} 0.7085 & 0.2185 & 0.1820 & 0.1820 \\ 0.2770 & 0.7117 & 0.2456 & 0.1745 \\ 0.0145 & 0.0699 & 0.5473 & 0.0875 \\ 0 & 0 & 0.0252 & 0.5560 \end{pmatrix},$$

$$P(:,:,2,2) = \begin{pmatrix} 0.5480 & 0.0580 & 0.0215 & 0.0215 \\ 0.4194 & 0.8540 & 0.3879 & 0.3168 \\ 0.0326 & 0.0880 & 0.5654 & 0.1056 \\ 0 & 0 & 0.0252 & 0.5560 \end{pmatrix},$$

$$P(:,:,3,2) = \begin{pmatrix} 0.5360 & 0.0461 & 0.0096 & 0.0096 \\ 0.2667 & 0.7014 & 0.2353 & 0.1642 \\ 0.1890 & 0.2444 & 0.7218 & 0.2620 \\ 0.0082 & 0.0082 & 0.0334 & 0.5643 \end{pmatrix},$$

$$P(:,:,4,2) = \begin{pmatrix} 0.5360 & 0.0461 & 0.0096 & 0.0096 \\ 0.2434 & 0.6781 & 0.2120 & 0.1409 \\ 0.0384 & 0.0938 & 0.5712 & 0.1114 \\ 0.1821 & 0.1821 & 0.2073 & 0.7381 \end{pmatrix},$$

$$P(:,:,1,3) = \begin{pmatrix} 0.6989 & 0.2089 & 0.1724 & 0.1724 \\ 0.1547 & 0.5894 & 0.1233 & 0.0522 \\ 0.1398 & 0.1951 & 0.6725 & 0.2127 \\ 0.0066 & 0.0066 & 0.0318 & 0.5626 \end{pmatrix},$$

$$P(:,:,2,3) = \begin{pmatrix} 0.5384 & 0.0484 & 0.0119 & 0.0119 \\ 0.2971 & 0.7317 & 0.2656 & 0.1946 \\ 0.1579 & 0.2132 & 0.6906 & 0.2308 \\ 0.0066 & 0.0066 & 0.0318 & 0.5626 \end{pmatrix},$$

$$P(:,:,3,3) = \begin{pmatrix} 0.5265 & 0.0365 & 0 & 0 \\ 0.1444 & 0.5791 & 0.1130 & 0.0419 \\ 0.3142 & 0.3696 & 0.8470 & 0.3872 \\ 0.0148 & 0.0148 & 0.0400 & 0.5709 \end{pmatrix},$$

$$P(:,:,4,3) = \begin{pmatrix} 0.5265 & 0.0365 & 0 & 0 \\ 0.1212 & 0.5558 & 0.0897 & 0.0186 \\ 0.1637 & 0.2190 & 0.6964 & 0.2366 \\ 0.1887 & 0.1887 & 0.2139 & 0.7447 \end{pmatrix},$$

$$P(:,:,1,4) = \begin{pmatrix} 0.6989 & 0.2089 & 0.1724 & 0.1724 \\ 0.1361 & 0.5707 & 0.1047 & 0.0336 \\ 0.0191 & 0.0745 & 0.5519 & 0.0921 \\ 0.1459 & 0.1459 & 0.1710 & 0.7019 \end{pmatrix},$$

$$P(:,:,2,4) = \begin{pmatrix} 0.5384 & 0.0484 & 0.0119 & 0.0119 \\ 0.2785 & 0.7131 & 0.2470 & 0.1759 \\ 0.0373 & 0.0926 & 0.5700 & 0.1102 \\ 0.1459 & 0.1459 & 0.1710 & 0.7019 \end{pmatrix},$$

$$P(:,:,3,4) = \begin{pmatrix} 0.5265 & 0.0365 & 0 & 0 \\ 0.1258 & 0.5604 & 0.0944 & 0.0233 \\ 0.1936 & 0.2490 & 0.7264 & 0.2666 \\ 0.1541 & 0.1541 & 0.1793 & 0.7101 \end{pmatrix},$$

$$P(:,:,4,4) = \begin{pmatrix} 0.5265 & 0.0365 & 0 & 0 \\ 0.1025 & 0.5372 & 0.0711 & 0 \\ 0.0430 & 0.0984 & 0.5758 & 0.1160 \\ 0.3280 & 0.3280 & 0.3531 & 0.8840 \end{pmatrix}.$$

In these three examples, the higher-order Markov chain is actually a fitted low-parameter version of the original data.

For the two transition probability tensors (v)–(vi), they are irreducible. The condition in Theorem 2.3 is satisfied. Each has unique and positive limiting probability distribution vector. In Table 4.2, we list γ for these two tensors. We can check that for (v) and (vi), the quantity $\gamma > 1$, but $\delta_3 \leq 1/2$. According to Theorem 3.1, we know that the iterative method converges linearly to the unique and positive probability vector. Their corresponding theoretical convergence rates $(2 - \gamma)$ are also shown in Table 4.2. For the transition probability tensor (vii), it is irreducible, but the condition in Theorem 2.4 is not satisfied (i.e., $\delta_4 \leq 2/3$).

We employ the iterative method to solve for limiting probability vectors of these transition probability tensors (v)–(vii), and find that the iterative method converges. The resulting limiting probability distribution vectors are reported in Table 4.2. To test the correctness of solution for the transition probability tensor in (vii), we use Mathematica to solve the corresponding system of polynomials of four variables, and find that the computed probability vector is the only solution.

