Applied Mathematics and Computation 303 (2017) 226–239

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Markov chains with memory, tensor formulation, and the dynamics of power iteration

Sheng-Jhih Wu[a], Moody T. Chu[b,*]

[a] Center for Advanced Statistics and Econometrics Research, School of Mathematical Sciences, Soochow University, Suzhou, China
[b] Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 USA

ARTICLE INFO

MSC:
15A18
15A51
15A69
60J99

Keywords:
Markov chain with memory
Transition probability tensor
Stationary distribution
Power method
Rate of convergence
Second dominant eigenvalue

ABSTRACT

A Markov chain with memory is no different from the conventional Markov chain on the product state space. Such a Markovianization, however, increases the dimensionality exponentially. Instead, Markov chain with memory can naturally be represented as a tensor, whence the transitions of the state distribution and the memory distribution can be characterized by specially defined tensor products. In this context, the progression of a Markov chain can be interpreted as variants of power-like iterations moving toward the limiting probability distributions. What is not clear is the makeup of the "second dominant eigenvalue" that affects the convergence rate of the iteration, if the method converges at all. Casting the power method as a fixed-point iteration, this paper examines the local behavior of the nonlinear map and identifies the cause of convergence or divergence. As an application, it is found that there exists an open set of irreducible and aperiodic transition probability tensors where the Z-eigenvector type power iteration fails to converge.

1. Introduction

A Markov chain is a stochastic process $\{X_t\}_{t=0}^{\infty}$ over a finite state space S, where the conditional probability distribution of future states in the process depends upon the present or past states. The classical "Markov property" specifies that the probability of transition to the next state depends only on the probability of the current state. That is, among the states $s_i \in S$, the model assumes that

$$\Pr(X_{t+1} = s_{t+1} | X_t = s_t, \dots, X_2 = s_2, X_1 = s_1) = \Pr(X_{t+1} = s | X_t = s_t).$$

For simplicity, identify the states as $S = \{1, 2, \dots, n\}$ and assume that the chain is time homogeneous. Then a transition probability matrix $P = [p_{ij}]$ defined by

$$p_{ij} := \Pr(X_{t+1} = i \mid X_t = j) \tag{1}$$

is independent of t and is column stochastic. The above process, generally characterized as memoryless[1], is a well studied subject in the literature.

There are situations where the data sequence does depend on past values. As can be expected, the additional history of memory often has the advantage of offering a more precise predictive value. By bringing more memory into the random

* Corresponding author.
E-mail addresses: szwu@suda.edu.cn (S.-J. Wu), chu@math.ncsu.edu, mtchu@ncsu.edu (M.T. Chu).

[1] Strictly speaking, such a process should be called a chain with memory 1 in accordance with the definition (2).

http://dx.doi.org/10.1016/j.amc.2017.01.030
0096-3003/

process, we can build a higher order Markov model. Interesting applications include packet video traffic in larger buffers [1], finance risk management [2–4], wind turbine design [5], alignment of DNA sequences or long-range correlated dynamic systems [6–8], growth of polymer chains [9,10], cloud data mining [11,12], and many others [13]. A Markov chain with memory m is a process satisfying

$$\Pr(X_{t+1} = s_{t+1} | X_t = s_t, \ldots, X_1 = s_1) = \Pr(X_{t+1} = s_{t+1} | X_t = s_t, \ldots, X_{t-m+1} = s_{t-m+1}) \tag{2}$$

for all $t \geq m$. By defining

$$Y_t = (X_t, X_{t-1}, \ldots, X_{t-m+1}) \tag{3}$$

and by taking the ordered m-tuples of X values as its product state space, it is easy to see that the chain $\{Y_t\}$ with suitable starting values satisfies the Markov property. In principle, upon exploiting the underlying structure, the transition process can be analyzed with the classical theory for memoryless Markov chain. Note, however, that the size of the aggregated chain, also known as the Markovianization, is considerably larger — of dimension n^{m-1}. Though mathematically equivalent, basic tasks such as bookkeeping multi-states and other associated operations will be fairly tedious[2].

