

Ergodicity Coefficients for Higher-Order Stochastic Processes*

Dario Fasino[†] and Francesco Tudisco[‡]

Abstract. The use of higher-order stochastic processes such as nonlinear Markov chains or vertex-reinforced random walks is significantly growing in recent years as they are much better at modeling high dimensional data and nonlinear dynamics in numerous application settings. In many cases of practical interest, these processes are identified with a stochastic tensor, and their stationary distribution is a tensor Z -eigenvector. However, fundamental questions such as the convergence of the process towards a limiting distribution and the uniqueness of such a limit are still not well understood and are the subject of rich recent literature. Ergodicity coefficients for stochastic matrices provide a valuable and widely used tool to analyze the long-term behavior of standard, first-order, Markov processes. In this work, we extend an important class of ergodicity coefficients to the setting of stochastic tensors. We show that the proposed higher-order ergodicity coefficients provide new explicit formulas that (a) guarantee the uniqueness of Perron Z -eigenvectors of stochastic tensors, (b) provide bounds on the sensitivity of such eigenvectors with respect to changes in the tensor, and (c) ensure the convergence of different types of higher-order stochastic processes governed by cubical stochastic tensors. Moreover, we illustrate the advantages of the proposed ergodicity coefficients on several example application settings, including the analysis of PageRank vectors for triangle-based random walks and the convergence of lazy higher-order random walks.

Key words. nonnegative tensors, stochastic tensors, higher-order Markov chain, ergodicity coefficient, Z -eigenvector, vertex-reinforced random walk, spacey random walk, multilinear PageRank

AMS subject classifications. 15B51, 65F35, 60J10, 65C40

DOI. 10.1137/19M1285214

1. Introduction. Markov processes are among the best known and most popular stochastic processes in computational mathematics and mathematics of data science. For these types of processes the state transitions only depend on the last state. This is modeled by a stochastic matrix P whose entries P_{ij} quantify the probability of the process of transitioning from state j to state i . Any such a matrix leaves the simplex $\mathcal{S}_1 = \{x \geq 0 : x^T \mathbb{1} = 1\}$ invariant, and the classical Brouwer's fixed point theorem thus implies that there exists at least one stationary distribution $x = Px$ for the Markov chain described by P . In other words, there exists at least one eigenvector x of P corresponding to the eigenvalue 1 such that x has nonnegative entries that sum up to one. While Brouwer's theorem holds in general for

*Received by the editors September 3, 2019; accepted for publication (in revised form) July 20, 2020; published electronically August 31, 2020.

<https://doi.org/10.1137/19M1285214>

Funding: The work of the first author was supported by INdAM-GNCS, Italy, and by the departmental research project ICON (Innovative Combinatorial Optimization in Networks), DMIF-PRID 2017, University of Udine, Italy. The work of the second author was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie individual fellowship "MAGNET" grant agreement 744014.

[†]Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy (dario.fasino@uniud.it).

[‡]GSSI – Gran Sasso Science Institute, 67100, L'Aquila (AQ), Italy (francesco.tudisco@gssi.it).

mappings leaving a closed convex set invariant, much more can be said for the specific case of stochastic matrices. In particular, if the Markov chain described by P is ergodic, then P has a unique nonnegative eigenvector x in \mathcal{S}_1 which corresponds to the eigenvalue 1, the magnitude of any other eigenvalue of P is strictly smaller than one, and the power method $x_{t+1} = Px_t$ converges to x for any choice of $x_0 \in \mathcal{S}_1$, with a convergence rate that depends on the largest subdominant eigenvalue. In a way, these properties characterize the concept of ergodic chain, and the so-called ergodicity coefficients were introduced to estimate whether or not a Markov chain is ergodic without resorting to spectral properties [31, 56].

A natural extension of a Markov process is to have the state transitions depend on several previous states, rather than just the last one. While the study and application of this kind of higher-order stochastic processes has a relatively long history (see, e.g., [7, 50]), their interest has grown significantly in more recent years due to their ability to improve the mathematical modeling and understanding of numerous problems in data and network sciences, such as detecting communities and analyzing spreading dynamics in networks [54, 61], understanding the behavior of web browsers and drivers trajectories [8, 17], defining new clustering algorithms in data mining that exploit motifs and nonbacktracking walks [10, 35], improving centrality and ranking models for networks and hypergraphs [3, 27, 42, 45], and forecasting the appearance of new links or finding missing links in networks [4, 44].

Many higher-order stochastic processes of practical interest can be modeled by hypermatrices, or tensors, with m modes $\mathbf{P} = (\mathbf{P}_{i_1, \dots, i_m})$. When $m = 3$, for example, we have a second-order Markov chain if \mathbf{P}_{ijk} quantifies the probability of transitioning to state i , given that the last state was j and the previous one k . Another example is given by linear vertex-reinforced random walks, where the transition probability from state j to state i is defined as $\sum_k \mathbf{P}_{ijk} y_k$ for a vector y which depends on the history of the states that have been visited. The stationary distribution equation for these higher-order stochastic processes boils down to the Z -eigenvector of the corresponding stochastic tensor. While the existence of such stationary distribution is ensured also in this setting by Brouwer's fixed point theorem, the ergodicity of those processes is much less understood than that of matrix-based processes. By extending the wide and influential literature on ergodicity coefficients for matrices, in this work we introduce a family of higher-order ergodicity coefficients for stochastic cubical tensors and discuss how these allow us to derive new conditions on the existence, uniqueness, and computability of stationary distributions for different type of higher-order stochastic processes described by tensors, in particular, second-order Markov chains and a new class of linear vertex-reinforced random walks for which, to the best of our knowledge, we provide the first convergence result for both the occupation vector and the density distribution. This class includes previously considered vertex-reinforced stochastic processes such as the spacey random walk [8].

From the linear algebraic perspective, our new conditions allow us to prove guarantees for existence, uniqueness, and computability of the Perron Z -eigenvector of a stochastic tensor of order three. Dominant Z -eigenvectors of nonnegative tensors appear in a large variety of applications, including diffusion kurtosis imaging in medical engineering [49], low-rank factorization and signal processing [1, 20], quantum processing, quantum geometry, and data mining [5, 9, 30]. Even though computing a prescribed Z -eigenvector is in general NP-hard [28], the use of higher-order ergodicity coefficients allows us to identify a class of nonnegative tensors for which the Perron eigenvector can be approximated efficiently to an arbitrary precision.

While we focus here on stochastic tensors of order three, we believe the results here presented can be further extended to more general eigenvector problems for nonnegative tensors.

The remainder of the paper is structured as follows: We fix the relevant notation in the next section. In section 3 we review the concept of higher-order Markov chain, its associated Z -eigenvector stationary distribution, and the issues related to the ergodicity of this type of higher-order stochastic process. In section 4 we recall the concept of ergodicity coefficient for a stochastic matrix and some of its properties. Then, in section 5, we introduce our new higher-order ergodicity coefficients for stochastic cubic tensors, and we prove some of our main results. In section 6 we show how these apply to the ergodicity of higher-order stochastic processes. In particular, after recalling the definition of vertex-reinforced and spacey random walks, we introduce in subsection 6.2 a general family of Markov processes with memory that includes the spacey random walk as a particular case, and we prove a new convergence result for this general stochastic process. In section 7 we compare the ergodicity coefficients introduced in section 5 with analogous coefficients found in the recent literature. Finally, in section 8, we discuss a number of application examples that showcase the advantages of the proposed results. In particular, we consider the computation of the multilinear PageRank and its application to triangle-based random walks in networks, and the convergence of the shifted higher-order power method.

2. Notation. Let e_i be the i th canonical basis vector in \mathbb{R}^n , and let $\mathbf{1}$ be the all-ones vector. Define the sets $\mathcal{S}_1 = \{x \in \mathbb{R}^n : x \geq 0, \|x\|_1 = 1\}$ and $\mathcal{Z}_1 = \{x \in \mathbb{R}^n : \mathbf{1}^T x = 0, \|x\|_1 = 1\}$. A real cubical tensor \mathbf{P} of order 3 (or, equivalently, with 3 modes) is a three-way array with real entries of size $n \times n \times n$. We denote by $\mathbb{R}^{[3,n]}$ the set of such tensors and use capital bold letters to denote its elements. The (i, j, k) -entry of $\mathbf{P} \in \mathbb{R}^{[3,n]}$ is denoted by \mathbf{P}_{ijk} . Matrices are tensors with only 2 modes and are denoted with standard capital letters.

Given a tensor $\mathbf{P} \in \mathbb{R}^{[3,n]}$, we write $\mathbf{P}xy$ to denote the tensor-vector multiplication over the second and third modes:

$$(\mathbf{P}xy)_i = \sum_{j,k=1}^n \mathbf{P}_{ijk} x_j y_k$$

for $i = 1, \dots, n$. Moreover, the product \mathbf{Px} denotes the matrix associated with the linear map $y \mapsto \mathbf{P}xy$, that is,

$$(2.1) \quad (\mathbf{Px})_{ij} = \sum_{k=1}^n \mathbf{P}_{ikj} x_k.$$

With this notation, it holds that $(\mathbf{Px})y = \mathbf{P}xy$. A Z -eigenvalue of a tensor $\mathbf{P} \in \mathbb{R}^{[3,n]}$ is a real number λ such that there exists a nonzero vector $x \in \mathbb{R}^n$ such that $\mathbf{P}xx = \lambda x$. Such vector x is a Z -eigenvector associated with λ ; see [48].

There are $6 = 3!$ possible transpositions of a tensor $\mathbf{P} \in \mathbb{R}^{[3,n]}$, each corresponding to a different permutation π of the set $\{1, 2, 3\}$. Using the notation proposed in [52], the transposed tensor corresponding to the permutation π can be denoted by $\mathbf{P}^{\langle \pi \rangle}$, namely, $(\mathbf{P}^{\langle \pi \rangle})_{ijk} = \mathbf{P}_{\pi(i), \pi(j), \pi(k)}$. As it will be of particular importance to us, we devote the special notation

\mathbf{P}^S to denote the tensor obtained by transposing the entries of \mathbf{P} over the second and third modes, namely,

$$\mathbf{P}^S = \mathbf{P}^{([132])}, \quad (\mathbf{P}^S)_{ijk} = \mathbf{P}_{ikj}.$$

Moreover, we say that a tensor \mathbf{P} is S -symmetric whenever $\mathbf{P} = \mathbf{P}^S$.

All inequalities in this work are meant entrywise. In particular, we write $\mathbf{P} \geq 0$ (resp., $\mathbf{P} > 0$) to denote a tensor such that $\mathbf{P}_{ijk} \geq 0$ (resp., $\mathbf{P}_{ijk} > 0$) for all indices $i, j, k = 1, \dots, n$. A tensor $\mathbf{P} \in \mathbb{R}^{[3,n]}$ is said to be column stochastic, or simply stochastic, if $\mathbf{P} \geq 0$ and its first mode entries all sum up to one, i.e., $\sum_{i=1}^n \mathbf{P}_{ijk} = 1$ for all $j, k = 1, \dots, n$. A tensor acting as the identity on the unit sphere $x^T x = 1$ can be defined in the case of tensors with an even number of modes; see [33]. For tensors with three modes we define the following two left \mathbf{E}^L and right \mathbf{E}^R “one-sided identity tensors”:

$$(2.2) \quad \mathbf{E}_{ijk}^L = \delta_{ij} \quad \text{and} \quad \mathbf{E}_{ijk}^R = \delta_{ik} \quad \text{for all } i, j, k = 1, \dots, n.$$

Both \mathbf{E}^L and \mathbf{E}^R are stochastic tensors, and for all $x \in \mathcal{S}_1$ and $v \in \mathbb{R}^n$ one has $\mathbf{E}^L v x = v$ and $\mathbf{E}^R x v = v$. Indeed, $(\mathbf{E}^L v x)_i = \sum_{jk} \mathbf{E}_{ijk}^L v_j x_k = \sum_{jk} \delta_{ij} v_j x_k = v_i \sum_k x_k = v_i$ and similarly for \mathbf{E}^R . Note that, letting $\mathbf{E} = \alpha \mathbf{E}^L + (1 - \alpha) \mathbf{E}^R$ for any $\alpha \in [0, 1]$, it holds that $\mathbf{E} x x = x$ for all $x \in \mathcal{S}_1$.

3. Higher-order Markov chains. Higher-order Markov chains are a natural extension of Markov chains, where the transition probabilities depend on the past few states, rather than just the last one. For a plain introduction, see, e.g., [8, 62]. For example, a discrete-time second-order Markov chain is defined by a third-order tensor $\mathbf{P} = (\mathbf{P}_{ijk})$ where \mathbf{P}_{ijk} is the conditional probability of transitioning to state i , given that the last state was j and the second last state was k . More precisely, if $X(t)$ is the random variable describing the status of the chain on the set $\{1, \dots, n\}$ at time $t = 0, 1, \dots$, then

$$\mathbf{P}_{ijk} = \mathbb{P}(X(t+1) = i | X(t) = j, X(t-1) = k),$$

where \mathbb{P} denotes probability. Hence, the sequence $\{X(t)\}$ obeys the rule

$$(3.1) \quad \mathbb{P}(X(t+1) = i) = \sum_{j,k} \mathbf{P}_{ijk} \mathbb{P}(X(t) = j, X(t-1) = k).$$

Obviously it must hold $\sum_i \mathbf{P}_{ijk} = 1$ for $j, k = 1, \dots, n$, i.e., the tensor \mathbf{P} is stochastic.

Let $x_t \in \mathcal{S}_1$ be the probability vector of the random variable $X(t)$, i.e., the vector with entries $(x_t)_i = \mathbb{P}(X(t) = i)$. Let Y_t denote the joint probability function $(Y_t)_{ij} = \mathbb{P}(X(t) = i, X(t-1) = j)$. Then, x_t is the marginal probability $Y_t \mathbf{1}$, i.e., the vector with entries $(x_t)_i = \sum_j (Y_t)_{ij}$.

Hence, the dynamic of the second-order Markov chain (3.1) is described by the two-phase process

$$(3.2) \quad \begin{cases} (Y_{t+1})_{ij} = \sum_k \mathbf{P}_{ijk} (Y_t)_{jk}, \\ (x_{t+1})_i = \sum_j (Y_{t+1})_{ij}. \end{cases}$$

Note that both steps in (3.2) are linear, and thus their convergence can be analyzed using standard ergodicity arguments. In fact, the second-order Markov chain over the state set $\{1, \dots, n\}$ can be easily reduced to a first-order Markov chain with state space $\{1, \dots, n\} \times \{1, \dots, n\}$; see, e.g., [8, 62]. Thus, under appropriate hypotheses on \mathbf{P} , the iteration (3.2) has a unique limit $Y \geq 0$ such that

$$(3.3) \quad Y_{ij} = \sum_{k=1}^n \mathbf{P}_{ijk} Y_{jk}.$$

However, this approach has a computational drawback: the size of the joint probability function of a second-order Markov chain is the square of the number of states. The situation gets even worse for an m th order Markov chain due to the “curse of dimensionality” effect: the memory space required by the joint density grows exponentially with the states space size, requiring n^m entries. Moreover, the convergence analysis of the iteration (3.2) and its natural extension to the $m > 2$ setting becomes cumbersome.