In Table 4.2, we give the maximum value of $\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1 / \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1$ over the iterations of the proposed iterative method for the transition probability tensors in (v)–(vii). We find that their maximum values are much larger than those of (i)–(iv), and the numbers of iterations required for convergence are also more than those required for (i)–(iv). For transition probability tensors in (v) and (vi), the estimated convergence rates $(2 - \gamma)$ are very close to the maximum value of $\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1 / \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1$. In Figure 4.1(b), we show the convergence history of the iterative method for the transition probability tensor (vi). It is clear from the figure that the method converges linearly. Similar observations can be found for the transition probability tensors (v) and (vii). By comparing Figures 4.1(a) and 4.1(b), we see that the convergence of the iterative method for the tensor (ii) is much faster than that for the tensor (vi).

Example	γ	$(2 - \gamma)$	δ_m	$\max \dfrac{\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1}{\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1}$	number of iterations
(v)	1.4150	0.5850	0.4150	0.5793	37
(vi)	1.1709	0.8291	0.1710	0.8248	105
(vii)	–	–	0	0.9290	208

Example	limiting probability distribution vector
(v)	$[0.0511, 0.1610, 0.7879]$
(vi)	$[0.4728, 0.2986, 0.2286]$
(vii)	$[0.1481, 0.4161, 0.3241, 0.1117]$

TABLE 4.2
The computed quantities of the examples of tensors (v)-(vii).

5. Concluding Remarks. In this paper, we have developed a model and proposed an iterative method to calculate limiting probability distribution vector of a transition probability tensor $\mathcal{P}$ arising from a higher-order Markov chain. Experimental results are reported to demonstrate the theoretical results such as existence

and uniqueness of limiting probability distribution vector of a higher-order Markov chain, and linear convergence of the proposed iterative method. Here we give several future research work based on the proposed framework. Theoretically, we require $\mathcal{P}$ to satisfy the condition in Theorem 2.3 or Theorem 2.4, in order to show that non-negative solution of $\mathcal{P}\mathbf{x}^{m-1} = \mathbf{x}$ with $\|\mathbf{x}\|_1 = 1$ is unique. We conjecture that the uniqueness result may be true for transition probability tensors with other properties, see the tensor (vii) in Section 4. We do not have a proof yet, so we leave it as an open problem for future analysis.

REFERENCES

[1] S. Adke and S. Deshmukh, Limit distribution of a high order Markov chain, *J. R. Statis. Soc. B*, 50 (1998), pp. 105-108.

[2] A. Berchtold and A. Raftery, The mixture transition distribution model for high-order Markov chains and non-Gaussian time series, *Statistical Science*, 7 (2002), pp. 328-356.

[3] A. Berman and R. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Classics in Applied Mathematics, SIAM, 1994.

[4] K. C. Chang, K. Pearson and T. Zhang, Perron-Frobenius theorem for nonnegative tensors, *Commu. Math. Sci.*, 6 (2008) 507-520.

[5] W. Ching and M. Ng, *Markov Chains: Models, Algorithms and Applications*, International Series on Operations Research and Management Science, Springer, 2006.

[6] L. De Lathauwer, B. De Moor and J. Vandewalle, A multilinear singular value decomposition, *SIAM J. Matrix Anal. Appl.*, 21 (2000) 1253-1278.

[7] S. Friedland, S. Gaubert and L. Han, Perron-Frobenius theorem for nonnegative multilinear forms and extensions, *Linear Algebra Appl.*, to appear

[8] P. Kutchukian, D. Lou and E. Shakhnovich, FOG: Fragment optimized growth algorithm for the de novo generation of molecules occupying druglike chemical, *Journal of Chemical Information and Modeling*, 49 (2009), pp. 1630-1642.

[9] A. N. Langville, C. D. Meyer, Deeper inside PageRank, *Internet Mathematics*, 1 (2005), pp. 335-380.

[10] X. Li, M. Ng and Y. Ye, Finding stationary probability vector of a transition probability tensor arising from a higher-order Markov chain, preprint, 8 February 2011, http://www.math.hkbu.edu.hk/∼mng/tensor-research/report1.pdf.

[11] X. Li, M. Ng and Y. Ye, HAR: hub, authority and relevance scores in multi-relational data for query search, *2012 SIAM International Conference on Data Mining*, CA, 141-152.

[12] L.-H. Lim, Singular values and eigenvalues of tensors: a variational approach, *Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP'05)*, 1(2005) 129-132.

[13] I. MacDonald and W. Zucchini, *Hidden Markov and Other Models for Discrete-valued Time Series*, Chapman & Hall, London, 1997.

[14] NCBI, http://www.ncbi.nlm.nih.gov/Sitmap/samplerecord.html, gene="REV7".

[15] M. Ng, X. Li and Y. Ye, MultiRank: co-ranking for objects and relations in multi-relational data, *The 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-2011)*, August 21-24, 2011, San Diego, CA., 1217-1225.

[16] M. Ng, L. Qi and G. Zhou, Finding the largest eigenvalue of a non-negative tensor, *SIAM J. Matrix Anal. Appl.*, 31 (2009) 1090-1099.

[17] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web, 1998.

[18] L. Qi, Eigenvalues of a real supersymmetric tensor, *Journal of Symbolic Computation*, 40 (2005) 1302-1324.

[19] A. Raftery, A model of high-order Markov chains, *Journal of the Royal Statistical Society, Series B*, 47 (1985), 528-539.

[20] A. Raftrey and S. Tavare, Estimation and modelling repeated patterns in high order Markov chains with the mixture transition distribution model, *Applied Statistics*, 43 (1994), 179-199.

[21] S. Ross, *Introduction to Probability Models*, Academic Press, 2003.

[22] M. Waterman, *Introduction to Computational Biology*, Chapman & Hall, Cambridge, 1995.

[23] Q. Yang, Z. Huang and M. Ng, A Data Cube Model for Prediction-based Web Prefetching, *Journal of Intelligent Information Systems*, 20 (2003), pp. 11-30.