ERGODICITY COEFFICIENTS DEFINED BY VECTOR NORMS

ILSE C.F. IPSEN* AND TERESA M. SELEE†

Abstract. Ergodicity coefficients for stochastic matrices determine inclusion regions for subdominant eigenvalues; estimate the sensitivity of the stationary distribution to changes in the matrix; and bound the convergence rate of methods for computing the stationary distribution.

We survey results for ergodicity coefficients that are defined by p-norms, for stochastic matrices as well as for general real or complex matrices. We express ergodicity coefficients in the one-, two-, and infinity-norms as norms of projected matrices, and we bound coefficients in any p-norm by norms of deflated matrices. We show that two-norm ergodicity coefficients of a matrix A are closely related to the singular values of A. In particular, the singular values determine the extreme values of the coefficients. We show that ergodicity coefficients can determine inclusion regions for subdominant eigenvalues of complex matrices, and that the tightness of these regions depends on the departure of the matrix from normality. In the special case of normal matrices, two-norm ergodicity coefficients turn out to be Lehmann bounds.

Key words. singular values, eigenvalues, stochastic matrices, nonnegative matrices, inclusion regions, projections

AMS subject classification. 15-02, 15A18, 15A42, 15A48, 15A51, 65F15, 65F35

1. Introduction.

Ergodicity, in its most general form, has to do with the long term behavior of dynamical systems. Here we concentrate on particular systems, namely finite inhomogeneous Markov chains, and try to understand measures of ergodicity from the point of view of linear algebra.

In the context of inhomogeneous Markov chains, ergodicity refers to the asymptotic behavior of products of stochastic matrices[1] where the number of factors grows unbounded. Very informally, a Markov chain is *ergodic* if the matrix products converge to a rank one matrix, that is, a stochastic matrix all of whose rows are equal. So-called coefficients of ergodicity were introduced to estimate how fast, if at all, these products converge to a matrix of rank one.

In the simplest case, all factors in the products are identical to the same stochastic matrix S. Order the eigenvalues $\lambda_i(S)$ in order of decreasing magnitude, $1 = \lambda_1(S) \geq |\lambda_2(S)| \geq \ldots$. If the subdominant eigenvalue is strictly smaller in magnitude than the dominant eigenvalue, i.e. $|\lambda_2(S)| < 1$, then $|\lambda_2(S^k)| = |\lambda_2(S)|^k \to 0$ as $k \to \infty$. This means, the powers S^k converge to a stochastic matrix of rank one, and the magnitude of the subdominant eigenvalue, $|\lambda_2(S)|$, estimates the asymptotic rate of convergence. In this situation $|\lambda_2(S)|$ could serve as a coefficient of ergodicity, see [24, 60].

Suppose now the products consist of different stochastic matrices S_j whose number is increasing and we would like to know at which rate, if at all, the products $S_1 \cdots S_j$ converge to a rank-one matrix as $j \to \infty$. The second eigenvalue is of no use here, since in general $\lambda_2(S_1 \cdots S_j) \neq \lambda_2(S_1) \cdots \lambda_2(S_j)$. We need a substitute for $|\lambda_2|$, with some kind of multiplicative property, and the ability to recognize when a matrix has rank one.

*Department of Mathematics, North Carolina State University, P.O. Box 8205, Raleigh, NC 27695-8205, USA (`ipsen@ncsu.edu`, `http://www4.ncsu.edu/~ipsen/`)

†Georgia Tech Resarch Institute, Sensors and Electromagnetic Applications Laboratory, 7220 Richardson Road, Smyrna, GA 30060, USA (`Teresa.Selee@gtri.gatech.edu`)

[1] A (row) stochastic matrix is a square matrix with nonnegative elements that sum to one in each row. A product of stochastic matrices is again stochastic.

An example of such a substitute is the one-norm coefficient of ergodicity,

$$\tau_1(S) = \max_{\|z\|_1=1,\, z^T\mathbb{1}=0} \|S^T z\|_1,$$

where the maximum ranges over real vectors z, the superscript T denotes the transpose, and $\mathbb{1}$ is the column vector of all ones. The coefficient $\tau_1(S)$ is simply the norm of the matrix S restricted to the subspace that is orthogonal to $\mathbb{1}$. As we will show in §3, the coefficient $\tau_1(S)$ bounds the non-unit eigenvalues of the stochastic matrix S: $|\lambda| \leq \tau_1(S)$ for any eigenvalue $\lambda \neq 1$ of S; and it is submultiplicative: $\tau_1(S_1S_2) \leq \tau_1(S_1)\tau_1(S_2)$ for any two stochastic matrices S_1 and S_2.

1.1. Motivation. We became interested in coefficients of ergodicity in the context of work on the Google matrix [35, 37, 36, 78].