In order to circumvent these issues, Raftery [50] proposed a technique to approximate higher-order Markov chains by means of a linear combination of first-order ones, by assuming that the joint probability distribution of the lagged random variables $X(t), \dots, X(t-m+1)$ can be replaced by a mixture of its marginals. In the second-order ($m = 2$) case, that assumption reduces to replacing the conditional probabilities \mathbf{P}_{ijk} by an expression of the form $\lambda Q_{ij} + (1 - \lambda) Q_{ik}$, where Q is a stochastic matrix and $\lambda \in [0, 1]$. This technique, known as the mixture transition distribution model, has been widely used to fit stochastic models with far fewer parameters than the fully parameterized model to multidimensional data in a variety of applications [11, 51].

A more recent approach, which maintains all the information contained in the transition tensor \mathbf{P} , is the one proposed by Li and Ng in [39]. Here, still in the $m = 2$ case, one assumes that the joint probability distribution of the higher-order Markov chain is the tensor product of its marginal distributions, that is, $Y_t = x_t x_{t-1}^T$. This hypothesis, which is equivalent to assuming that the random variables $X(t)$ and $X(t-1)$ are independent, is a conceptual simplification of the Markov chain formalism that is introduced in order to obtain a computationally tractable extension to the second-order case. The resulting process is the quadratic version of a nonlinear Markov process [34], and it is still called a second-order Markov chain by many authors; see, e.g., [29, 36, 39]. In this work, we will follow this well-established convention.

Using our tensor-vector product notation, this “reduced” higher-order Markov process boils down to the iteration

$$(3.4) \quad x_{t+1} = \mathbf{P} x_t x_{t-1},$$

which replaces (3.2) and is the higher-order counterpart of the usual Markov process for a stochastic matrix in the classical (first-order) Markov chain setting. The limit of this sequence, if it exists, is a nonnegative vector $x \in \mathcal{S}_1$ such that

$$(3.5) \quad x = \mathbf{P} x x,$$

that is, x is a Z -eigenvector of \mathbf{P} associated with the Z -eigenvalue 1. Thus, it is natural to consider any such vector as a stationary density of the Markov chain (3.1).

Note that the limit matrix Y of (3.2) is such that $Y\mathbf{1} = Y^T\mathbf{1}$. Indeed, from (3.3) we have

$$(\mathbf{1}^T Y)_j = \sum_i Y_{ij} = \sum_k Y_{jk} \sum_i P_{ijk} = \sum_k Y_{jk} = (Y\mathbf{1})_j.$$

But that row-column sum is generally different from the vector x in (3.5). In fact, that solution corresponds to a case where Y has rank one, namely, $Y = xx^T$. Indeed, if Y in (3.3) is such that $\text{rank}(Y) = 1$, then $x = Y\mathbf{1}$ must solve (3.5).

On the other hand, the converse implication is false in general; that is, if x solves (3.5), then the matrix $Y = xx^T$ may not be a solution of (3.3). Indeed, extensive numerical experiments reported in [62] show that the vector x is strongly correlated with the row-column sum vector of Y , but the matrix Y has full rank in general and $x \neq Y\mathbf{1}$.

3.1. Ergodicity of higher-order Markov chains. In the matrix case, a Markov chain is called ergodic whenever it has a unique stationary vector and, for any initial probability distribution, that vector is the limiting distribution of the chain. Necessary and sufficient conditions for ergodicity of Markov chains are well known and are essentially related to spectral and structural properties, e.g., irreducibility and aperiodicity, of the transition matrix [56].

The situation complicates significantly when moving from matrices to tensors and, more generally, from linear to nonlinear cases [34]. In fact, even though the existence of a solution $x \in \mathcal{S}_1$ to (3.5) is a direct consequence of Brouwer's fixed point theorem, the properties that characterize uniqueness and convergence of the process to the stationary distribution do not extend straightforwardly from the matrix case. For example, unlike the matrix case, the irreducibility of \mathbf{P} is not enough to ensure the uniqueness of x and additional assumptions are required. In fact, it is not too difficult to produce examples of entrywise positive stochastic tensors for which (3.5) has multiple solutions, or the solution of (3.5) is unique but the iteration (3.4) fails to converge to that solution. For instance, a $\mathbf{P} \in \mathbb{R}^{[4,2]}$ example is provided by Chang and Zhang in [15], while several $\mathbf{P} \in \mathbb{R}^{[3,3]}$ examples are provided by Saburov in [55].

A sufficient condition that ensures ergodicity is the existence of a metric with respect to which the system is contractive. Even though, as in the linear case, this is a sufficient but not a necessary requirement in general, suitable choices of the metric can provide valuable conditions for the ergodicity of higher-order stochastic processes that can be given in terms of the entries of the tensor \mathbf{P} .

By considering the ℓ^1 and the Hilbert metrics on \mathcal{S}_1 , in the following we introduce a family of ergodicity coefficients for stochastic cubic tensors of order three, and we show, in section 6, how they allow us to prove new conditions for the ergodicity of various higher-order stochastic processes. The conditions we obtain in this way can be easily computed and are, to the best of our knowledge, among the weakest conditions available in the literature so far.

4. Coefficients of ergodicity. Let $d : \mathcal{S}_1 \times \mathcal{S}_1 \rightarrow \mathbb{R}_+$ be a metric on $\mathcal{S}_1 = \{x \geq 0 : x^T \mathbf{1} = 1\}$, and consider a mapping $f : \mathcal{S}_1 \rightarrow \mathcal{S}_1$. Although other notions of ergodicity coefficient are available in the literature (see, e.g., [31]), for the purpose of this work a coefficient of ergodicity for f is the best Lipschitz constant of f with respect to d , that is,

$$(4.1) \quad \tau_d(f) = \sup_{\substack{x,y \in \mathcal{S}_1 \\ x \neq y}} \frac{d(f(x), f(y))}{d(x, y)}.$$

Different choices of the metric d give rise to different notions of ergodicity coefficients. For example, if d is the Hilbert projective distance

$$(4.2) \quad d_H(x, y) = \log \left(\max_i \frac{x_i}{y_i} \max_i \frac{y_i}{x_i} \right),$$

then (4.1) is the so-called Birkhoff contraction ratio [12], which we denote by $\tau_H(f)$. This choice of metric is particularly interesting because it extends very naturally to the case of mappings f that leave the slice of a generic proper cone invariant (not just \mathcal{S}_1). Moreover, when f is a linear map described by the matrix A , the Birkhoff–Hopf theorem [22] provides an explicit formula for $\tau_H(f) = \tau_H(A)$, which we recall below:

$$\tau_H(A) = \tanh \left(\frac{1}{4} \log \left(\max_{ijk} \frac{A_{ij} A_{hk}}{A_{ik} A_{hj}} \right) \right),$$

where $\tanh(\lambda) = (e^\lambda - e^{-\lambda})/(e^\lambda + e^{-\lambda})$ denotes the hyperbolic tangent. An equivalent formula can be found also in [56, section 3.4]. More recently, in [24], an analogous explicit formula has been proved for the case where f is a (weakly) multilinear mapping induced by a nonnegative tensor. In particular, this formula holds for the case of Z -eigenvectors of cubic stochastic tensors, and we will review it in this setting in subsection 7.1.

Another popular and successful choice for the distance in (4.1) is $d(x, y) = \|x - y\|_p$, where $\|\cdot\|_p$ is the p -norm on \mathbb{R}^n . Norm-based coefficients were introduced by Dobrushin in 1956 [21] for the case of linear mappings and have been the subject of numerous investigations afterwards; see, e.g., [31, 55, 56].

In section 5 we analyze properties of norm-based coefficients for mappings defined by a stochastic tensor \mathbf{P} . To this end, we first review some relevant properties of these coefficients for the case of linear maps.

4.1. Norm-based ergodicity coefficients for matrices. Let P be a stochastic matrix and $p \geq 1$. The p -norm ergodic coefficient of P is

$$\tau_p(P) = \sup_{\substack{x, y \in \mathcal{S}_1 \\ x \neq y}} \frac{\|Px - Py\|_p}{\|x - y\|_p}.$$

This definition extends obviously to any matrix $P \in \mathbb{R}^{n \times n}$, when appropriate. The linearity of P , the continuity of $\|\cdot\|_p$, and the fact that the set $\{z \in \mathbb{R}^n : z = (x - y)/\|x - y\|_p, x, y \in \mathcal{S}_1\}$ coincides with $\mathcal{Z}_p = \{z \in \mathbb{R}^n : \|z\|_p = 1, \mathbf{1}^T z = 0\}$, which is compact, yield the equivalent formula

$$\tau_p(P) = \max_{\substack{\|x\|_p=1 \\ x^T \mathbf{1}=0}} \|Px\|_p.$$

Norm-based ergodicity coefficients for stochastic matrices P are particularly useful for three reasons: they provide sufficient conditions for the ergodicity of the Markov chain associated with P ; they can be used to derive bounds on the variation of the stationary distribution of the Markov chain, when the transition probabilities face a small perturbation; and, in the case $p = 1$, they yield easily computable upper bounds on the convergence rate of the Markov

process $x_{t+1} = Px_t$. We review these properties in the next Theorems 4.2, 4.3, and 4.4. Then, in sections 5 and 6, we will use the 1-norm ergodicity coefficients for matrices to introduce what we call higher-order ergodicity coefficients for stochastic tensors of order three, and we will show that the above three fundamental properties carry over to the tensor case. We refer to [31, 56, 60] for more details on $\tau_p(P)$.

The following properties follow directly from the definition of p -norm ergodic coefficient.

Theorem 4.1. *If $P, Q \in \mathbb{R}^{n \times n}$ are stochastic, then*

1. $0 \leq \tau_p(P) \leq \|P\|_p$,
2. $|\tau_p(P) - \tau_p(Q)| \leq \tau_p(P - Q)$,
3. $\tau_p(P) = 0$ if and only if $\text{rank}(P) = 1$.

Moreover, the following perturbation bound holds (see, e.g., [57] or [31, Theorem 3.14]).

Theorem 4.2. *Let P, P' be two stochastic irreducible matrices, and let x, x' be their corresponding stationary probability vectors. Then*

$$\|x - x'\|_p \leq \frac{\|P - P'\|_p}{1 - \tau_p(P)}.$$

As an immediate consequence of the definition (4.1), the inequality $\tau_p(P) < 1$ implies that the map $f : \mathcal{S}_1 \mapsto \mathcal{S}_1$ defined by $f(x) = Px$ is a contraction. This observation implies the following result.

Theorem 4.3. *If P is a stochastic matrix with $\tau_p(P) < 1$ for some $p \geq 1$, then P is ergodic, i.e., there exists a unique eigenvector $x \in \mathcal{S}_1$ such that $Px = x$. Moreover, the power method $x_{t+1} = Px_t$ converges to x for any $x_0 \in \mathcal{S}_1$, and*

$$\|x_t - x\|_p \leq \tau_p(P)^t \|x_0 - x\|_p.$$

The theorem above gives a sufficient condition for the ergodicity of P which is very useful in practice when combined with a number of explicit formulas that allow us to compute $\tau_p(P)$ using only the entries of P . Here we recall those for the particular case $p = 1$ [21].

Theorem 4.4. *Let $P \in \mathbb{R}^{n \times n}$. Then*

$$\tau_1(P) = \frac{1}{2} \max_{j,k} \sum_i |P_{ij} - P_{ik}|.$$

Moreover, if P is stochastic, then

$$\tau_1(P) = 1 - \min_{jk} \sum_i \min\{P_{ij}, P_{ik}\} = 1 - \min_{\mathcal{I} \subset \{1, \dots, n\}} \left(\min_j \sum_{i \notin \mathcal{I}} P_{ij} + \min_k \sum_{i \in \mathcal{I}} P_{ik} \right).$$

We will devote sections 5 and 6 to extend the ergodicity coefficient $\tau_1(P)$ to three-mode tensors, to prove analogous theorems to the preceding ones, and to discuss further properties and applications.

5. Ergodicity coefficients for third-order tensors. Let $\mathbf{P} \in \mathbb{R}^{[3,n]}$ be a cubic stochastic tensor. We define the following higher-order ergodicity coefficients:

$$(5.1) \quad \begin{aligned} \mathcal{T}_L(\mathbf{P}) &= \max_{x \in \mathcal{S}_1} \max_{y \in \mathcal{Z}_1} \|\mathbf{P}xy\|_1, \\ \mathcal{T}_R(\mathbf{P}) &= \max_{x \in \mathcal{S}_1} \max_{y \in \mathcal{Z}_1} \|\mathbf{Pyx}\|_1, \\ \mathcal{T}(\mathbf{P}) &= \max_{x \in \mathcal{S}_1} \max_{y \in \mathcal{Z}_1} \|\mathbf{P}xy + \mathbf{Pyx}\|_1. \end{aligned}$$

The preceding definitions are extended obviously to any tensor $\mathbf{P} \in \mathbb{R}^{[3,n]}$, when appropriate. We remark the following immediate identities:

$$\mathcal{T}_L(\mathbf{P}) = \mathcal{T}_R(\mathbf{P}^S), \quad \mathcal{T}(\mathbf{P}) = 2\mathcal{T}_L\left(\frac{1}{2}\mathbf{P} + \frac{1}{2}\mathbf{P}^S\right) = 2\mathcal{T}_R\left(\frac{1}{2}\mathbf{P} + \frac{1}{2}\mathbf{P}^S\right).$$

In particular, for an S -symmetric tensor \mathbf{P} we have $\mathcal{T}_L(\mathbf{P}) = \mathcal{T}_R(\mathbf{P}) = \frac{1}{2}\mathcal{T}(\mathbf{P})$.

The relationship between the preceding definitions and the norm-based ergodicity coefficients considered in section 4.1 can be revealed by considering the matrices associated with the tensor-vector products \mathbf{Px} and \mathbf{P}^Sx defined as in (2.1). In fact, it is not difficult to see that the following identities hold:

$$\mathcal{T}_L(\mathbf{P}) = \max_{x \in \mathcal{S}_1} \tau_1(\mathbf{Px}), \quad \mathcal{T}_R(\mathbf{P}) = \max_{x \in \mathcal{S}_1} \tau_1(\mathbf{P}^Sx), \quad \mathcal{T}(\mathbf{P}) = \max_{x \in \mathcal{S}_1} \tau_1(\mathbf{Px} + \mathbf{P}^Sx).$$

The above formulas yield a characterization of the three coefficients in (5.1) which, for example, was used in [55] to define $\mathcal{T}(\mathbf{P})$ in the case of S -symmetric tensors. Hereafter, we exploit these formulas to derive explicit expressions for computing the coefficients above from the knowledge of the tensor entries and provide the tensor equivalent of Theorem 4.4.