In recent years higher-order tensor analysis have become an effective way to address high-throughput and multi-dimensional data by different disciplines. Markov chain with memory fits naturally such a tensor formulation. Assuming again time homogeneity, a Markov chain with memory $m-1$ can be conveniently represented via the order-m tensor $\mathcal{P} = [p_{i_1 i_2 \ldots i_m}]$ defined by

$$p_{i_1 i_2 \ldots i_m} := \Pr(X_{t+1} = i_1 | X_t = i_2, \ldots, X_{t-m+2} = i_m), \tag{4}$$

where $\mathcal{P}$ is called a transition probability tensor. Necessarily we have the properties that $0 \leq p_{i_1 i_2 \ldots i_m} \leq 1$ and that

$$\sum_{i_1=1}^{n} p_{i_1 i_2 \ldots i_m} = 1 \tag{5}$$

for every fixed $(m-1)$-tuple $(i_2, \ldots, i_m)$. What is most interesting is that the transitions among the states as well as the history of memory can be characterized by specially defined tensor products. Our goal in this paper is to recast such a process under the tensor formulation. In particular, we are interested in understanding the dynamics of the transition to the stationary distribution and the associated 2-phase power iteration scheme in the context of tensor operations.

While some classical results in matrix theory can be extended naturally to tensors, there are cases where the nonlinearity of tensors makes the generalization far more cumbersome. The notion of eigenvalue is one such incident. Depending on the applications, there are several ways to mull over how an eigenvalue of a tensor should be defined [14–17]. Markov chain with memory and the associated transition probability tensor can serve as a practical model for exploring the following two notions of eigenvalues and their implications:

1. The classical concept of eigenvalues when characterizing the evolution of the joint probability mass functions.
2. The notion of Z-eigenvalue[3] when approximating the evolution of the state probability distribution.

In this context, we study the role of the "second dominant eigenvalue" in such a dynamics of a Markov chain with memory. We also intend to address some practical issues arisen from a recent discussion in [10] which proposes to short cut the computation of the stationary state distribution by approximating the stationary joint probability mass function. These issues include whether the assumption used in proposing the Z-eigenvector computation is statistically justifiable and the anatomy of the true cause that affects the rate of convergence. The tool we are about to develop gives some insight into this limiting behavior. It is possible to generalize our framework to other types of eigenvalues for tensors, e.g., the so called H-eigenvalues [18]. For demonstration, we choose to concentrate only on the application to the transition probability tensors in this presentation.

This paper is organized as follows. We begin in Section 2 with some basic properties of transition probability tensors. We review two types of dynamics necessarily involved in a Markov chain with memory, each of which entails a particular kind of tensor product. The evolution of the joint probability mass function itself follows a scheme similar to the conventional power method, whereas finding the stationary probability distributions of the state vector requires a 2-phase iteration. In Sections 3, we argue that an appropriate rearrangement of the transition probability tensor reveals the proper cause of convergence for this classical type of evolution. In Section 4 we address some concerns arisen from the recent notion of approximating the stationary distribution by the dominant Z-eigenvector. We identify the true makeup of the "second" dominant eigenvalue in the tensor setting. Most importantly, we show by counter examples that the convergence of this shortcut type of power method proposed in [10] is not always guaranteed. Included in the Appendix is the local analysis in a similar spirit for matrices, which probably offers an alternative explanation of convergence for the classical power method.

[2] See an example of vectorizing a Markov chain with memory 2 in Section 3.

[3] Given an order-m tensor $A = [a_{i_1 \ldots i_m}] \in \mathbb{R}^{n \times \ldots \times n}$ and an n-dimensional vector $\mathbf{x}$, real or complex, let the tensor product $A\mathbf{x}^{m-1}$ denote an n-dimensional vector whose ith entry is defined by $\sum_{i_2,\ldots,i_m=1}^{n} a_{i i_2 \ldots i_m} x_{i_2} \ldots x_{i_m}$, $1 \leq i \leq n$. If there exist a non-zero vector $\mathbf{x}$ and scalar λ such that $A\mathbf{x}^{m-1} = \lambda\mathbf{x}$, then λ is called a Z-eigenvalue and $\mathbf{x}$ the corresponding Z-eigenvector of A. See [15,17] for the initial exploration of this subject.