The Google matrix is a convex combination of a stochastic matrix S and a rank-one stochastic matrix, $G \equiv \alpha S + (1-\alpha)\mathbb{1}v^T$, where v is some nonnegative column vector whose elements sum to one, and $0 \leq \alpha < 1$. Various algorithms have been proposed to compute the stationary distribution of G, that is a column vector $\pi \neq 0$ with $\pi^T G = \pi^T$. In [35] we analyzed a so-called aggregation-disaggregation algorithm and showed that its asymptotic convergence rate is bounded by the ergodicity coefficient τ_1 of the aggregated matrix.

Alternatively, the stationary distribution π can be computed by applying the power method to G. The power method has an asymptotic convergence rate of $|\lambda_2(G)|$, where $\lambda_i(G)$ are the eigenvalues of G labeled in descending order, $1 = \lambda_1(G) \geq |\lambda_2(G)| \geq \ldots$. A derivation "from scratch" [22, 30, 75] shows that $|\lambda_2(G)| \leq \alpha$, but it also follows follows immediately from $|\lambda_2(G)| \leq \tau_1(G) = \alpha\tau_1(S) \leq \alpha$.

Note that the vector $\mathbb{1}$ in the expression for G is a dominant eigenvector of S and also of G, since both matrices are stochastic. Hence the rank-one matrix $\mathbb{1}v^T$ is almost a spectral projector, but not quite. This helped us to realize that τ_1 implicitly deflates a matrix by removing the dominant spectral projector through the constraint $z^T\mathbb{1} = 0$.

The Google matrix form has been extended to general complex matrices [32]. Let A be a complex square matrix with dominant eigenvalue λ and right eigenvector w, i.e. $Aw = \lambda w$. Set $H \equiv \gamma A + (1-\gamma)wx^*$, where the superscript $*$ denotes the conjugate transpose, γ is a complex scalar, and x a complex column vector with $x^*w = 1$. Then one can show [32, Theorem 1.2] that $\lambda_2(H) = \gamma\lambda_2(A)$. With a more general ergodicity coefficient $\tau_p(w, A)$ that maximizes over vectors orthogonal to w, we obtain readily that $|\lambda_2(H)| \leq |\gamma|\,\tau_p(w, A)$. Again, as for the Google matrix above, the rank-one matrix wx^* approximates a spectral projector.

The connection of ergodicity coefficients to deflated (or downdated) matrices sparked our interest. When we started looking at the literature on ergodicity coefficients, we found many scattered results; it was not always clear how they were related; and the notation was at times inconsistent and not always transparent. That's why we decided to write this paper.

1.2. Overview. We survey coefficients of ergodicity that are defined by vector norms, from the vantage point of numerical linear algebra. We try to present a coherent discussion of existing results, with simplified and complete proofs. We argue that ergodicity coefficients can be viewed as norms of deflated matrices. For two-norm coefficients we present new explicit expressions and establish connections to singular values and eigenvalue bounds.

We restrict our attention to ergodicity coefficients of finite dimensional matrices. Ergodicity coefficients for stochastic matrices of infinite dimension have been studied by, among others, Isaacson and Madsen [38], Pax [50], Paz and Reichaw [52], and Rhodius [59].

We could have started this survey with ergodicity coefficients in their most general form and derived the results for stochastic matrices as corollaries. Instead, we decided to follow the historical development a bit, which began with coefficients for stochastic matrices: in the one norm (§3), infinity norm (§4), and any p-norm (§5, §6). Subsequently we discuss extensions to real matrices (§7), and to complex matrices with maximization over arbitrary subspaces (§8). We illustrate applications of ergodicity coefficients to: estimating the sensitivity of stationary distributions (§3.3), and determining inclusion regions for eigenvalues of nonnegative matrices (§7.5), general complex matrices (§8.1), and normal matrices (§8.4).

1.3. New Results. We present a much simpler proof for the explicit form of the one-norm ergodicity coefficient (second proof of Theorem 3.4).

We represent ergodicity coefficients as norms of (obliquely) projected matrices, for the one-norm (Corollary 3.5), and for the infinity-norm (Corollaries 4.3 and 6.6). For the two-norm, we derive explicit expressions in terms of orthogonal projections of the matrix (Corollary 6.8, Theorems 7.12, 7.15, 8.6 and 8.7). We show that general ergodicity coefficients in any p-norm can be bounded by the norm of a deflated matrix (Theorem 8.2).

We show that two-norm coefficients can reproduce any singular value (Corollaries 7.16 and 8.8), and that their extreme values are determined by singular values (Theorem 8.9). We apply ergodicity coefficients to determine inclusion regions for subdominant eigenvalues of general complex matrices (Theorems 8.5 and 8.11) and show that the tightness of the inclusion regions depends on the departure of the matrix from normality. In the special case of normal matrices, the two-norm ergodicity coefficients turn out to be Lehmann bounds (Theorem 8.13).

Notation. The identity matrix is I with columns e_i. The column vector of all ones is $\mathbb{1}$. The transpose of a matrix A is A^T, and the conjugate transpose is A^*.