Theorem 5.1. *Let $\mathbf{P} \in \mathbb{R}^{[3,n]}$. Then,*

$$(5.2) \quad \mathcal{T}_L(\mathbf{P}) = \frac{1}{2} \max_{j,k_1,k_2} \sum_i |\mathbf{P}_{ijk_1} - \mathbf{P}_{ijk_2}|.$$

Moreover, if \mathbf{P} is stochastic, then

$$(5.3) \quad \mathcal{T}_L(\mathbf{P}) = 1 - \min_{j,k_1,k_2} \sum_i \min\{\mathbf{P}_{ijk_1}, \mathbf{P}_{ijk_2}\}$$

$$(5.4) \quad = 1 - \min_{\mathcal{I} \subset [n]} \min_j \left(\min_{k_1} \sum_{i \notin \mathcal{I}} \mathbf{P}_{ijk_1} + \min_{k_2} \sum_{i \in \mathcal{I}} \mathbf{P}_{ijk_2} \right).$$

Proof. For $i = 1, \dots, n$ let $P^{(i)}$ be the stochastic matrix $P_{jk}^{(i)} = \mathbf{P}_{jik}$. Hence, for any $x \in \mathcal{S}_1$ and $y \in \mathcal{Z}_1$ we have $\mathbf{P}xy = \sum_i x_i P^{(i)}y$. By the triangle inequality,

$$\|\mathbf{P}xy\|_1 = \left\| \sum_i x_i P^{(i)}y \right\|_1 \leq \sum_i x_i \|P^{(i)}y\|_1 \leq \max_i \tau_1(P^{(i)}).$$

On the other hand, if $x = e_i$ where i is an index such that $\tau_1(P^{(i)}) \geq \tau_1(P^{(j)})$ for $j = 1, \dots, n$ and $y \in \mathcal{Z}_1$ is a vector such that $\|P^{(i)}y\|_1 = \tau_1(P^{(i)})$, then the inequalities above are actually equalities. Thus all the claims follow at once from Theorem 4.4. ■

The analogous formulas for the other higher-order coefficients in (5.1) are derived hereafter.

Corollary 5.2. *Let $\mathbf{P} \in \mathbb{R}^{[3,n]}$. The following identities hold:*

$$(5.5) \quad \mathcal{T}_R(\mathbf{P}) = \frac{1}{2} \max_{j_1, j_2, k} \sum_{i=1}^n |\mathbf{P}_{ijk} - \mathbf{P}_{ij_2k}|,$$

$$(5.6) \quad \mathcal{T}(\mathbf{P}) = \frac{1}{2} \max_{j, k_1, k_2} \sum_i |\mathbf{P}_{ijk_1} - \mathbf{P}_{ijk_2} + \mathbf{P}_{ik_1j} - \mathbf{P}_{ik_2j}|.$$

Moreover, if \mathbf{P} is stochastic, then

$$(5.7) \quad \mathcal{T}_R(\mathbf{P}) = 1 - \min_{j_1, j_2, k} \sum_{i=1}^n \min\{\mathbf{P}_{ijk}, \mathbf{P}_{ij_2k}\}$$

$$(5.8) \quad = 1 - \min_{\mathcal{I} \subset [n]} \min_{j_1, j_2, k} \left(\sum_{i \notin \mathcal{I}} \mathbf{P}_{ijk} + \sum_{i \in \mathcal{I}} \mathbf{P}_{ij_2k} \right),$$

$$(5.9) \quad \mathcal{T}(\mathbf{P}) = 2 - \min_{j, k_1, k_2} \sum_i \min\{\mathbf{P}_{ijk_1} + \mathbf{P}_{ik_1j}, \mathbf{P}_{ijk_2} + \mathbf{P}_{ik_2j}\}$$

$$(5.10) \quad = 2 - \min_{\mathcal{I} \subset [n]} \min_j \left(\min_{k_1} \sum_{i \in \mathcal{I}} (\mathbf{P}_{ijk_1} + \mathbf{P}_{ik_1j}) + \min_{k_2} \sum_{i \notin \mathcal{I}} (\mathbf{P}_{ijk_2} + \mathbf{P}_{ik_2j}) \right).$$

Proof. Equations (5.5), (5.7), and (5.8) derive from the identity $\mathcal{T}_R(\mathbf{P}) = \mathcal{T}_L(\mathbf{P}^S)$ and (5.2), (5.3), and (5.4), respectively. Now, define $\mathbf{Q} = \frac{1}{2}(\mathbf{P} + \mathbf{P}^S)$. Note that if \mathbf{P} is stochastic, then also \mathbf{Q} is stochastic. Since

$$\|\mathbf{P}xy + \mathbf{P}yx\|_1 = \|\mathbf{P}xy + \mathbf{P}^Sxy\|_1 = 2 \left\| \frac{1}{2} (\mathbf{P} + \mathbf{P}^S) xy \right\|_1 = 2\|\mathbf{Q}xy\|_1,$$

we have $\mathcal{T}(\mathbf{P}) = 2\mathcal{T}_L(\mathbf{Q})$. Hence, (5.6), (5.9), and (5.10) derive from the latter identity and (5.2), (5.3), and (5.4), respectively. ■

By (5.3), (5.7), and (5.9), it is immediate to observe that for a stochastic tensor \mathbf{P} it holds that $0 \leq \mathcal{T}_L(\mathbf{P}), \mathcal{T}_R(\mathbf{P}) \leq 1$ and

$$(5.11) \quad 0 \leq \mathcal{T}(\mathbf{P}) \leq \mathcal{T}_L(\mathbf{P}) + \mathcal{T}_R(\mathbf{P}) \leq 2.$$

Stronger inequalities can be easily obtained for positive tensors, as shown in the next result.

Corollary 5.3. *Let $\mathbf{P} \in \mathbb{R}^{[3,n]}$ be a stochastic tensor. If there exists a positive number $\alpha > 0$ such that $\mathbf{P}_{ijk} \geq \alpha$ for all i, j, k , then*

$$\mathcal{T}_L(\mathbf{P}) \leq 1 - n\alpha, \quad \mathcal{T}_L(\mathbf{P}) \leq 1 - n\alpha, \quad \mathcal{T}(\mathbf{P}) \leq 2(1 - n\alpha).$$

Proof. The three inequalities in the claim follow immediately from (5.3), (5.7), and (5.9), respectively. ■

Remark 5.4. A close look at Theorem 5.1 reveals that, for any tensor $\mathbf{P} \in \mathbb{R}^{[3,n]}$, we have $\mathcal{T}_L(\mathbf{P}) = 0$ if and only if $\mathbf{P}_{ijk} = A_{ij}$ for some matrix $A \in \mathbb{R}^{n \times n}$. In particular, \mathbf{P} is stochastic if and only if A is stochastic. Analogously, from Corollary 5.2 we derive that $\mathcal{T}_R(\mathbf{P}) = 0$ if and only if $\mathbf{P}_{ijk} = A_{ik}$ for some matrix A . Consequently, $\mathcal{T}_L(\mathbf{P}) + \mathcal{T}_R(\mathbf{P}) = 0$ if and only if $\mathbf{P}_{ijk} = v_i$ for some vector v . It is not difficult to prove that the latter is also equivalent to $\mathcal{T}(\mathbf{P}) = 0$. Hence, if \mathbf{P} is nonzero, we have

$$\mathcal{T}_L(\mathbf{P}) + \mathcal{T}_R(\mathbf{P}) = 0 \iff \mathcal{T}(\mathbf{P}) = 0 \iff \text{rank}(\mathbf{P}) = 1.$$

In the matrix case, a coefficient of ergodicity τ is called proper when the identity $\tau(P) = 0$ for a stochastic matrix P is equivalent to the condition $\text{rank}(P) = 1$; see [31, 56]. For example, both the Birkhoff coefficient τ_H and all the norm-based ergodicity coefficients τ_p are proper. By extending that definition to the tensor case, we can say that \mathcal{T} is proper, while \mathcal{T}_L and \mathcal{T}_R are not proper.

The remark above shows that Property 3 of Theorem 4.1 carries over to the higher-order setting. In the next subsection 5.1 we show that also properties 1 and 2 of that theorem enjoy a tensor counterpart. In subsection 6.1, instead, we show how the perturbation result of Theorem 4.2 transfers to stochastic tensors.

5.1. Bounding the variation of higher-order coefficients. When working with stochastic tensors, it is quite natural to endow $\mathbb{R}^{[3,n]}$ with the norm

$$\|\mathbf{P}\|_1 = \max_{\|x\|_1 = \|y\|_1 = 1} \|\mathbf{P}xy\|_1 = \max_{j,k} \sum_i |\mathbf{P}_{ijk}|,$$

so that, if \mathbf{P} is stochastic, we have $\|\mathbf{P}\|_1 = 1$.

With the next theorem we prove a Lipschitz-continuity condition for the higher-order ergodicity coefficients with respect to the tensor 1-norm above.

Theorem 5.5. *For arbitrary $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{[3,n]}$ we have*

$$|\mathcal{T}_*(\mathbf{P}) - \mathcal{T}_*(\mathbf{Q})| \leq \mathcal{T}_*(\mathbf{P} - \mathbf{Q}) \leq \|\mathbf{P} - \mathbf{Q}\|_1$$

where \mathcal{T}_* is any of \mathcal{T}_L or \mathcal{T}_R . Moreover,

$$|\mathcal{T}(\mathbf{P}) - \mathcal{T}(\mathbf{Q})| \leq \mathcal{T}(\mathbf{P} - \mathbf{Q}) \leq 2\|\mathbf{P} - \mathbf{Q}\|_1.$$

Proof. Let $\mathcal{T}_* = \mathcal{T}_L$, the other case being completely analogous. Suppose that $\mathcal{T}_L(\mathbf{P}) \geq \mathcal{T}_L(\mathbf{Q})$. Hence, for some $x \in \mathcal{S}_1$ and $y \in \mathcal{Z}_1$ we have

$$\mathcal{T}_L(\mathbf{P}) = \|\mathbf{P}xy\|_1 \leq \|(\mathbf{P} - \mathbf{Q})xy\|_1 + \|\mathbf{Q}xy\|_1 \leq \mathcal{T}_L(\mathbf{P} - \mathbf{Q}) + \mathcal{T}_L(\mathbf{Q}).$$

Hence, $\mathcal{T}_L(\mathbf{P}) - \mathcal{T}_L(\mathbf{Q}) \leq \mathcal{T}_L(\mathbf{P} - \mathbf{Q})$. By reversing the roles of \mathbf{P} and \mathbf{Q} we obtain $\mathcal{T}_L(\mathbf{Q}) - \mathcal{T}_L(\mathbf{P}) \leq \mathcal{T}_L(\mathbf{P} - \mathbf{Q})$, and we arrive at the first claim. Analogously, for some $x \in \mathcal{S}_1$ and $y \in \mathcal{Z}_1$ we have

$$\begin{aligned} \mathcal{T}(\mathbf{P}) &= \|\mathbf{P}xy + \mathbf{P}yx\|_1 \leq \|(\mathbf{P} - \mathbf{Q})xy + (\mathbf{P} - \mathbf{Q})yx\|_1 + \|\mathbf{Q}xy + \mathbf{Q}yx\|_1 \\ &\leq \mathcal{T}(\mathbf{P} - \mathbf{Q}) + \mathcal{T}(\mathbf{Q}). \end{aligned}$$

The inequality $\mathcal{T}(\mathbf{Q}) - \mathcal{T}(\mathbf{P}) \leq \mathcal{T}(\mathbf{P} - \mathbf{Q})$ follows from the preceding one by exchanging \mathbf{P} and \mathbf{Q} , and the second claim follows. The rightmost inequalities follow immediately from the definition of the ergodicity coefficients. \blacksquare

6. Second-order stochastic processes and Z -eigenvectors. In this section we prove an analogous of Theorem 4.3 for tensor Z -eigenvectors. Precisely, given \mathbf{P} stochastic, we provide a new condition that ensures the existence and uniqueness of a positive vector $x \in \mathcal{S}_1$ such that $x = \mathbf{P}xx$. Moreover, we show that under the same condition the higher-order power method $x_{t+1} = \mathbf{P}x_tx_t$, which is the prototypical nonlinear Markov process [34], always converges to x , and we provide an analogous, but stronger, condition that guarantees the global convergence of the alternate scheme $x_{t+1} = \mathbf{P}x_tx_{t-1}$.

The next theorem provides the tensor analogous of Theorem 4.3.

Theorem 6.1. *If \mathbf{P} is stochastic, then $\mathcal{T}(\mathbf{P})$ is the best Lipschitz constant of the quadratic map $f : \mathcal{S}_1 \mapsto \mathcal{S}_1$ given by $f(x) = \mathbf{P}xx$, that is,*

$$\mathcal{T}(\mathbf{P}) = \tau_1(f) = \sup_{x,y \in \mathcal{S}_1} \frac{\|\mathbf{P}xx - \mathbf{P}yy\|_1}{\|x - y\|_1}.$$

Therefore, if $\mathcal{T}(\mathbf{P}) < 1$, then there exists a unique Z -eigenvector $x \in \mathcal{S}_1$ such that $\mathbf{P}xx = x$. Moreover, the higher-order power method $x_{t+1} = \mathbf{P}x_tx_t$ converges to x for any $x_0 \in \mathcal{S}_1$, and

$$\|x_t - x\|_1 \leq \mathcal{T}(\mathbf{P})^t \|x_0 - x\|_1.$$

Proof. Let $f : \mathcal{S}_1 \mapsto \mathcal{S}_1$ be given by $f(x) = \mathbf{P}xx$. Let $\mathbf{Q} = \frac{1}{2}(\mathbf{P} + \mathbf{P}^S)$. Note that \mathbf{Q} is a stochastic tensor such that $\mathbf{Q} = \mathbf{Q}^S$. Moreover, the equation $f(x) = x$ is equivalent to $\mathbf{Q}xx = x$. Then, for all $x, y \in \mathcal{S}_1$ we have

$$\begin{aligned} f(x) - f(y) &= \mathbf{Q}xx - \mathbf{Q}yy + \mathbf{Q}xy - \mathbf{Q}xy \\ &= \mathbf{Q}xx - \mathbf{Q}yy + \mathbf{Q}yx - \mathbf{Q}xy = \mathbf{Q}(x + y)(x - y). \end{aligned}$$

Hence,

$$\begin{aligned} \tau_1(f) &= \max_{x,y \in \mathcal{S}_1} \frac{\|f(x) - f(y)\|_1}{\|x - y\|_1} = \max_{x,y \in \mathcal{S}_1} \frac{2\|\mathbf{Q}\left(\frac{1}{2}x + \frac{1}{2}y\right)(x - y)\|_1}{\|x - y\|_1} \\ &= \max_{v \in \mathcal{S}_1} \max_{w \in \mathcal{Z}_1} 2\|\mathbf{Q}vw\|_1 = 2\mathcal{T}_L(\mathbf{Q}). \end{aligned}$$

Since $2\mathcal{T}_L(\mathbf{Q}) = \mathcal{T}(\mathbf{P})$, we obtain the first part of the claim. In particular, we get $\|f(x) - f(y)\|_1 \leq \mathcal{T}(\mathbf{P})\|x - y\|_1$ for any $x, y \in \mathcal{S}_1$. Hence, if $\mathcal{T}(\mathbf{P}) < 1$, then f is contractive with respect to the 1-norm. By the Banach fixed point theorem, there exists a unique fixed point $x \in \mathcal{S}_1$ such that $x = f(x)$. Moreover, the iteration $x_{t+1} = f(x_t)$ converges to x with $\|x_t - x\|_1 \leq \mathcal{T}(\mathbf{P})\|x_{t-1} - x\|_1$ for any $x_0 \in \mathcal{S}_1$, and the proof is complete. ■

We note in passing that the following result, which has been derived from a well-known uniqueness result in the fixed point theory several times by different authors [19, 38, 39], is a direct consequence of the theorem above.