Let x be a $n \times 1$ column vector with elements x_i, $1 \leq i \leq n$. The one-norm of x is $\|x\|_1 = \sum_{1 \leq i \leq n} |x_i|$, and the infinity-norm is $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$. The matrix $\operatorname{diag}(x)$ denotes a $n \times n$ diagonal matrix whose i diagonal element equals x_i, $1 \leq i \leq n$. The componentwise inequality $x \geq 0$ $(x > 0)$ means that all elements of x are nonnegative (positive), while $|x| > 0$ means that all elements of x are nonzero. The vector x is *stochastic* if $x \geq 0$ and $x^T \mathbb{1} = 1$. By $x_{1:k}$ we mean the $k \times 1$ vector consisting of elements $x_1, \ldots, x_k$.

The elements of a $n \times n$ matrix S are denoted by s_{ij}, $1 \leq i, j \leq n$. The column space of a $n \times n$ matrix A is $\operatorname{range}(A)$, and its orthogonal complement in $\mathbb{R}^n$ or $\mathbb{C}^n$, i.e. the left null space of A, is denoted by $\operatorname{range}(A)^\perp$.

2. General Ergodicity Coefficients for Stochastic Matrices.

We present a formal definition of ergodicity, and we introduce two very general classes of ergodicity coefficients.

2.1. Weak Ergodicity.

Ergodicity, in general, refers to the long term behavior of dynamical systems. In the context of finite, inhomogeneous Markov chains, ergodicity describes the long-term behavior of products of stochastic matrices where the number of factors is increasing. Seneta attributes the following definition of weak ergodicity to a 1931 paper by Kolmogorov.

DEFINITION 2.1 (§4 in [42], §1 in [65]). *Let $\{S_k\}$ be a sequence of $n \times n$ stochastic matrices, $k \geq 1$, and let $t_{ij}^{(p,r)}$ be the (i,j) entry of the forward product $T^{(p,r)} = S_{p+1} S_{p+2} \cdots S_{p+r}$. The sequence $\{S_k\}$ is called* weakly ergodic *if for all $1 \leq i, j, k \leq n$ and $p \geq 0$,*

$$t_{ik}^{(p,r)} - t_{jk}^{(p,r)} \to 0 \qquad \text{as} \quad r \to \infty.$$

This means, a sequence of stochastic matrices is weakly ergodic if the rows of the products tend to equalize, as the number of factors in the product increases.

Conditions for weak ergodicity appear in numerous sources. Among the earliest works we found were papers by Bernštein [8, 9, 10, 11, 12, 13], Doeblin [20], Dobrushin [18], Dynkin [21], Sapogov [61, 62] and Sarymsakov [63]; these papers appeared between 1920 and 1950. More recent papers by Cohn [16, 17], Dobrushin [19], Hajnal [25, 26], Kingman [40], and Paz and Reichaw [52] are written in English and are more easily accessible. Excellent summaries can be found in papers by Seneta [65, §1], [66, §1], and in his book [67, Chapter 3-4].

2.2. First Class of Ergodicity Coefficients.

This class of ergodicity coefficient is defined specifically for stochastic matrices.

DEFINITION 2.2 (p 509 in [65]). *A* coefficient of ergodicity*, or* ergodicity coefficient*, is a continuous scalar function $\mu(\cdot)$ defined for stochastic matrices S that* *satisfies* $0 \le \mu(S) \le 1$*. A coefficient of coefficient is* proper *if*

$$\mu(S) = 0 \iff S = \mathbb{1}v^T,$$

where v *is a stochastic vector.*

A proper coefficient of ergodicity is equal to zero if all rows of the stochastic matrix are identical. This, in turn, is the case if and only if the rank of the stochastic matrix equals one. Sometimes one finds an alternative definition, where the ergodicity coefficient is defined instead as $\hat{\mu}(S) \equiv 1 - \mu(S)$ and is called proper if: $\hat{\mu}(S) = 1 \iff \operatorname{rank}(S) = 1$ [14], [26, §2], [34, p 56], [40, §4], [65, §2].

For the particular case of doubly stochastic[2] matrices, a proper ergodicity coefficient is zero for both the matrix and its transpose at the same time. This is because the rank of a matrix is equal to the rank of its transpose. The corresponding statement below was shown for $\tau_1(\cdot)$ in [50, Property 1].

THEOREM 2.3. *If* $S_d \in \mathbb{R}^{n\times n}$ *is a doubly stochastic matrix and* $\mu(\cdot)$ *a proper coefficient of ergodicity then*

$$\mu(S_d) = \mu\left(S_d^T\right) = 0 \iff \operatorname{rank}(S_d) = 1.$$

Proof. According to Definition 2.2, $\mu(S_d) = 0 \iff \operatorname{rank}(S_d) = 1$. Since S_d^T is also stochastic, $\mu(S_d^T) = 0 \iff \operatorname{rank}(S_d^T) = 1$. The desired statement follows from the fact that $\operatorname{rank}(S_d) = \operatorname{rank}(S_d^T)$. □

Although a doubly stochastic matrix of rank one equals $S_d = \frac{1}{n}\mathbb{1}\mathbb{1}^T$ and is symmetric, doubly stochastic matrices of larger rank are in general not symmetric.