Corollary 6.2. *If $\mathbf{P} \in \mathbb{R}^{[3,n]}$ is a stochastic tensor such that $\mathbf{P}_{ijk} > 1/(2n)$ for all i, j, k , then there exists a unique Z -eigenvector $x \in \mathcal{S}_1$ such that $\mathbf{P}xx = x$ and the higher-order power method $x_{t+1} = \mathbf{P}x_tx_t$ converges to x for any $x_0 \in \mathcal{S}_1$.*

Proof. In the stated hypotheses we have $\mathcal{T}(\mathbf{P}) < 1$ by virtue of Corollary 5.3. Hence, the claim is a direct consequence of Theorem 6.1. \blacksquare

Given a stochastic tensor \mathbf{P} and two initial points $x_0, x_{-1} \in \mathcal{S}_1$, the following alternate higher-order power method has been considered in [29]:

$$(6.1) \quad x_{t+1} = \mathbf{P}x_t x_{t-1}, \quad t = 0, 1, 2, \dots$$

Note that this coincides with the second-order stochastic process described in (3.4). In [29] the convergence of (6.1) has been proven when $P_{ijk} > 1/(2n)$ and under restrictive hypotheses on the choice of x_0 and x_{-1} . The following theorem provides a condition in terms of $\mathcal{T}_L(\mathbf{P})$ and $\mathcal{T}_R(\mathbf{P})$ that ensures that (6.1) converges globally to the unique fixed point of \mathbf{P} .

Theorem 6.3. *Let \mathbf{P} be a stochastic tensor, and let $s = \mathcal{T}_L(\mathbf{P}) + \mathcal{T}_R(\mathbf{P})$. If $s < 1$, then the iteration (6.1) converges to the unique Z-eigenvector $x \in \mathcal{S}_1$ such that $x = \mathbf{P}xx$. In fact, for all $x_0, x_{-1} \in \mathcal{S}_1$ and $t = 0, 1, \dots$ it holds that*

$$\|x_{t+1} - x\|_1 \leq s^{\lceil(t+1)/2\rceil} \max\{\|x_0 - x\|_1, \|x_{-1} - x\|_1\}.$$

Proof. First notice that the assumption $\mathcal{T}_L(\mathbf{P}) + \mathcal{T}_R(\mathbf{P}) < 1$ implies $\mathcal{T}(\mathbf{P}) < 1$; thus, by Theorem 6.1, there exists a unique positive $x \in \mathcal{S}_1$ such that $x = \mathbf{P}xx$. We have

$$\begin{aligned} x_{t+1} - x &= \mathbf{P}x_t x_{t-1} - \mathbf{P}xx = \mathbf{P}x_t x_{t-1} - \mathbf{P}x_t x + \mathbf{P}x_t x - \mathbf{P}xx \\ &= \mathbf{P}x_t(x_{t-1} - x) + \mathbf{P}(x_t - x)x. \end{aligned}$$

Thus, for any $t \geq 0$,

$$\begin{aligned} \|x_{t+1} - x\|_1 &\leq \mathcal{T}_L(\mathbf{P})\|x_{t-1} - x\|_1 + \mathcal{T}_R(\mathbf{P})\|x_t - x\|_1 \\ &\leq (\mathcal{T}_L(\mathbf{P}) + \mathcal{T}_R(\mathbf{P})) \max\{\|x_t - x\|_1, \|x_{t-1} - x\|_1\}. \end{aligned}$$

In particular, the claim is true for $t = 0$. The proof is completed by a simple inductive argument. Indeed, let $m = \max\{\|x_0 - x\|_1, \|x_{-1} - x\|_1\}$ and $\epsilon_t = \|x_t - x\|_1$ to simplify notations. For $t > 0$, suppose the claim true up to $t - 1$. Then,

$$\epsilon_{t+1} \leq s \max\{\epsilon_t, \epsilon_{t-1}\} \leq s \max\left\{s^{\lceil t/2 \rceil}, s^{\lceil(t-1)/2 \rceil}\right\} m = s^{\lceil(t+1)/2 \rceil} m,$$

and the theorem is proved. \blacksquare

6.1. A perturbation result for the stochastic Z-eigenvector. A fundamental perturbation analysis problem is to obtain quality bounds on the variation of the ergodic distribution of the nonnegative stochastic tensor \mathbf{P} , when \mathbf{P} is perturbed. The following result provides a bound in terms of the higher-order norm-based ergodicity coefficients and represents a tensor counterpart of Theorem 4.2.

Theorem 6.4. *Let \mathbf{P} and its perturbation \mathbf{P}' be two stochastic tensors in $\mathbb{R}^{[3,m]}$. If $\mathcal{T}(\mathbf{P}) < 1$, then the stochastic solution of $x = \mathbf{P}xx$ is unique, and for any stochastic vector x' such that $x' = \mathbf{P}'x'x'$ it holds that*

$$\|x - x'\|_1 \leq \frac{\|\mathbf{P} - \mathbf{P}'\|_1}{1 - \mathcal{T}(\mathbf{P})}.$$

Proof. Suppose first that both \mathbf{P} and \mathbf{P}' are S -symmetric. By adding and subtracting $\mathbf{P}x'x'$ we have

$$\begin{aligned}\|x - x'\|_1 &= \|\mathbf{P}xx - \mathbf{P}'x'x' + \mathbf{P}x'x' - \mathbf{P}x'x'\|_1 \\ &\leq \|\mathbf{P}x(x - x') + \mathbf{P}(x - x')x'\|_1 + \|(\mathbf{P} - \mathbf{P}')x'x'\|_1 \\ &= 2 \left\| \mathbf{P} \left(\frac{1}{2}x + \frac{1}{2}x' \right) (x - x') \right\|_1 + \|(\mathbf{P} - \mathbf{P}')x'x'\|_1 \\ &\leq 2\mathcal{T}_L(\mathbf{P})\|x - x'\|_1 + \|\mathbf{P} - \mathbf{P}'\|_1 = \mathcal{T}(\mathbf{P})\|x - x'\|_1 + \|\mathbf{P} - \mathbf{P}'\|_1.\end{aligned}$$

Rearranging terms we find $\|x - x'\|_1(1 - \mathcal{T}(\mathbf{P})) \leq \|\mathbf{P} - \mathbf{P}'\|_1$, and the claim follows.

In the general case, define $\mathbf{Q} = \frac{1}{2}(\mathbf{P} + \mathbf{P}^S)$ and $\mathbf{Q}' = \frac{1}{2}(\mathbf{P}' + \mathbf{P}'^S)$, and repeat the previous arguments. Finally, note that $\mathcal{T}(\mathbf{Q}) = \mathcal{T}(\mathbf{P})$ and $\|\mathbf{Q} - \mathbf{Q}'\|_1 \leq \|\mathbf{P} - \mathbf{P}'\|_1$. ■

6.2. Convergence of a class of vertex-reinforced random walks. Vertex-reinforced random walks are another important example of higher-order discrete-time stochastic process $\{X(t)\}_t$ on the state space $\{1, \dots, n\}$, where the state transitions at time t depend on the whole history $X(0), \dots, X(t-1)$ [7, 47]. Starting from an initial state $X(0) \in \{1, \dots, n\}$, the process evolves according to the formulas

$$(6.2) \quad \begin{aligned}\mathbb{P}(X(t+1) = i | \mathcal{F}_t) &= \mathcal{M}(y_t)_{i, X(t)}, \\ (y_t)_j &= \frac{1 + \sum_{i=1}^t [X(i) = j]}{t+n},\end{aligned}$$

where \mathcal{F}_t is the σ -field generated by $X(1), \dots, X(t)$ and \mathcal{M} is a map from \mathcal{S}_1 to the set of stochastic $n \times n$ matrices. The vector y_t , which is called the occupation vector, is an auxiliary stochastic vector that is introduced in order to record the history of the process. Indeed, the i th entry of y_t is proportional to the number of times the process visited state i up to the t th time step, plus one. Now, let x_t be the probability vector of $X(t)$, that is, the n -vector whose i th entry is $\mathbb{P}(X(t) = i)$. Then, the process (6.2) can be equivalently described via the coupled equations

$$x_{t+1} = \mathcal{M}(y_t)x_t, \quad y_t = \frac{1}{t+n} \sum_{s=1}^t x_s + \frac{1}{t+n} \mathbb{1}.$$

When \mathcal{M} is linear, there exists a stochastic tensor \mathbf{P} such that $\mathcal{M}(v)_{ij} = \sum_k \mathbf{P}_{ijk} v_k$ and the corresponding stochastic process is the so-called spacey random walk, introduced in [8]. In this case, with minor notation changes with respect to the original version, the previous iteration can be recast as

$$(6.3) \quad \begin{cases} x_{t+1} = \mathbf{P}x_t y_t, \\ y_{t+1} = \frac{1}{t+1}x_t + \frac{t}{t+1}y_t. \end{cases}$$

On the basis of key results by Benaïm [7], Benson, Gleich, and Lim established the convergence of the spacey random walk in terms of the convergence of a certain ordinary differential equation to a stable equilibrium, and one auxiliary condition placed on \mathbf{P} [8, Theorem 9].

However, only the convergence of the occupation vectors $\{y_t\}$ (which corresponds to the convergence in the Cesàro average sense of the random variables $X(t)$) can be derived from the results in [7, 8]. In fact, the second equation in (6.3) yields

$$y_{t+1} - y_t = \frac{1}{t+1}(x_t - y_t).$$

Hence, even if the sequence $\{y_t\}$ has a limit and the left-hand side converges to zero, that does not imply the convergence of the sequence $\{x_t\}$.

In what follows, we consider the following generalization of (6.3),

$$(6.4) \quad \begin{cases} x_{t+1} = \mathbf{P}x_ty_t, \\ y_{t+1} = c_tx_t + (1 - c_t)y_t \end{cases}$$

with $c_t \in [0, 1]$, and we show in the next theorem that, if the higher-order ergodicity coefficients $\mathcal{T}_L(\mathbf{P})$ and $\mathcal{T}_R(\mathbf{P})$ are small enough, then the stochastic process (6.4) is globally convergent, provided that the sequence $\{c_t\}$ is not too small. This requirement on $\{c_t\}$ can be seen as a condition that avoids the process from freezing along the way on a limit point that is far away from the Z -eigenvector of \mathbf{P} . In fact, the possibility of such a behavior has been shown in [13] for a stochastic process closely related to (6.4).

Theorem 6.5. *Let the sequence $\{c_t\}$ in (6.4) be nonincreasing and such that*

$$(6.5) \quad \sum_{t=1}^{\infty} c_t = +\infty.$$

If $\mathcal{T}_L(\mathbf{P}) + \mathcal{T}_R(\mathbf{P}) < 1$, then the vertex-reinforced random walk (6.4) converges globally, i.e., for any starting points $x_0, y_0 \in \mathcal{S}_1$ we have

$$\lim_{t \rightarrow \infty} \|x_t - x\|_1 = \lim_{t \rightarrow \infty} \|y_t - x\|_1 = 0,$$

where x is the unique stochastic solution of $x = \mathbf{P}xx$. Moreover, if there exists a positive constant α such that $c_t \geq \alpha$, then the convergence is linear.

Proof. Firstly, note that, in the stated hypotheses, the vector x exists and is unique owing to Theorem 6.1. Subtracting the identity $x = \mathbf{P}xx$ from (6.4) we obtain

$$\begin{aligned} x_{t+1} - x &= \mathbf{P}(x_t - x)y_t + \mathbf{P}x(y_t - x), \\ y_{t+1} - x &= c_t(x_t - x) + (1 - c_t)(y_t - x). \end{aligned}$$

Let $\alpha_t = \|x_t - x\|_1$ and $\beta_t = \|y_t - x\|_1$. Using vector inequalities, we have

$$\begin{bmatrix} \alpha_{t+1} \\ \beta_{t+1} \end{bmatrix} \leq \begin{bmatrix} \mathcal{T}_L(\mathbf{P}) & \mathcal{T}_R(\mathbf{P}) \\ c_t & 1 - c_t \end{bmatrix} \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix}.$$

Let $\gamma_t = \|(\alpha_t, \beta_t)^T\|_\infty = \max\{\alpha_t, \beta_t\}$. For notational simplicity, let $\ell = \mathcal{T}_L(\mathbf{P})$, $r = \mathcal{T}_R(\mathbf{P})$, and define

$$A_t = \begin{bmatrix} \ell & r \\ c_t & 1 - c_t \end{bmatrix}.$$

Hence, for $t = 1, 2, \dots$ we have

$$\gamma_{t+1} \leq \|A_t \cdots A_1 A_0\|_\infty \gamma_0.$$

Moreover, since $\gamma_{t+1} \leq \|A_t\|_\infty \gamma_t$ and $\|A_t\|_\infty = 1$, we have $\gamma_{t+1} \leq \gamma_t$, that is, the sequence $\{\gamma_t\}$ is nonincreasing. Now, for $t \geq 1$, consider the product $A_t A_{t-1}$. Simple computations show that

$$A_t A_{t-1} = \begin{bmatrix} \ell^2 + r c_{t-1} & r(\ell + 1 - c_{t-1}) \\ \ell c_t + (1 - c_t)c_{t-1} & r c_t + (1 - c_t)(1 - c_{t-1}) \end{bmatrix},$$

$$\|A_t A_{t-1}\|_\infty = \max\{r + \ell(\ell + r), 1 - c_t(\ell + r)\} < 1.$$

In particular, if $\lim_{t \rightarrow \infty} c_t = 0$, then there exists an integer t_* such that for $t \geq t_*$ it holds that $\|A_{2t} A_{2t-1}\|_\infty = 1 - c_{2t}(\ell + r)$. Consequently, we have

$$\gamma_{2t+1} \leq \left(\prod_{j=t_*}^t \|A_{2j} A_{2j-1}\|_\infty \right) \|A_{2t_*-2} \cdots A_1 A_0\|_\infty \gamma_0 = C \prod_{j=t_*}^t (1 - c_{2j}(\ell + r)),$$

where $C = \|A_{2t_*-2} \cdots A_1 A_0\|_\infty \gamma_0$. In order to prove that $\lim_{t \rightarrow \infty} \gamma_t = 0$ it is sufficient to discuss the limit

$$\lim_{t \rightarrow \infty} \prod_{j=1}^t (1 - c_{2j}(\ell + r)),$$

which exists and is nonnegative since all factors belong to $(0, 1)$. By a known result on the convergence of infinite products (see, e.g., [32, page 223]), the preceding limit is positive if and only if the series

$$\sum_{j=1}^{\infty} c_{2j}(\ell + r)$$

is convergent. Hence, if (6.5) holds, then $\lim_{t \rightarrow \infty} \gamma_t = 0$, and we are done.

On the other hand, if $c_t \geq \alpha > 0$, then there exists a number $s \in (0, 1)$ such that $\|A_t A_{t-1}\|_\infty \leq s$. Hence,

$$\gamma_{2t} \leq \prod_{j=0}^{t-1} \|A_{2j+1} A_{2j}\|_\infty \gamma_0 \leq s^t \gamma_0,$$

and the last claim follows. ■

Note that both the spacey random walk (6.3) and the second-order Markov chain (6.1) are particular cases of the stochastic processes (6.4), corresponding to the choices $c_t = \frac{1}{t+1}$ and $c_t = 1$, respectively. Observe that both these choices satisfy the assumption (6.5). Thus, the convergence condition for the second-order Markov chain of Theorem 6.3 also follows as a consequence of Theorem 6.5. Moreover, we obtain the following convergence result for the spacey random walk which, to the best of our knowledge, is the first result that gives explicit conditions that guarantee the convergence of both the occupation vector and the density distribution for this stochastic process.

Corollary 6.6. *If $\mathcal{T}_L(\mathbf{P}) + \mathcal{T}_R(\mathbf{P}) < 1$, then the spacey random walk (6.3) converges globally, i.e., for any starting points $x_0, y_0 \in \mathcal{S}_1$ we have*

$$\lim_{t \rightarrow \infty} \|x_t - x\|_1 = \lim_{t \rightarrow \infty} \|y_t - x\|_1 = 0,$$

where x is the unique stochastic solution of $x = \mathbf{P}xx$.

Proof. It suffices to observe that the coefficient sequence $\{c_t\}$ of the spacey random walk is a trailing subsequence of the harmonic sequence; hence the hypothesis (6.5) is fulfilled. ■

7. Comparison with previous works. In this section we discuss how the newly proposed higher-order ergodicity coefficient $\mathcal{T}(\mathbf{P})$, based on the 1-norm, compares with previous works. In particular, we compare it with the contraction ratios proposed by Gautier and Tudisco in [24], where the Hilbert metric is used to quantify the contractivity of multilinear operators, and with the coefficients introduced by Li and Ng in [39] in order to characterize the uniqueness of stationary distributions of stochastic tensors.

7.1. Higher-order Birkhoff coefficients. When d is the Hilbert projective metric d_H defined in (4.2), the ergodicity coefficient (4.1) is known as Birkhoff contraction ratio, and the renowned Birkhoff–Hopf theorem provides an explicit formula for such a coefficient when f is a linear map. Recently, the Birkhoff–Hopf theorem has been extended to the case of multilinear mappings [24]. We review that theorem in the following, for the case of a bilinear map $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ described by a cubic tensor \mathbf{P} as $f(x, y) = \mathbf{P}xy$.

Theorem 7.1. *Let $\mathbf{P} \in \mathbb{R}^{[3,n]}$ be a nonnegative tensor, let*

$$\Delta(\mathbf{P}) = \max_{i_1, j_1, k_1, i_2, j_2, k_2} \frac{\mathbf{P}_{i_1 j_1 k_1} \mathbf{P}_{i_2 j_2 k_2}}{\mathbf{P}_{i_1 j_2 k_1} \mathbf{P}_{i_2 j_1 k_2}},$$

and let $\kappa(\mathbf{P}) = \tanh(\frac{1}{4} \log \Delta(\mathbf{P}))$. Then

$$d_H(\mathbf{P}xy, \mathbf{P}x'y') \leq \kappa(\mathbf{P})d_H(x, x') + \kappa(\mathbf{P}^S)d_H(y, y').$$

From Theorem 7.1 we immediately derive a formula for the higher-order Birkhoff ergodicity coefficient for stochastic tensors and the corresponding analogous of Theorem 6.1. Precisely, we have the following result.

Corollary 7.2. *Let $\mathbf{P} \in \mathbb{R}^{[3,n]}$ be a stochastic tensor, and let*

$$\mathcal{T}_H(\mathbf{P}) = 2\kappa(\mathbf{P} + \mathbf{P}^S) = 2 \tanh\left(\frac{1}{4} \log \widehat{\Delta}(\mathbf{P})\right),$$

where

$$\widehat{\Delta}(\mathbf{P}) = \max_{i_1, j_1, k_1, i_2, j_2, k_2} \frac{(\mathbf{P}_{i_1 j_1 k_1} + \mathbf{P}_{i_1 k_1 j_1})(\mathbf{P}_{i_2 j_2 k_2} + \mathbf{P}_{i_2 k_2 j_2})}{(\mathbf{P}_{i_1 j_1 k_2} + \mathbf{P}_{i_1 k_2 j_1})(\mathbf{P}_{i_2 j_2 k_1} + \mathbf{P}_{i_2 k_1 j_2})}.$$

If $\mathcal{T}_H(\mathbf{P}) < 1$, then there exists a unique Z -eigenvector $x \in \mathcal{S}_1$ such that $\mathbf{P}xx = x$ and the higher-order power iteration $x_{t+1} = \mathbf{P}x_t x_t$ converges to x for any starting point $x_0 \in S_1$.

Proof. Consider the S -symmetric tensor $\mathbf{Q} = \frac{1}{2}(\mathbf{P} + \mathbf{P}^S)$. Note that $\Delta(\mathbf{Q}) = \widehat{\Delta}(\mathbf{P})$ and thus $\kappa(\mathbf{Q}) = \frac{1}{2}\mathcal{T}_H(\mathbf{P})$. Therefore, using the identity $\mathbf{P}xx = \mathbf{Q}xx$, which holds for all $x \in \mathbb{R}^n$, the triangle inequality for d_H , and Theorem 7.1, we have

$$\begin{aligned} d_H(\mathbf{P}xx, \mathbf{P}yy) &= d_H(\mathbf{Q}xx, \mathbf{Q}yy) \leq d_H(\mathbf{Q}xx, \mathbf{Q}xy) + d_H(\mathbf{Q}xy, \mathbf{Q}yy) \\ &\leq \kappa(\mathbf{Q})[d_H(x, x) + d_H(x, y) + d_H(x, y) + d_H(y, y)] = \mathcal{T}_H(\mathbf{P})d_H(x, y). \end{aligned}$$

This shows that $x \mapsto \mathbf{P}xx$ is a contraction with respect to the Hilbert metric. As (\mathcal{S}_1, d_H) is a complete metric space, the proof continues as that of Theorem 6.1. ■

Note that, similarly to the 1-norm case, $\mathcal{T}_H(\mathbf{P}) = 0$ if and only if \mathbf{P} has rank one, that is, \mathcal{T}_H is proper. However, while $\mathcal{T}_H(\mathbf{P}) = 2$ for any tensor \mathbf{P} not of rank one and having at least one zero entry, $\mathcal{T}(\mathbf{P})$ can be smaller than one even for sparse tensors. For example, if \mathbf{P} is the tensor

$$\mathbf{P} = \frac{1}{2} \left[\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \right],$$

then one easily verifies that $\mathcal{T}_H(\mathbf{P}) = 2$, while $\mathcal{T}(\mathbf{P}) = 1/2$.

The left panel of Figure 7.1 scatter plots these two coefficients computed on a set of ten thousand random stochastic $n \times n \times n$ tensors with size n between 2 and 10. In the matrix case it is well known that, for any stochastic matrix P , it holds that $\tau_1(P) \leq \mathcal{T}_H(P)$; see [56, section 3.4]. While the numerical comparison shown in Figure 7.1 suggests the inequality $\mathcal{T}(\mathbf{P}) \leq \mathcal{T}_H(\mathbf{P})$, an explicit comparison between the 1-norm and the Birkhoff higher-order coefficients $\mathcal{T}(\mathbf{P})$ and $\mathcal{T}_H(\mathbf{P})$, for general tensors, is out of scope and is left open to future work.

7.2. Li and Ng's coefficients. Given a stochastic tensor $\mathbf{P} \in \mathbb{R}^{[3,n]}$, consider the following quantities introduced in [37, 39]:

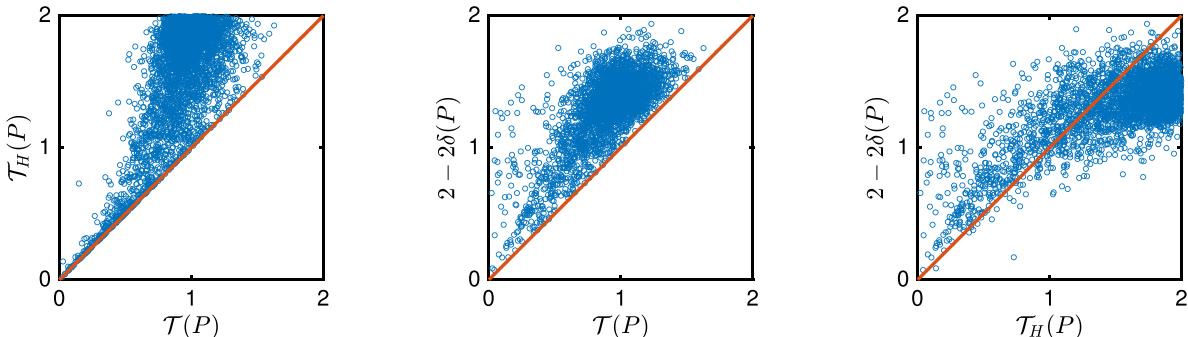


Figure 7.1. Scatter plot of different coefficients over 10,000 random $n \times n \times n$ stochastic tensors \mathbf{P} with size n chosen uniformly at random within $\{2, \dots, 10\}$.

$$(7.1) \quad \gamma(\mathbf{P}) = \min_{\mathcal{I} \subset [n]} \left\{ \min_k \left(\min_{j \in \mathcal{I}} \sum_{i \notin \mathcal{I}} \mathbf{P}_{ijk} + \min_{j \notin \mathcal{I}} \sum_{i \in \mathcal{I}} \mathbf{P}_{ijk} \right) + \min_j \left(\min_{k \in \mathcal{I}} \sum_{i \notin \mathcal{I}} \mathbf{P}_{ijk} + \min_{k \notin \mathcal{I}} \sum_{i \in \mathcal{I}} \mathbf{P}_{ijk} \right) \right\},$$

$$(7.2) \quad \delta(\mathbf{P}) = \min_{\mathcal{I} \subset [n]} \left(\min_{j,k} \sum_{i \notin \mathcal{I}} \mathbf{P}_{ijk} + \min_{j,k} \sum_{i \in \mathcal{I}} \mathbf{P}_{ijk} \right).$$

Li and Ng proved in [39] two conditions for the uniqueness of the stationary distribution and the convergence of the iteration $x_{t+1} = \mathbf{P}x_t x_t$ in terms of the entries of \mathbf{P} , which we review in the following.

Theorem 7.3 (see [39]). *Let $\mathbf{P} \in \mathbb{R}^{[3,n]}$ be a stochastic tensor. If $\gamma(\mathbf{P}) > 1$, then there exists an unique solution $x \in \mathcal{S}_1$ of the equation $x = \mathbf{P}xx$. Moreover, the iteration $x_{t+1} = \mathbf{P}x_t x_t$ converges to x .*

As $\gamma(\mathbf{P}) \geq 2\delta(\mathbf{P})$, the following consequence is immediate.

Corollary 7.4. *Let $\mathbf{P} \in \mathbb{R}^{[3,n]}$ be a stochastic tensor. If $\delta(\mathbf{P}) > 1/2$, then all the claims in the preceding theorem are true.*

Moreover, we recall from [37, Theorem 4] the three-mode case of a perturbation bound for the stationary probability vector of a stochastic tensor of order $m > 2$.

Theorem 7.5 (see [37]). *Let \mathbf{P} and its perturbation \mathbf{P}' be two stochastic tensors in $\mathbb{R}^{[3,m]}$. If $\delta(\mathbf{P}) > 1/2$, then the stochastic solution of $x = \mathbf{P}xx$ is unique, and for any stochastic vector x' such that $x' = \mathbf{P}'x'x'$ it holds that*

$$\|x - x'\|_1 \leq \frac{\|\mathbf{P} - \mathbf{P}'\|_1}{2\delta(\mathbf{P}) - 1}.$$

In what follows, we aim to compare the above results with the ones we proved in the previous sections. First, we prove a special characterization of $\delta(\mathbf{P})$ in (7.2), which provides an explicit formula for $\delta(\mathbf{P})$ in terms of the entries of \mathbf{P} .

Lemma 7.6. *Let $\mathbf{P} \in \mathbb{R}^{[3,n]}$ be a stochastic tensor. Then*

$$\delta(\mathbf{P}) = 1 - \frac{1}{2} \max_{j_1, j_2, k_1, k_2} \|\mathbf{P}e_{j_1}e_{k_1} - \mathbf{P}e_{j_2}e_{k_2}\|_1.$$

Proof. First, note that for any zero-sum vector $y \in \mathbb{R}^n$ it holds that $\sum_i |y_i| = 2 \max \{ \sum_{i \in \mathcal{I}} y_i : \mathcal{I} \subseteq \{1, \dots, n\} \}$. Let j_1, j_2, k_1, k_2 be fixed. Then we have

$$\begin{aligned} \|\mathbf{P}e_{j_1}e_{k_1} - \mathbf{P}e_{j_2}e_{k_2}\|_1 &= \sum_i |\mathbf{P}_{ij_1k_1} - \mathbf{P}_{ij_2k_2}| = 2 \max_{\mathcal{I} \subset [n]} \sum_{i \in \mathcal{I}} (\mathbf{P}_{ij_1k_1} - \mathbf{P}_{ij_2k_2}) \\ &= 2 \max_{\mathcal{I} \subset [n]} \left(1 - \sum_{i \notin \mathcal{I}} \mathbf{P}_{ij_1k_1} - \sum_{i \in \mathcal{I}} \mathbf{P}_{ij_2k_2} \right) = 2 - 2 \min_{\mathcal{I} \subset [n]} \left(\sum_{i \notin \mathcal{I}} \mathbf{P}_{ij_1k_1} + \sum_{i \in \mathcal{I}} \mathbf{P}_{ij_2k_2} \right). \end{aligned}$$

Therefore

$$\begin{aligned} 1 - \frac{1}{2} \max_{j_1, j_2, k_1, k_2} \|\mathbf{P}e_{j_1}e_{k_1} - \mathbf{P}e_{j_2}e_{k_2}\|_1 &= \frac{1}{2} \min_{j_1, j_2, k_1, k_2} (2 - \|\mathbf{P}e_{j_1}e_{k_1} - \mathbf{P}e_{j_2}e_{k_2}\|_1) \\ &= \min_{j_1, j_2, k_1, k_2} \min_{\mathcal{I} \subset [n]} \left(\sum_{i \notin \mathcal{I}} \mathbf{P}_{ij_1 k_1} + \sum_{i \in \mathcal{I}} \mathbf{P}_{ij_2 k_2} \right), \end{aligned}$$

which coincides with (7.2), after rearranging terms. \blacksquare

Using the characterization of $\delta(\mathbf{P})$ in the preceding lemma, the following theorem compares $\delta(\mathbf{P})$ and $\gamma(\mathbf{P})$ with the higher-order ergodic coefficient $\mathcal{T}(\mathbf{P})$.