Definition 2.2 allows us to express the condition for weak ergodicity in terms of proper ergodicity coefficients.

THEOREM 2.4 (p 136 in [67]). *Let* $\{S_k\}$ *be a sequence of* $n\times n$ *stochastic matrices,* $k \ge 1$*, and* $T^{(p,r)} = S_{p+1}S_{p+2}\cdots S_{p+r}$*. The sequence* $\{S_k\}$ *is weakly ergodic if for* $p \ge 0$

$$\mu\left(T^{(p,r)}\right) \to 0 \qquad \text{as} \quad r \to \infty,$$

where $\mu(\cdot)$ *is a proper coefficient of ergodicity.*

EXAMPLE 2.5. *Let* S *be a stochastic matrix. The following are proper ergodicity coefficients [67, p 137]*

$$\begin{aligned}
\tau_1(S) &= \frac{1}{2}\max_{ij}\sum_k |s_{ik} - s_{jk}| = 1 - \min_{ij}\sum_k \min\{s_{ik}, s_{jk}\}\\
\alpha(S) &= \max_k \ \max_{ij} |s_{ik} - s_{jk}|\\
\beta(S) &= 1 - \sum_k \min_i s_{ik}.
\end{aligned}$$

The equality of the two expressions for $\tau_1(S)$ *will be proved in Corollary 3.6.*

The following ergodicity coefficient is not proper [67, p 137],

$$\gamma(S) \equiv 1 - \max_k \ \min_i s_{ik}.$$

[2] A stochastic matrix S_d is doubly stochastic if S_d^T is also stochastic.

This is because the $n \times n$ stochastic matrix $S = \frac{1}{n}\mathbb{1}\mathbb{1}^T$ has $\operatorname{rank}(S) = 1$ *but* $\gamma(S) = 1 - \frac{1}{n} \neq 0$.

The ergodicity coefficients in Example 2.5 can be related to each other in a number of ways [50], [65, §1]. Below is one of the simpler relations.

THEOREM 2.6 (p 56 in [34], pp 137–138 in [67]). *If S is a stochastic matrix then the ergodicity coefficients in Example 2.5 satisfy*

$$\alpha(S) \leq \tau_1(S) \leq \beta(S) \leq \gamma(S).$$

Proof. We start with the bound $\alpha(S) \leq \tau_1(S)$. Choose indices l, m and r so that $\alpha(S) = s_{mr} - s_{lr} \geq 0$. Then

$$\tau_1(S) = \frac{1}{2} \max_{ij} \sum_k |s_{ik} - s_{jk}| \geq \frac{1}{2} \sum_k |s_{mk} - s_{lk}|.$$

We remove the absolute values by breaking up the sum as follows. Let $\mathcal{P}_m$ be the set of all indices k with $s_{mk} \geq s_{lk}$, and $\mathcal{P}_l$ the set of all indices k with $s_{mk} < s_{lk}$. Then

$$\sum_k |s_{mk} - s_{lk}| = \sum_{k \in \mathcal{P}_m} (s_{mk} - s_{lk}) + \sum_{k \in \mathcal{P}_l} (s_{lk} - s_{mk}).$$

Since the elements in each row of S sum to one,

$$\sum_{k \in \mathcal{P}_l} s_{mk} = 1 - \sum_{k \in \mathcal{P}_m} s_{mk}, \qquad \sum_{k \in \mathcal{P}_l} s_{lk} = 1 - \sum_{k \in \mathcal{P}_m} s_{lk}.$$

Applying these two equalities to $\sum_{k \in \mathcal{P}_l} (s_{lk} - s_{mk})$ gives

$$\sum_{k \in \mathcal{P}_l} (s_{lk} - s_{mk}) = \sum_{k \in \mathcal{P}_m} (s_{mk} - s_{lk}),$$

hence

$$\tau_1(S) \geq \frac{1}{2} \sum_k |s_{mk} - s_{lk}| = \sum_{k \in \mathcal{P}_m} (s_{mk} - s_{lk}).$$

We have expressed the sum in terms of $\mathcal{P}_m$, because $\alpha(S) = s_{mr} - s_{lr} \geq 0$ implies that the index r must be in $\mathcal{P}_m$. Extracting the rth term from the sum gives

$$\sum_{k \in \mathcal{P}_m} (s_{mk} - s_{lk}) = (s_{mr} - s_{lr}) + \sum_{\substack{k \in \mathcal{P}_m \\ k \neq r}} (s_{mk} - s_{lk}) = \alpha(S) + \sum_{\substack{k \in \mathcal{P}_m \\ k \neq r}} (s_{mk} - s_{lk}).$$

The indices $k \in \mathcal{P}_m$ are those for which $s_{mk} \geq s_{lk}$, hence $\sum_{\substack{k \in \mathcal{P}_m \\ k \neq r}} (s_{mk} - s_{lk}) \geq 0$, and $\tau_1(S) \geq \alpha(S)$.