Theorem 7.7. *Let $\mathbf{P} \in \mathbb{R}^{[3,n]}$ be stochastic. Then $\mathcal{T}(\mathbf{P}) \leq 2 - 2\delta(\mathbf{P})$. Moreover, if $\mathbf{P} = \mathbf{P}^S$, then $2 - \gamma(\mathbf{P}) \leq \mathcal{T}(\mathbf{P})$.*

Proof. The formulas (5.2) and (5.5) can be rewritten as

$$\mathcal{T}_L(\mathbf{P}) = \frac{1}{2} \max_{j, k_1, k_2} \|\mathbf{P}e_j(e_{k_1} - e_{k_2})\|_1, \quad \mathcal{T}_R(\mathbf{P}) = \frac{1}{2} \max_{j_1, j_2, k} \|\mathbf{P}(e_{j_1} - e_{j_2})e_k\|_1,$$

respectively. Using the preceding formulas and (5.11) it is immediate to obtain

$$\max_{j_1, j_2, k_1, k_2} \|\mathbf{P}e_{j_1}e_{k_1} - \mathbf{P}e_{j_2}e_{k_2}\|_1 \geq 2 \max\{\mathcal{T}_L(\mathbf{P}), \mathcal{T}_R(\mathbf{P})\} \geq \mathcal{T}_L(\mathbf{P}) + \mathcal{T}_R(\mathbf{P}) \geq \mathcal{T}(\mathbf{P}).$$

From Lemma 7.6 we conclude $1 - \delta(\mathbf{P}) \geq \frac{1}{2}\mathcal{T}(\mathbf{P})$, and this proves the first part of the claim. Furthermore, using the symmetry $\mathbf{P} = \mathbf{P}^S$, the formulas (7.1) and (5.10) simplify to

$$\begin{aligned} \gamma(\mathbf{P}) &= 2 \min_{\mathcal{I} \subset [n]} \min_j \left(\min_{k \in \mathcal{I}} \sum_{i \notin \mathcal{I}} \mathbf{P}_{ijk} + \min_{k \notin \mathcal{I}} \sum_{i \in \mathcal{I}} \mathbf{P}_{ijk} \right), \\ \mathcal{T}(\mathbf{P}) &= 2 - 2 \min_{\mathcal{I} \subset [n]} \min_j \left(\min_k \sum_{i \in \mathcal{I}} \mathbf{P}_{ijk} + \min_k \sum_{i \notin \mathcal{I}} \mathbf{P}_{ijk} \right). \end{aligned}$$

The inequality $2 - \gamma(\mathbf{P}) \leq \mathcal{T}(\mathbf{P})$ follows, and the proof is complete. \blacksquare

We conclude with several important remarks that we obtain as a consequence of the preceding results.

First, notice that the requirement $\delta(\mathbf{P}) > \frac{1}{2}$ appearing in Theorem 7.5 is stronger than the one of Theorem 6.4, namely, if $\delta(\mathbf{P}) > \frac{1}{2}$ holds, then $\mathcal{T}(\mathbf{P}) < 1$ must hold as well. Moreover, $2\delta(\mathbf{P}) - 1 \geq 1 - \mathcal{T}(\mathbf{P})$. Thus the right-hand side of Theorem 7.5 is larger than the one of Theorem 6.4. This shows that Theorem 6.4 is an improvement over Theorem 7.5.

On the other hand, the condition $\gamma(\mathbf{P}) > 1$ is weaker than $\mathcal{T}(\mathbf{P}) < 1$. Hence, the hypothesis in Theorem 7.3 ensuring uniqueness of the solution of $x = \mathbf{P}xx$ and convergence of the higher-order power method can be more general than the one in Theorem 6.1, at least when $\mathbf{P} = \mathbf{P}^S$. Additionally, it is important to point out that the inequality $\mathcal{T}(\mathbf{P}) < 1$ can be checked using $\mathcal{O}(n^4)$ arithmetic operations, while the computation of $\gamma(\mathbf{P})$ requires the solution of a nontrivial combinatorial optimization problem, which is in general significantly more expensive.

The central and the rightmost panels of Figure 7.1 compare numerically, via scatter plots, the condition $2 - 2\delta(\mathbf{P}) < 1$ and the ergodicity conditions $\mathcal{T}(\mathbf{P}) < 1$ and $\mathcal{T}_H(\mathbf{P}) < 1$ obtained via the higher-order ergodicity coefficients, on a test set of 10,000 randomly generated tensors with varying size.

8. Examples. We conclude with a number of example applications of Theorem 6.1. The examples here below further demonstrate the usefulness of the newly introduced higher-order ergodicity coefficients in a variety of contexts.

8.1. Multilinear PageRank. Given a stochastic tensor \mathbf{P} , a scalar $0 < \alpha < 1$, and a probability vector $v \in \mathcal{S}_1$, the multilinear PageRank is a solution of the equation

$$(8.1) \quad \alpha \mathbf{P}xx + (1 - \alpha)v = x.$$

This definition has been introduced by Gleich, Lim, and Yu [27] in analogy to the renowned Google's PageRank vector, defined as the solution of $\alpha Px + (1 - \alpha)v = x$ where P is a stochastic transition probability matrix. Pursuing that analogy, the solution of (8.1) gives the stationary probability of a stochastic process that with probability α behaves like the second-order Markov chain (6.1) and with probability $1 - \alpha$ teleports to a random state chosen according to the discrete density v .

A detailed analysis of the possibly multiple nonnegative solutions to (8.1) is provided by Meini and Poloni in [43]. They also discuss various first- and second-order iterative methods to compute a solution to (8.1). In particular, fixed point type methods are often a choice of preference, due to their inexpensive iterations and simple implementation. Also, these types of methods can be easily extrapolated achieving fast converge rates; see [18]. However, in practice one is interested in values of α not too far from 1, but, unlike the matrix case, requiring $\alpha < 1$ is not enough to ensure the uniqueness of the multilinear PageRank or the convergence of the fixed point iterates. In the original paper [27], the condition $\alpha < 1/2$ is proved to be sufficient to ensure both these properties (8.1). More recently, a tighter sufficient condition for the uniqueness of the multilinear PageRank has been proved by Li, Ng, and Vong [38], in terms of the following quantity:

$$\theta(\mathbf{P}, \sigma) = \max_{j, k_1, k_2} \sum_i |\mathbf{P}_{ijk_1} - \sigma_i| + |\mathbf{P}_{ik_2 j} - \sigma_i|,$$

where σ is any real vector. Precisely, Theorems 1 and 2 in [38] show that if there exists $\sigma \in \mathbb{R}^n$ such that $\alpha \theta(\mathbf{P}, \sigma) < 1$, then (8.1) has a unique nonnegative solution and the fixed point iteration for (8.1) converges to such a solution.

Theorem 6.1 provides a new condition that improves the range of values of α for which we can guarantee both the uniqueness of a nonnegative solution of (8.1) and the convergence of the associated fixed point iteration, as shown by the following result.

Corollary 8.1. *If $\alpha \mathcal{T}(\mathbf{P}) < 1$, then (8.1) has a unique solution $x \in \mathcal{S}_1$. Moreover, the fixed point iteration $x_{t+1} = \alpha \mathbf{P}x_t x_t + (1 - \alpha)v$ converges linearly to x with a convergence rate of at least $\alpha \mathcal{T}(\mathbf{P})$. Finally, it holds that $\|x - v\|_1 \leq 2\alpha$.*

Proof. Let $\mathbf{V} \in \mathbb{R}^{[3,n]}$ be the rank-one tensor $\mathbf{V}_{ijk} = v_i$. Since $\mathbf{V}xx = v$ for any $x \in \mathcal{S}_1$, (8.1) can be rewritten as $x = \mathbf{P}_\alpha xx$ where

$$(8.2) \quad \mathbf{P}_\alpha = \alpha \mathbf{P} + (1 - \alpha) \mathbf{V}.$$

By Theorem 6.1, the condition $\mathcal{T}(\mathbf{P}_\alpha) < 1$ guarantees uniqueness of the solution and convergence of the fixed point iteration. However, $\mathcal{T}(\mathbf{P}_\alpha) \leq \alpha \mathcal{T}(\mathbf{P}) + (1 - \alpha) \mathcal{T}(\mathbf{V}) = \alpha \mathcal{T}(\mathbf{P})$, due to the fact that $\mathcal{T}(\mathbf{V}) = 0$, as noted in Remark 5.4.

Finally, note that the vector v is characterized by the identity $v = \mathbf{V}vv = \mathbf{P}_0vv$. Hence, by considering \mathbf{P}_α as a perturbation of $\mathbf{V} = \mathbf{P}_0$, from Theorem 6.4 we get

$$\|x - v\|_1 \leq \frac{1}{1 - \mathcal{T}(\mathbf{V})} \|\mathbf{P}_\alpha - \mathbf{V}\|_1 = \alpha \|\mathbf{P} - \mathbf{V}\|_1 \leq 2\alpha$$

since $\mathcal{T}(\mathbf{V}) = 0$, and the proof is complete. \blacksquare

Note that the condition for the uniqueness given by Corollary 8.1 is always an improvement with respect to the one of [38]. In fact, using the formula (5.6) for $\mathcal{T}(\mathbf{P})$, for any $\sigma \in \mathbb{R}^n$ we have

$$\begin{aligned} 2\mathcal{T}(\mathbf{P}) &= \max_{j,k_1,k_2} \sum_i |\mathbf{P}_{ijk_1} - \mathbf{P}_{ijk_2} + \mathbf{P}_{ik_1j} - \mathbf{P}_{ik_2j} + 2\sigma_i - 2\sigma_i| \\ &\leq \max_{j,k_1,k_2} \sum_i |\mathbf{P}_{ijk_1} - \sigma_i| + |\mathbf{P}_{ijk_2} - \sigma_i| + |\mathbf{P}_{ik_1j} - \sigma_i| + |\mathbf{P}_{ik_2j} - \sigma_i| \\ &\leq \left[\max_{j,k_1,k_2} \sum_i |\mathbf{P}_{ijk_1} - \sigma_i| + |\mathbf{P}_{ik_2j} - \sigma_i| \right] + \left[\max_{j,k_1,k_2} \sum_i |\mathbf{P}_{ijk_2} - \sigma_i| + |\mathbf{P}_{ik_1j} - \sigma_i| \right] \\ &= 2\theta(\mathbf{P}, \sigma). \end{aligned}$$

In order to illustrate how the various conditions differ in practice, we consider two small example tensors borrowed from [27]:

$$(8.3) \quad \begin{aligned} \mathbf{P}_1 &= \left[\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1/3 & 0 & 0 \\ 1/3 & 0 & 1/2 \\ 1/3 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \right], \\ \mathbf{P}_2 &= \left[\begin{bmatrix} 0 & 0 & 1/3 \\ 0 & 0 & 1/3 \\ 1 & 1 & 1/3 \end{bmatrix} \begin{bmatrix} 1/3 & 0 & 0 \\ 1/3 & 0 & 0 \\ 1/3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 & 1/2 \\ 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} \right]. \end{aligned}$$

Figure 8.1 compares the range of values of α that guarantee uniqueness of the multilinear PageRank and convergence of the corresponding fixed point iteration for the two tensors \mathbf{P}_1 and \mathbf{P}_2 , according to the original Theorem 5.1 in [27], Corollary 7.2, Theorem 6.1, and Theorem 1 in [38]. For the latter result, we show the value of the quantities $\alpha \theta(\mathbf{P}, \sigma_k)$, $k = 1, 2, 3$, obtained with the three choices of vectors

$$(8.4) \quad (\sigma_1)_i = \max_{jk} \mathbf{P}_{ijk}, \quad (\sigma_2)_i = \min_{jk} \mathbf{P}_{ijk}, \quad \sigma_3 = \frac{\sigma_1 + \sigma_2}{2},$$

as proposed in Corollaries 1 and 2 in the same paper. The interesting ranges are those where the corresponding graphs stay below the dashed line.

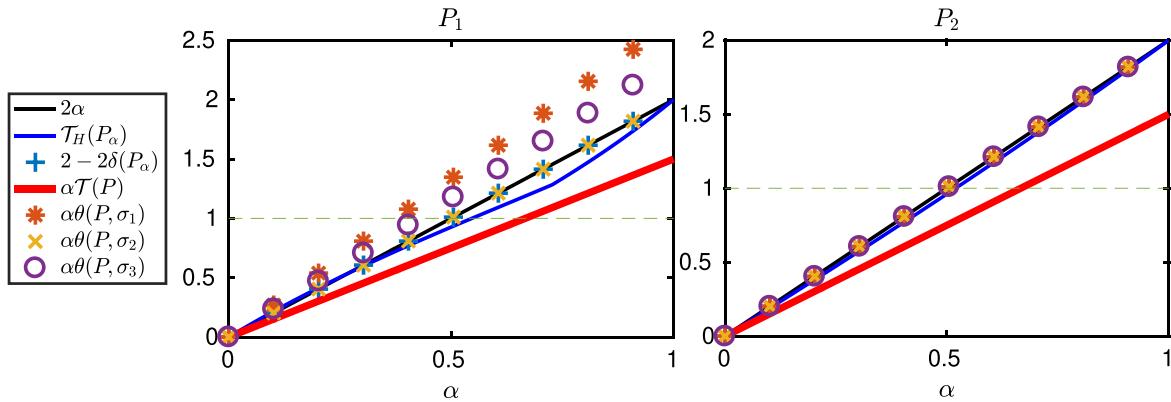


Figure 8.1. This figure compares the results in [27, Theorem 5.1], Corollary 7.2, Theorem 6.1, and [38, Corollaries 1 and 2] by comparing the values of 2α , $\mathcal{T}_H(\mathbf{P}_\alpha)$, $2 - 2\delta(\mathbf{P}_\alpha)$, $\alpha\mathcal{T}(\mathbf{P})$, $\alpha\theta(\mathbf{P}, \sigma_k)$, $k = 1, 2, 3$, where \mathbf{P}_α is defined as in (8.2), the vectors $\sigma_1, \sigma_2, \sigma_3$ defined as in (8.4), and \mathbf{P} is either of the example tensors \mathbf{P}_1 and \mathbf{P}_2 of (8.3).

8.2. Triangle-based PageRank on networks. Random walks are an important tool for exploratory network analysis. For example, they are at the basis of widely used methods for local clustering [2, 58], link prediction [6, 40], and network centrality [41, 46]. The typical random walk on a network is a Markov process where the probability to move from a node j to a node i is inversely proportional to the number of outgoing edges leaving from j . This classical first-order process only takes into account pairwise node-node relationships. However, recent work has highlighted that many important network features arise by exploiting the interaction of larger groups of nodes acting together; see, e.g., [4, 10, 44, 45].

In order to account for this type of second-order node interaction, we can consider a second-order stochastic process where the probability to move to a node i depends on the number of triangles that point towards i . We consider in this section a centrality index for networks that combines the standard and the triangle-based random walks, introduced in [4, 27] as a special case of the multilinear PageRank equation (8.1), and we show how the higher-order ergodicity coefficients for stochastic tensors help choose the parameters that enter into the model equation.

Let $G = (V, E)$ be an undirected graph, with $V = \{1, \dots, n\}$, and consider the tensor

$$\mathbf{T}_{ijk} = \begin{cases} \frac{1}{\Delta(j,k)} & \text{if } i, j, k \text{ form a triangle in } G, \\ 0 & \text{otherwise,} \end{cases}$$

where $\Delta(j, k)$ is the number of triangles that contain both nodes j and k . This tensor is the triangle-based version of the transition matrix of the standard random walk in G ,

$$A_{ij} = \begin{cases} \frac{1}{d(j)} & \text{if } i, j \text{ form an edge in } G, \\ 0 & \text{otherwise,} \end{cases}$$

where $d(i) = \sum_j A_{ij}$ is the degree of node i . Clearly, \mathbf{T} has many vanishing columns as in general two nodes $j, k \in V$ may not participate in any triangle in G . In order to make

\mathbf{T} stochastic we follow the standard practice of replacing null columns with the uniform probability vector; see, e.g., [27, 45]. That is, we modify the previous definition by setting $\mathbf{T}_{ijk} = 1/n$ for all $i = 1, \dots, n$ whenever $\Delta(j, k) = 0$. Similarly, we set $A_{ij} = 1/n$ for all $i = 1, \dots, n$ if j is an isolated node in G (i.e., if the j th column of A is zero). Now, define the tensor \mathbf{A} as $\mathbf{A}_{ijk} = A_{ij}$, and, for $\beta \in [0, 1]$, consider the stochastic tensor

$$(8.5) \quad \mathbf{P} = \beta\mathbf{T} + (1 - \beta)\mathbf{A}.$$

The next result specializes Theorem 6.1 to the multilinear PageRank problem associated with the tensor in (8.5) and, additionally, provides a bound on the distance between the solution to that problem and the standard PageRank vector.

Corollary 8.2. *Let \mathbf{P} be defined as in (8.5), and let $\gamma = \alpha(1 + \beta)$. If $\gamma < 1$, then (8.1) has a unique solution $x \in \mathcal{S}_1$, and the fixed point iteration $x_{t+1} = \alpha\mathbf{P}x_t + (1 - \alpha)v$ converges linearly to x with a convergence rate of at least γ .*

Moreover, let $z \in \mathcal{S}_1$ be the solution of the ordinary PageRank problem corresponding to the transition matrix A and the teleportation vector v ,

$$(8.6) \quad z = \alpha Az + (1 - \alpha)v.$$

Then,

$$\|x - z\|_1 \leq \frac{\alpha\beta}{1 - \alpha} \|\mathbf{T} - \mathbf{A}\|_1.$$

Proof. Observe that, as $\mathcal{T}(A) = \tau_1(A) \leq 1$, we have the trivial upper bound $\mathcal{T}(\mathbf{P}) \leq \beta + 1$; hence the first part of the claim follows from Corollary 8.1. Moreover, simple passages allow us to recast x as the stochastic solution of the equation $x = \mathbf{P}_{\alpha,\beta}xx$ where

$$\mathbf{P}_{\alpha,\beta} = \alpha\beta\mathbf{T} + \alpha(1 - \beta)\mathbf{A} + (1 - \alpha)\mathbf{V}$$

and \mathbf{V} is the rank-one tensor $\mathbf{V}_{ijk} = v_i$. Analogously, the vector z can also be considered as the stochastic solution of $z = (\alpha\mathbf{A} + (1 - \alpha)\mathbf{V})zz$, that is, $z = \mathbf{P}_{\alpha,0}zz$. We have $\mathcal{T}(\mathbf{P}_{\alpha,0}) \leq \alpha$; hence from Theorem 6.4 we get

$$\|x - z\|_1 \leq \frac{1}{1 - \alpha} \|\mathbf{P}_{\alpha,\beta} - \mathbf{P}_{\alpha,0}\|_1 = \frac{\alpha\beta}{1 - \alpha} \|\mathbf{T} - \mathbf{A}\|_1,$$

which completes the proof. ■

Together with Corollary 8.1, the previous result shows that small values of α and β produce a multilinear PageRank vector that does not differ sensibly from the ordinary PageRank vector. Note that the uniqueness and convergence condition $\gamma < 1$ in Corollary 8.2 can be fulfilled by any value $\alpha < 1/(1 + \beta)$. This condition is evidently less restrictive than the better known inequality $\alpha < 1/2$. This is one of several implications of Corollary 8.2. Below we consider an example real-world network to further showcase the advantages of that corollary.

The socfb-Carnegie49 network is a Facebook graph considered, for example, in the study of the social structure of Facebook users [59], and available online on Network Repository [53]. The graph has 6637 nodes and 249967 undirected edges. The largest connected component

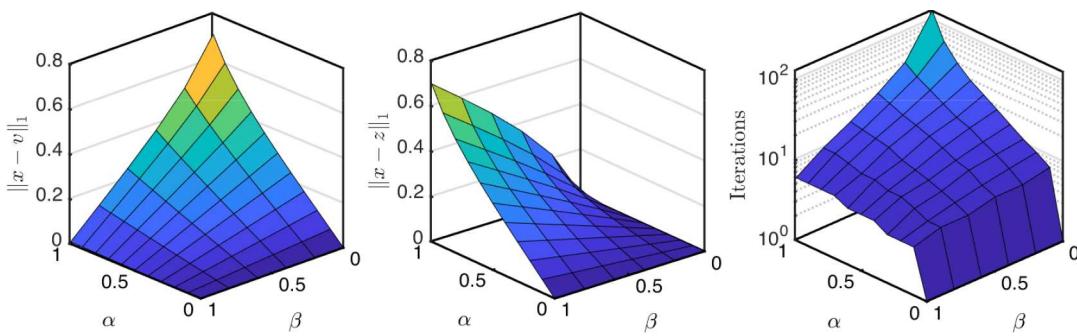


Figure 8.2. Triangle-based PageRank analysis on the socfb-Carnegie49 Facebook network, with varying α and β . Left to right: $\|x - v\|_1$, $\|x - z\|_1$, and number of iterations $x_{t+1} = \mathbf{P}_{\alpha,\beta}x_t x_t$ to reach $\|x_{t+1} - x_t\| < 10^{-8}$.

consists of 6621 vertices, and the triangle tensor \mathbf{T} has 13860318 nonzero entries. Figure 8.2 shows the results of a number of multilinear PageRank problems with coefficients α and β varying in $[0, 1]$. The equation $x = \alpha \mathbf{P}xx + (1 - \alpha)v$, with $\mathbf{P} = \beta\mathbf{T} + (1 - \beta)\mathbf{A}$ as in (8.5) and with uniform teleportation vector $v = \mathbb{1}/n$, has been solved via the fixed point iteration $x_{t+1} = \alpha \mathbf{P}x_t x_t + (1 - \alpha)v$ endowed with the stopping criterion $\|x_{t+1} - x_t\|_1 < 10^{-8}$.

The leftmost panel in Figure 8.2 shows the distance $\|x - v\|_1$, whereas the central panel shows $\|x - z\|_1$ where z is the usual PageRank vector, defined as in (8.6), with the same α value chosen for the multilinear version. The iteration number to convergence is shown in the rightmost panel.

While the overall behavior of $\|x - z\|_1$ reflects the estimate in Corollary 8.2, the panel on the left shows that x approaches v not only when $\alpha \approx 0$ but also when β is large. This is due to the fact that the triangle tensor \mathbf{T} has many zero columns. Consequently, the vast majority of the columns of \mathbf{P} coincide with the uniform vector v , and, for an arbitrary vector $x \in \mathcal{S}_1$, the product $\mathbf{P}xx$ is in general very close to v , in the sense that the 1-norm of the vector $r(x) = \mathbf{P}xx - v$ is rather small. With this notation we obtain

$$x - v = \alpha(1 - \beta)\{\mathbf{A}xx - v\} + \alpha\beta r(x) = \alpha(1 - \beta)\{Ax - v\} + \alpha\beta r(x).$$

Broadly speaking, when \mathbf{T} is very sparse $\|r(x)\|_1$ is usually negligible, and we can adopt the estimate $\|x - v\|_1 \approx \alpha(1 - \beta)\|Ax - v\|_1$. This approximation justifies the small error $\|x - v\|_1$ observed when $\beta \approx 1$.

In conclusion, this method yields most meaningful results when both α and β are neither too small nor too close to 1, as in those cases the computed vector x boils down to the standard PageRank or to the uniform vector. Extensive numerical experiments we performed on several real-world networks suggest the ‘‘reference’’ choice $\alpha, \beta \approx 0.6$. These values yield a good balance between first- and second-order information, fulfill the condition $\alpha(1 + \beta) < 1$ in Corollary 8.2, and ensure a fast convergence of the fixed point iteration. As an illustration, in Figure 8.3 we compare via scatter plots the reference solution for $\alpha = \beta = 0.6$ (on the x-axis) with other multilinear PageRank vectors (on the y-axis) for different values of α and β chosen as follows: in the leftmost panel the y-axis corresponds to the solution with $\alpha = 0.6$ and $\beta = 0$, i.e., the standard PageRank vector; in the central panel the vertical coordinates correspond to the purely triangle-based case $\alpha = 0.6$ and $\beta = 1$; in the rightmost panel the vector corresponding to $\alpha = \beta = 0.6$ is plotted against the solution for the three

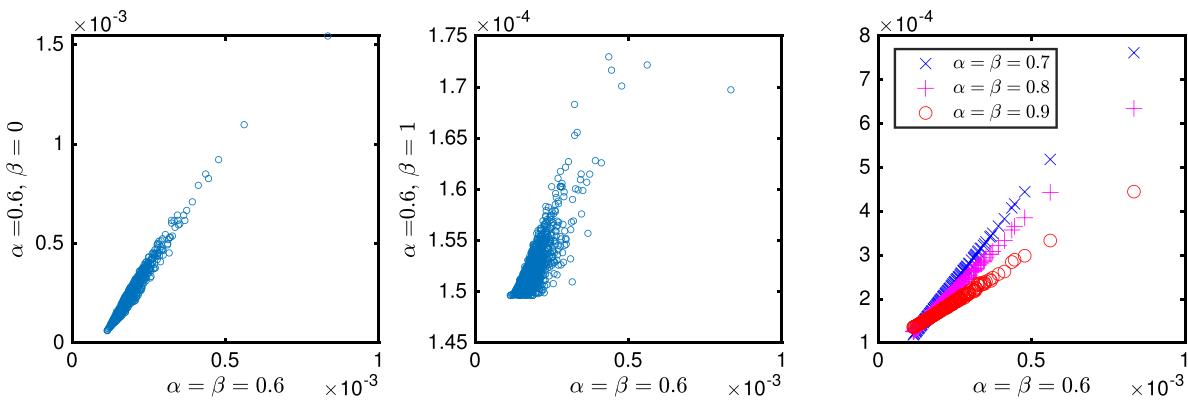


Figure 8.3. Scatter plots of triangle-based PageRank vectors of the socfb-Carnegie49 network for different choices of the parameters α and β . From left to right: comparisons between the solution with $\alpha = \beta = 0.6$ and the standard PageRank vector (left); the purely triangle-based solution $\alpha = 0.6$ and $\beta = 1$ (center); and other solutions with $\alpha = \beta$ (right).

choices $\alpha = \beta \in \{0.7, 0.8, 0.9\}$. The first two panels show the sensitivity of the solution with respect to different choices of β highlighting, in particular, the contribution of both edge- and triangle-based walks in the graph. The reference solution correlates with the classic PageRank, although the two vectors perceptibly differ, while it is significantly different than the purely triangle-based solution. The latter has a much smaller range and is more sensitive to changes in the graph topology, apart from being possibly nonunique. The last panel on the right, instead, shows that the reference solution for $\alpha = \beta = 0.6$ correlates almost perfectly with other numerical solutions obtained with larger values of the coefficients $\alpha = \beta$. Together with the timings shown in Figure 8.2, this illustrates that larger choices of the coefficients $\alpha = \beta$ result in essentially the same information on the nodes but require a much larger iteration count.

8.3. Higher-order shifted power method and lazy random walk. Let \mathbf{P} be symmetric, that is, $\mathbf{P} = \mathbf{P}^{\langle \pi \rangle}$ for every permutation π of $\{1, 2, 3\}$. In [33], Kolda and Mayo analyzed the convergence of the “shifted symmetric higher-order power method”

$$(8.7) \quad \hat{x}_{t+1} = \mathbf{P}x_t x_t + \alpha x_t, \quad x_{t+1} = \frac{\hat{x}_{t+1}}{\|\hat{x}_{t+1}\|_2}.$$

Their starting point is the optimization of the cubic form $f(x) = x^T(\mathbf{P}xx)$ over the sphere $x^T x = 1$, whose stationary points are, for symmetric tensors, Z-eigenvectors of \mathbf{P} related to the best symmetric rank-one approximation of \mathbf{P} . The coefficient α can be chosen positive or negative, in order to make the modified function $f(x) + \alpha x^T x$ convex or concave, respectively. Using fixed point theory, the authors of [33] prove that, given an appropriate shift α , the iterates in (8.7) generically converge to some Z-eigenvector. The shifting technique has been considered also for tensors that are not symmetric. For example, it has been considered in the framework of the multilinear PageRank [27] or in the case of ℓ^p -eigenvalue computation [26].

Let the coefficient $\beta(\mathbf{P})$ be defined as $\beta(\mathbf{P}) = 2 \max_{\|x\|_2=1} \rho(\mathbf{P}x)$, where $\rho(\mathbf{P}x)$ denotes the spectral radius of the matrix $\mathbf{P}x$. One of the main results from [33] is that, if \mathbf{P} is symmetric and $\alpha > \beta(\mathbf{P})$, then the method (8.7) converges to some stationary point of f , which is a Z-eigenvector of \mathbf{P} .

If \mathbf{P} is stochastic (but not necessarily symmetric), then it is natural to replace the sphere $x^T x = 1$ with the simplex \mathcal{S}_1 and the vector 2-norm with the 1-norm. With these replacements, and other minor notation changes, the iteration (8.7) boils down to

$$(8.8) \quad x_{t+1} = \sigma \mathbf{P} x_t x_t + (1 - \sigma) x_t, \quad \sigma \in (0, 1),$$

which, for an initial stochastic vector x_0 , will remain in \mathcal{S}_1 throughout. This iteration coincides with the higher-order power method $x_{t+1} = \mathbf{P}_\sigma x_t x_t$ for the “shifted tensor”

$$\mathbf{P}_\sigma = \sigma \mathbf{P} + (1 - \sigma) \mathbf{E},$$

where \mathbf{E} is any tensor such that $\mathbf{E}xx = x$ for all $x \in \mathcal{S}_1$. For example, \mathbf{E} can be chosen as a convex combination of the left and right identities \mathbf{E}^L and \mathbf{E}^R , defined in (2.2).

Note that \mathbf{P}_σ is stochastic for any choice of $\sigma \in (0, 1)$, and thus the iteration (8.8) can be interpreted as a form of higher-order lazy random walk. In fact, recall that if $P \in \mathbb{R}^{n \times n}$ is a stochastic matrix, then the Markov chain associated with $\sigma P + (1 - \sigma)I$ is called a lazy random walk, as it describes a walker that, with probability σ , performs a transition according to P and remains in its current state otherwise. Hence, we can use Theorem 6.1 to provide a condition on σ , in terms of the entries of \mathbf{P} , that guarantees global convergence of the shifted power method (8.8).

Even though the higher-order ergodicity coefficient $\mathcal{T}(\mathbf{P})$ of the original tensor may be larger than one, suitable values of σ can ensure that Theorem 6.1 holds for \mathbf{P}_σ . In fact, it is interesting to note that the function $\sigma \mapsto \mathcal{T}(\mathbf{P}_\sigma)$ is continuous, piecewise linear, and convex, with $\mathcal{T}(\mathbf{P}_0) = \mathcal{T}(\mathbf{E}) = 1$. As $x = \mathbf{P}_\sigma xx$ if and only if $x = \mathbf{P}xx$, we deduce the following.