To prove $\tau_1(S) \leq \beta(S)$, let i_0 and j_0 be indices that achieve the minimum in the second expression for τ_1 in Example 2.5. From $\min\{s_{i_0,k}, s_{j_0,k}\} \leq \min_i s_{ik}$ follows

$$1 - \tau_1(S) = \sum_k \min\{s_{i_0,k}, s_{j_0,k}\} \geq \sum_k \min_i s_{i,k} = 1 - \beta(S).$$

At last, the bound $\beta(S) \leq \gamma(S)$ follows from $\max_k \min_i s_{ik} \leq \sum_k \min_i s_{ik}$. □

2.3. Second Class of Ergodicity Coefficients. This second class of coefficients gives rise to many of the popular ergodicity coefficients.

DEFINITION 2.7 (§2 in [66]). *Let $S \in \mathbb{R}^{n\times n}$ be a stochastic matrix, and let d be a metric defined on the set of stochastic vectors $\mathcal{D} = \{x : x \in \mathbb{R}^n,\ x \geq 0,\ x^T \mathbb{1} = 1\}$. The quantity*

$$\tau(S) = \sup_{\substack{x,y\in D\\ x\neq y}} \frac{d\left(x^T S, y^T S\right)}{d\left(x^T, y^T\right)}$$

is a coefficient of ergodicity.

For instance, $\sup_{x,y\in\mathcal{D},x\neq y} \|x^T S - y^T S\|_1/\|x^T - y^T\|_1$ is a coefficient of ergodicity in the sense of Definition 2.7, as well as Definition 2.2, because it is continuous, takes on values in $[0,1]$, and is generated by a metric, see §3 for more details.

However, there are situations where Definitions 2.2 and 2.7 are not consistent [44]. The coefficient $\mu(S) = s_{11}$ satisfies Definition 2.2 because it is continuous and $0 \leq \mu(S) \leq 1$; but $\mu(S)$ does not satisfy Definition 2.7 because it cannot be generated by a metric on $\mathcal{D}$. The coefficient $\sup_{x,y\in\mathcal{D},x\neq y} \|x^T S - y^T S\|_\infty/\|x^T - y^T\|_\infty$ satisfies Definition 2.7, but it does not satisfy Definition 2.2 because it can take on values greater than 1, see Section 4.

There are many ways to choose a metric d [44, 56, 57, 60, 66]. Two popular choices are presented below.

Birkhoff's Contraction Coefficient. The projective distance between between two vectors $x > 0$ and $y > 0$ is

$$d_B(x,y) = \ln\left(\frac{\max_i \frac{x_i}{y_i}}{\min_i \frac{x_i}{y_i}}\right) = \max_{ij} \ln\left(\frac{x_i y_j}{x_j y_i}\right).$$

The corresponding ergodicity coefficient

$$\tau_B(S) = \sup \frac{d_B\left(x^T S, y^T S\right)}{d_B\left(x^T, y^T\right)},$$

whose supremum ranges over all $x > 0$ and $y > 0$ that are not multiples of each other, is called *Birkhoff's Contraction Coefficient* [67, §3.1, 3.4]. It can actually be defined for the larger class of row-allowable matrices, which are nonnegative square matrices with at least one positive entry per row [67, §3.1, 3.4]. Hajnal [27, (7)] presents basic properties of τ_B, and Seneta [67, §3.1, 3.4] derives explicit expressions. More recently, Artzrouni and others have studied τ_B in the context of more general dynamical systems [2, 3, 4].

Ergodicity Coefficients Defined by Vector Norms. Norm-based coefficients appear as early as 1956 in a paper by Dobrushin [19, §1.4, 1.5]. For an operator S derived from a transition probability function and a particular norm $\|\cdot\|$, Drobushin chooses the metric $d(x,y) = \|x - y\|$, so that

$$\tau(S) = \sup_{\substack{x,y,\in D\\ x\neq y}} \frac{\|S(x-y)\|}{\|x-y\|}.$$

The continuity of norms, together with $z = x - y$ implies

$$\tau(S) = \sup_{\substack{\|z\|=1\\ z^T \mathbb{1}=0}} \|Sz\|.$$

For the remainder of the paper we focus on ergodicity coefficients defined by vector norms, because such coefficients appear most often in the context of stochastic matrices. Rhodius [57, §1] credits Seneta [66, §2] with introducing these coefficients. Unfortunately, there is some confusion associated with their definition, because many authors assume that a vector and its transpose have the same vector norm, i.e. they assume $\|x\|_p = \|x^T\|_p$ [24, 28, 29, 43, 44, 55, 57, 58, 59, 60, 66, 68, 70, 73, 76, 77]. We do not make this assumption.