Corollary 8.3. *If \mathbf{P} is stochastic and $\mathcal{T}(\mathbf{P}_\sigma) < 1$ for some $\sigma \geq 0$, then \mathbf{P} has a unique positive Z-eigenvector $x \in \mathcal{S}_1$ and the method (8.7) converges to x , for any starting point $x_0 \in \mathcal{S}_1$, with a convergence rate of at least $\mathcal{T}(\mathbf{P}_\sigma)$.*

Proof. The claim follows straightforwardly from Theorem 6.1 applied to \mathbf{P}_σ . ■

In Figure 8.4 we show the value of $\mathcal{T}(\mathbf{P}_\sigma)$ as a function of σ for the two example tensors (8.3) and for the choice $\mathbf{E} = \frac{1}{2}(\mathbf{E}^L + \mathbf{E}^R)$. Notice that for both the examples shown there exists an optimal σ_* such that $\min_\sigma \mathcal{T}(\mathbf{P}_\sigma) = \mathcal{T}(\mathbf{P}_{\sigma_*}) < 1$. Thus, although the higher-order

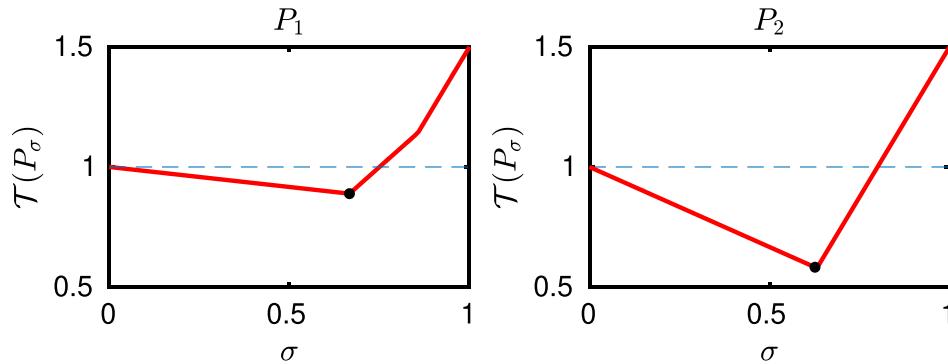


Figure 8.4. Variation of $\mathcal{T}(\mathbf{P}_\sigma)$ as σ varies within $[0, 1]$ for the two example tensors in (8.3).

ergodicity coefficient $\mathcal{T}(\mathbf{P})$ of the original tensor is larger than one, by Corollary 8.3 there exists a unique positive $x \in \mathcal{S}_1$ such that $x = \mathbf{P}xx$, and we can compute it with a method that converges as $\|x_{t+1} - x\|_1 \leq \mathcal{T}(\mathbf{P}_{\sigma^*})^t \|x_0 - x\|_1$ for an arbitrary $x_0 \in \mathcal{S}_1$.

9. Conclusions. This work adds to the long and continuing history of applications of tensor methods to data science by providing a novel analysis of the long-term behavior of higher-order stochastic processes governed by stochastic tensors. These types of processes are used in a large number of network science and data mining applications due to their ability to improve the underlying models and offer additional valuable insights. Even though stationary distributions of these processes are often required, fundamental mathematical questions such as the uniqueness of the distribution and the convergence behavior of the stochastic process remain unanswered. In fact, this is a relatively newly born and actively growing research area with many open questions.

Following a natural extension of the widely used ergodicity coefficients for Markov chains, we have introduced a new family of higher-order ergodicity coefficients for higher-order processes that provides new and easily computable conditions to ensure existence, uniqueness, and convergence towards the corresponding stationary distribution. The proposed analysis adds to previous work on uniqueness of Z -eigenvectors of stochastic tensors [14, 37, 39] and nonnegative tensors in general [16, 23, 25, 26] by providing new conditions that are either less restrictive or computationally easier to verify, or both.

Acknowledgments. The main results of this work were developed during a visiting period that F.T. spent at the Department of Mathematics, Computer Science and Physics of the University of Udine, Italy. He would like to thank the department and D.F. for the warm hospitality he received during that period.

REFERENCES

- [1] A. ANANDKUMAR, R. GE, D. HSU, S. M. KAKADE, AND M. TELGARSKY, *Tensor decompositions for learning latent variable models*, J. Mach. Learn. Res., 15 (2014), pp. 2773–2832.
- [2] R. ANDERSEN, F. CHUNG, AND K. LANG, *Local graph partitioning using PageRank vectors*, in Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS’06), IEEE, 2006, pp. 475–486, <https://doi.org/10.1109/FOCS.2006.44>.
- [3] F. ARRIGO, D. J. HIGHAM, AND V. NOFERINI, *Non-backtracking PageRank*, J. Sci. Comput., 80 (2019), pp. 1419–1437.
- [4] F. ARRIGO, D. J. HIGHAM, AND F. TUDISCO, *A framework for second-order eigenvector centralities and clustering coefficients*, Proc. A, 476 (2020), 20190724.
- [5] F. ARRIGO AND F. TUDISCO, *Multi-dimensional, multilayer, nonlinear and dynamic HITS*, in Proceedings of the 2019 SIAM International Conference on Data Mining, SIAM, 2019, pp. 369–377.
- [6] L. BACKSTROM AND J. LESKOVEC, *Supervised random walks: Predicting and recommending links in social networks*, in Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, 2011, pp. 635–644.
- [7] M. BENAÏM, *Vertex-reinforced random walks and a conjecture of Pemantle*, Ann. Probab., 25 (1997), pp. 361–392.
- [8] A. BENSON, D. F. GLEICH, AND L.-H. LIM, *The spacey random walk: A stochastic process for higher-order data*, SIAM Rev., 59 (2017), pp. 321–345.
- [9] A. R. BENSON, *Three hypergraph eigenvector centralities*, SIAM J. Math. Data Sci., 1 (2019), pp. 293–312.

- [10] A. R. BENSON, D. F. GLEICH, AND J. LESKOVEC, *Tensor spectral clustering for partitioning higher-order network structures*, Proceedings of the 2015 SIAM International Conference on Data Mining, 2015, pp. 118–126.
- [11] A. BERCHTOLD AND A. RAFTERY, *The mixture transition distribution model for high-order Markov chains and non-Gaussian time series*, Statist. Sci., 17 (2002), pp. 328–356.
- [12] G. BIRKHOFF, *Extensions of Jentzsch’s theorem*, Trans. Amer. Math. Soc., 85 (1957), pp. 219–227.
- [13] F. BOUGUET AND B. CLOEZ, *Fluctuations of the empirical measure of freezing Markov chains*, Electron. J. Probab., 23 (2018), 2.
- [14] H. BOZORGMANESH AND M. HAJARIAN, *Convergence of a transition probability tensor of a higher-order Markov chain to the stationary probability vector*, Numer. Linear Algebra Appl., 23 (2016), pp. 972–988.
- [15] K. CHANG AND T. ZHANG, *On the uniqueness and non-uniqueness of the positive Z-eigenvector for transition probability tensors*, J. Math. Anal. Appl., 408 (2013), pp. 525–540.
- [16] K.-C. CHANG, K. PEARSON, AND T. ZHANG, *Perron–Frobenius theorem for nonnegative tensors*, Commun. Math. Sci., 6 (2008), pp. 507–520.
- [17] F. CHIERICETTI, R. KUMAR, P. RAGHAVAN, AND T. SARLOS, *Are web users really Markovian?*, in Proceedings of the 21st International Conference on World Wide Web, ACM, 2012, pp. 609–618.
- [18] S. CIPOLLA, M. REDIVO-ZAGLIA, AND F. TUDISCO, *Extrapolation methods for fixed-point multilinear PageRank computations*, Numer. Linear Algebra Appl., 27 (2020), e2280.
- [19] L.-B. CUI AND Y. SONG, *On the uniqueness of the positive Z-eigenvector for nonnegative tensors*, J. Comput. Appl. Math., 352 (2019), pp. 72–78.
- [20] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *On the best rank-1 and rank- (R_1, R_2, \dots, R_N) approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342.
- [21] R. L. DOBRUSHIN, *Central limit theorem for nonstationary Markov chains. I, II*, Theory Probab. Appl., 1 (1956), pp. 65–80, 329–383.
- [22] S. P. EVESON AND R. D. NUSSBAUM, *An elementary proof of the Birkhoff–Hopf theorem*, Math. Proc. Cambridge Philos. Soc., 117 (1995), pp. 31–55.
- [23] S. FRIEDLAND, S. GAUBERT, AND L. HAN, *Perron–Frobenius theorem for nonnegative multilinear forms and extensions*, Linear Algebra Appl., 438 (2013), pp. 738–749.
- [24] A. GAUTIER AND F. TUDISCO, *The contractivity of cone-preserving multilinear mappings*, Nonlinearity, 32 (2019), pp. 4713–4728.
- [25] A. GAUTIER, F. TUDISCO, AND M. HEIN, *The Perron–Frobenius theorem for multihomogeneous mappings*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 1179–1205.
- [26] A. GAUTIER, F. TUDISCO, AND M. HEIN, *A unifying Perron–Frobenius theorem for nonnegative tensors via multihomogeneous maps*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 1206–1231.
- [27] D. F. GLEICH, L.-H. LIM, AND Y. YU, *Multilinear PageRank*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1507–1541.
- [28] C. J. HILLAR AND L.-H. LIM, *Most tensor problems are NP-hard*, J. ACM, 60 (2013), 45.
- [29] S. HU AND L. QI, *Convergence of a second order Markov chain*, Appl. Math. Comput., 241 (2014), pp. 183–192.
- [30] S. HU, L. QI, AND G. ZHANG, *Computing the geometric measure of entanglement of multipartite pure states by means of non-negative tensors*, Phys. Rev. A, 93 (2016), 012304.
- [31] I. C. F. IPSEN AND T. M. SELEE, *Ergodicity coefficients defined by vector norms*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 153–200.
- [32] K. KNOPP, *Infinite Sequences and Series*, Dover Publications, Inc., New York, 1956.
- [33] T. G. KOLDA AND J. R. MAYO, *Shifted power method for computing tensor eigenpairs*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 1095–1124.
- [34] V. N. KOLOKOLTSEV, *Nonlinear Markov Processes and Kinetic Equations*, Cambridge University Press, Cambridge, UK, 2010.
- [35] F. KRZAKALA, C. MOORE, E. MOSEL, J. NEEMAN, A. SLY, L. ZDEBOROVÁ, AND P. ZHANG, *Spectral redemption in clustering sparse networks*, Proc. Natl. Acad. Sci. USA, 110 (2013), pp. 20935–20940.
- [36] C.-K. LI AND S. ZHANG, *Stationary probability vectors of higher-order Markov chains*, Linear Algebra Appl., 473 (2015), pp. 114–125.

- [37] W. LI, L.-B. CUI, AND M. K. NG, *The perturbation bound for the Perron vector of a transition probability tensor*, Numer. Linear Algebra Appl., 20 (2013), pp. 985–1000.
- [38] W. LI, D. LIU, M. K. NG, AND S.-W. VONG, *The uniqueness of multilinear PageRank vectors*, Numer. Linear Algebra Appl., 24 (2017), e2107.
- [39] W. LI AND M. K. NG, *On the limiting probability distribution of a transition probability tensor*, Linear Multilinear Algebra, 62 (2014), pp. 362–385.
- [40] Y. LIU, H. TONG, L. XIE, AND Y. TANG, *Supervised link prediction using random walks*, in Communications in Computer and Information Science, Commun. Comput. Inf. Sci. 568, Springer-Verlag, Cham, 2015, pp. 107–118, https://doi.org/10.1007/978-981-10-0080-5_10.
- [41] N. MASUDA, M. A. PORTER, AND R. LAMBIOTTE, *Random walks and diffusion on networks*, Phys. Rep., 716/717 (2017), pp. 1–58.
- [42] Q. MEI, J. GUO, AND D. RADEV, *Divrank: The interplay of prestige and diversity in information networks*, in Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2010, pp. 1009–1018.
- [43] B. MEINI AND F. POLONI, *Perron-based algorithms for the multilinear PageRank*, Numer. Linear Algebra Appl., 25 (2018), e2177.
- [44] H. NASSAR, A. R. BENSON, AND D. F. GLEICH, *Pairwise Link Prediction*, preprint, arXiv:1907.04503, 2019.
- [45] M. K. NG, X. LI, AND Y. YE, *Multirank: Co-ranking for objects and relations in multi-relational data*, in Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2011, pp. 1217–1225.
- [46] J. D. NOH AND H. RIEGER, *Random walks on complex networks*, Phys. Rev. Lett., 92 (2004), 118701, <https://doi.org/10.1103/PhysRevLett.92.118701>.
- [47] R. PEMANTLE, *Vertex-reinforced random walk*, Probab. Theory Related Fields, 92 (1992), pp. 117–136.
- [48] L. QI AND Z. LUO, *Tensor Analysis: Spectral Theory and Special Tensors*, SIAM, Philadelphia, PA, 2017.
- [49] L. QI, Y. WANG, AND E. X. WU, *D-eigenvalues of diffusion kurtosis tensors*, J. Comput. Appl. Math., 221 (2008), pp. 150–157.
- [50] A. E. RAFTERY, *A model for high-order Markov chains*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 47 (1985), pp. 528–539.
- [51] A. E. RAFTERY AND S. TAVARÉ, *Estimation and modelling repeated patterns in high order Markov chains with the mixture transition distribution model*, J. R. Stat. Soc. Ser. C. Appl. Stat., 43 (1994), pp. 179–199.
- [52] S. RAGNARSSON AND C. F. VAN LOAN, *Block tensor unfoldings*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 149–169.
- [53] R. A. ROSSI AND N. K. AHMED, *The network data repository with interactive graph analytics and visualization*, in Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15, AAAI Press, 2015, p. 4292–4293, <http://networkrepository.com> (accessed 2020-04-26).
- [54] M. ROSVALL, A. V. ESQUIVEL, A. LANCICHINETTI, J. D. WEST, AND R. LAMBIOTTE, *Memory in network flows and its effects on spreading dynamics and community detection*, Nature Commun., 5 (2014), 4630.
- [55] M. SABUROV, *Ergodicity of p -majorizing nonlinear Markov operators on the finite dimensional space*, Linear Algebra Appl., 578 (2019), pp. 53–74.
- [56] E. SENETA, *Non-negative matrices and Markov chains*, Springer-Verlag, Cham, 1981.
- [57] E. SENETA, *Perturbation of the stationary distribution measured by ergodicity coefficient*, Adv. Appl. Probab., 20 (1988), pp. 228–230.
- [58] D. A. SPIELMAN AND S.-H. TENG, *A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning*, SIAM J. Comput., 42 (2013), pp. 1–26.
- [59] A. L. TRAUD, P. J. MUCHA, AND M. A. PORTER, *Social structure of Facebook networks*, Phys. A, 391 (2012), pp. 4165–4180.
- [60] F. TUDISCO, *A note on certain ergodicity coefficients*, Spec. Matrices, 3 (2015), pp. 175–185.
- [61] O. E. WILLIAMS, F. LILLO, AND V. LATOURA, *Effects of memory on spreading processes in non-Markovian temporal networks*, New J. Phys., 21 (2019), 043028.
- [62] S.-J. WU AND M. T. CHU, *Markov chains with memory, tensor formulation, and the dynamics of power iteration*, Appl. Math. Comput., 303 (2017), pp. 226–239.