To be consistent with the commonly used indices for ergodicity coefficients, we define the p-norm ergodicity coefficient for stochastic matrices S as

$$\tau_p(S) = \max_{\substack{\|z\|_p=1 \\ z^T \mathbb{1}=0}} \left\|S^T z\right\|_p , \tag{2.1}$$

where the maximum ranges over $z \in \mathbb{R}^n$. The subscript p indicates the dependence on the norm, and the order of the matrix-vector multiplication has been reversed so that the norm always applies to a column vector. Since $\tau_p(\cdot)$ is a continuous real-valued function on a finite dimensional real vector vector space, there is a vector that achieves the maximum, and the supremum reduces to a maximum.

3. One-Norm Ergodicity Coefficients for Stochastic Matrices.

The coefficient (2.1) in the one-norm applied to a stochastic matrix S is [66, §2], [67, §4.3],

$$\tau_1(S) = \max_{\substack{\|z\|_1=1 \\ z^T \mathbb{1}=0}} \left\|S^T z\right\|_1 \tag{3.1}$$

where the maximum ranges over $z \in \mathbb{R}^n$. This coefficient is also called the "Dobrushin coefficient" or "delta coefficient" [39, 51, 76]. In §3.2 we show that this $\tau_1(S)$ is identical to the explicit expressions in Example 2.5.

3.1. Properties.

We show that τ_1 in (3.1) is a proper coefficient of ergodicity in the sense of Definition 2.2 and that it is submultiplicative and a bound for subdominant eigenvalues.

The ergodicity coefficient $\tau_1(S)$ represents the norm of S restricted to the subspace spanned by the left eigenvectors associated with eigenvalues $\lambda \neq 1$. This is because the elements in each row of the stochastic matrix S sum to one, $S\mathbb{1} = \mathbb{1}$, so that $\mathbb{1}$ is a right eigenvector for $\lambda = 1$, and left eigenvectors associated with eigenvalues $\lambda \neq 1$ are orthogonal to $\mathbb{1}$.

By construction τ_1 is an ergodicity coefficient in the sense of Definition 2.7. Below we show that τ_1 is also an ergodicity coefficient in the sense of Definition 2.2.

THEOREM 3.1 (§4.3 in [67]). *If S, S_1 and S_2 are stochastic matrices then*

1. $0 \leq \tau_1(S) \leq 1$
2. $|\tau_1(S_1) - \tau_1(S_2)| \leq \tau_1(S_1 - S_2)$
3. $\tau_1(S) = 0 \iff \operatorname{rank}(S) = 1$.

Therefore, τ_1 is a proper coefficient of ergodicity in the sense of Definition 2.2.

Proof.

1. The first set of equalities follows from
$$0 \leq \tau_1(S) = \max_{\|z\|_1=1, z^T \mathbb{1}=0} \left\|S^T z\right\|_1 \leq \max_{\|z\|_1=1} \|S^T z\|_1 = \|S^T\|_1 = \|S\|_\infty = 1.$$
2. Let $\tau_1(S_1) \geq \tau_1(S_2)$, and $\tau_1(S_1) = \left\|S_1^T y\right\|_1$ for a vector y with $\|y\|_1 = 1$ and $y^T \mathbb{1} = 0$. Then
$$0 \leq \tau_1(S_1) - \tau_2(S_2) = \|S_1^T y\|_1 - \max_{\substack{\|z\|_1=1 \\ z^T \mathbb{1}=0}} \left\|S_2^T z\right\|_1 \leq \left\|S_1^T y\right\|_1 - \left\|S_2^T y\right\|_1 .$$

The triangle inequality implies

$$\left\|S_1^T y\right\|_1 - \left\|S_2^T y\right\|_1 \le \left\|(S_1 - S_2)^T y\right\|_1 \le \max_{\substack{\|z\|_1=1 \\ z^T \mathbb{1}=0}} \left\|(S_1 - S_2)^T z\right\|_1 = \tau_1(S_1 - S_2).$$

3. If $\tau_1(S) = 0$ then $\left\|S^T z\right\|_1 = 0$ for any z with $z^T \mathbb{1} = 0$, and in particular for $z = \frac{1}{2}(e_i - e_j)$ with $i \neq j$. This means that any two rows of S^T are identical and $\operatorname{rank}(S) = 1$. Conversely, if $\operatorname{rank}(S) = 1$, then $S = \mathbb{1} v^T$ for some stochastic vector v. For any vector z with $z^T \mathbb{1} = 0$ this implies $S^T z = v \mathbb{1}^T z = 0$. Hence $\tau_1(S) = 0$.

Item 1 implies that τ_1 takes on values in $[0, 1]$, and item 2 implies that $\tau_1(S)$ is a continuous function of S. Therefore τ_1 is a coefficient of ergodicity in the sense of Definition 2.2. Furthermore, item 3 implies that τ_1 is a proper coefficient of ergodicity. ∎

Item 2 in Theorem 3.1 also implies that $\tau_1(S)$ is well-conditioned in the absolute sense with respect to absolute changes in S. Since $\tau_1(\cdot)$ is a proper ergodicity coefficient, Theorem 2.3 implies for a doubly stochastic matrix S_d that $\tau_1(\cdot)$ is zero at the same time for S_d and S_d^T [50, Property 1],

$$\tau_1(S_d) = \tau_1(S_d^T) = 0 \iff \operatorname{rank}(S_d) = 1.$$

We now show that $\tau_1(S)$ is an upper bound on all non-unit eigenvalues of a stochastic matrix S, and that τ_1 is submultiplicative.

THEOREM 3.2 (§5.2 in [5], §2 [66]). *If S, S_1 and S_2 are stochastic matrices then*

1. *$|\lambda| \le \tau_1(S)$ for all eigenvalues $\lambda \neq 1$ of S*
2. *$\tau_1(S_1 S_2) \le \tau_1(S_1)\tau_1(S_2)$.*

Proof.

1. If $\lambda \neq 1$ is a real eigenvalue of S then there is a real left eigenvector v with $v^T S = \lambda v^T$ and $\|v\|_1 = 1$. Since v is a left eigenvector, and $\mathbb{1}$ is a right eigenvector for a different eigenvalue, v and $\mathbb{1}$ must be orthogonal, i.e. $v^T \mathbb{1} = 0$. Hence

$$|\lambda| = |\lambda| \|v\|_1 = \|\lambda v\|_1 = \left\|S^T v\right\|_1 \le \max_{\substack{\|z\|=1 \\ z^T \mathbb{1}=0}} \left\|S^T z\right\|_1 = \tau_1(S).$$

If $\lambda \neq 1$ is a complex eigenvalue, then its eigenvectors v are complex, too. Seneta [66, §3] proves $|\lambda| \le \tau_1(S)$ by making use of Lemma 3.3 below and showing that the explicit expression for τ_1 in Theorem 3.4 also holds when the maximum ranges over complex vectors. A different proof in [5, §5.2] constructs a seminorm on $\mathbb{C}^n$ that is equal to $\tau_1(S)$.

2. Let y be a vector with $\tau_1(S_1 S_2) = \|(S_1 S_2)^T y\|_1$, $\|y\|_1 = 1$ and $y^T \mathbb{1} = 0$. The vector $x \equiv S_1^T y / \left\|S_1^T y\right\|_1$ satisfies $\|x\|_1 = 1$ and $x^T \mathbb{1} = 0$. Then

$$\tau_1(S_1 S_2) = \left\|(S_1 S_2)^T y\right\|_1 = \left\|S_2^T S_1^T y\right\|_1 = \left\|S_1^T y\right\|_1 \left\|S_2^T x\right\|_1 \le \tau_1(S_1)\, \tau_1(S_2).$$

∎

The interesting feature of Theorem 3.2 is that the magnitude of complex eigenvalues can be bounded by an expression that is maximized over real vectors. A more general such eigenvalue bound for nonnegative matrices is presented in Theorem 7.17. Seneta [71, p 191] credits the submultiplicative property to Dobrushin [19]; it also appears in a 1975 paper by Kingman [40, (4.10)].

3.2. Explicit Expressions. We show that τ_1 in (3.1) is identical to the two expressions in Example 2.5.

Markov, in 1906, may have been the first to present an explicit expression for the ergodicity coefficient $\tau_1(S)$, as part of a construction of a Weak Law of Large Numbers [46, pp 358-359], [45, 47]. For a stochastic matrix S Markov introduces a quantity $0 < H < 1$ that satisfies

$$H = \frac{1}{2} \max_{ij} \sum_k |s_{ik} - s_{jk}|.$$

In fact $0 \leq H \leq 1$. The quantity H is equal to the first expression for $\tau_1(S)$ in Example 2.5. Seneta comments on this use of H in Markov's works [72, §5], [73, §7]. In the context of the Central Limit Theorem for Markov chains [19, p 70] Dobrushin shows that

$$\sup_{x,y} \frac{\|S^T x - S^T y\|_1}{\|x - y\|_1} = 1 - \min_{ij} \sum_k \min\{s_{ik}, s_{jk}\}.$$

This explains why Seneta refers to $\tau_1(S)$ as the "Markov-Dobrushin coefficient of ergodicity" [73, p 10]. In 1971, Paz [51] refers to $1 - \min_{ij} \sum_k \min\{s_{ik}, s_{jk}\}$ as the δ-coefficient. It was later adopted by other authors, including Tan [77] and Rhodius [55, 58, 59].

We present two proofs to show that the expressions for $\tau_1(S)$ in (3.1) and Example 2.5 are identical, i.e. to show that

$$\max_{\substack{\|z\|_1=1 \\ z^T \mathbb{1}=0}} \left\|S^T z\right\|_1 = \frac{1}{2} \max_{ij} \sum_k |s_{ik} - s_{jk}|.$$

The first proof makes use of an expression for real vectors whose elements sum to zero. Such vectors can be represented as linear combinations with nonnegative coefficients of vectors $e_i - e_j$.

LEMMA 3.3 (Lemma 2.4 in [67]). *If $x \in \mathbb{R}^n$ satisfies $x \neq 0$ and $x^T \mathbb{1} = 0$ then*

$$x = \sum_{i \neq j} y_{ij} \, \frac{e_i - e_j}{2}, \qquad \text{where} \qquad y_{ij} \geq 0, \quad \sum_{i \neq j} y_{ij} = \|x\|_1.$$

Proof. The proof proceeds by induction over the dimension n of x.

If $n = 2$ assume $x_1 > 0$. Then $x^T \mathbb{1} = 0$ implies $x = x_1 \begin{pmatrix} 1 & -1 \end{pmatrix}^T$. Setting $y_{12} \equiv 2x_1$ gives $x = y_{12}(e_1 - e_2)/2$ and $y_{12} = \|x\|_1 > 0$.

Now assume the lemma holds for $n \geq 2$ and we will show it holds for $n + 1$. Let $x \neq 0$ be a vector of dimension $n + 1$ with $x^T \mathbb{1} = 0$, and assume it has been permuted so that $x_n > 0$ and $x_{n+1} < 0$. Without loss of generality we also assume $x_n = \max_{1 \leq i \leq n+1} |x_i|$. Then

$$x = \begin{pmatrix} x_{1:n-1} \\ x_n \\ x_{n+1} \end{pmatrix} = \begin{pmatrix} x_{1:n-1} \\ x_n + x_{n+1} \\ 0 \end{pmatrix} - x_{n+1} \begin{pmatrix} 0_{1:n-1} \\ 1 \\ -1 \end{pmatrix}.$$

Define the vector $\hat{x} = \begin{pmatrix} x_{1:n-1}^T & x_n + x_{n+1} \end{pmatrix}^T$ of dimension n, which satisfies $\hat{x}^T \mathbb{1} = x^T \mathbb{1} = 0$. If $\hat{x} = 0$ then the conclusion follows as in the case $n = 2$. If $\hat{x} \neq 0$ then we apply the induction hypothesis to $\hat{x}$ and obtain

$$\hat{x} = \sum_{i \neq j} y_{ij} \, \frac{e_i - e_j}{2} \qquad \text{where} \quad y_{ij} \geq 0, \quad \sum_{i \neq j} y_{ij} = \|\hat{x}\|_1 \, .$$

Applying the definition of $\hat{x}$ and setting $y_{n,n+1} = -2x_{n+1} > 0$ gives

$$\sum_{i \neq j} y_{ij} + y_{n,n+1} = \|\hat{x}\|_1 - 2x_{n+1} = \|x_{1:n-1}\|_1 + |x_n + x_{n+1}| - 2x_{n+1}.$$

From $x_n = \max_{1 \leq i \leq n+1} |x_i| > 0$ and $x_{n+1} < 0$ follows $x_n + x_{n+1} > 0$. Hence $|x_n + x_{n+1}| = x_n + x_{n+1}$ and $\sum_{i \neq j} y_{ij} + y_{n,n+1} = \|x\|_1$. Furthermore, from the definition of $y_{n,n+1}$ we obtain the desired expression for x,

$$x = \begin{pmatrix} \hat{x} \\ 0 \end{pmatrix} - x_{n+1} \begin{pmatrix} 0_{1:n-1} \\ 1 \\ -1 \end{pmatrix} = \sum_{i \neq j} y_{ij} \, \frac{e_i - e_j}{2} + y_{n,n+1} \, \frac{e_n - e_{n+1}}{2}.$$

□

With the help of Lemma 3.3 we show that (3.1) is identical to Markov's H and to the first expression for $\tau_1(S)$ from Example 2.5.

THEOREM 3.4 (§3.1, §4.3 in [67]). *If $S \in \mathbb{R}^{n \times n}$ is a stochastic matrix then*

$$\tau_1(S) = \frac{1}{2} \max_{ij} \left\| S^T (e_i - e_j) \right\|_1 = \frac{1}{2} \max_{ij} \sum_{k=1}^{n} |s_{ik} - s_{jk}|.$$

Proof. Let $\tau_1(S) = \left\| S^T x \right\|_1$ where $x^T \mathbb{1} = 0$ and $\|x\|_1 = 1$. Applying Lemma 3.3 to x gives $x = \sum_{i \neq j} y_{ij} \, (e_i - e_j) \, /2$ where $\sum_{i \neq j} y_{ij} = \|x\|_1 = 1$. The triangle inequality applied to $\|S^T x\|_1$ yields

$$\tau_1(S) = \left\| S^T x \right\|_1 \leq \sum_{i \neq j} \frac{y_{ij}}{2} \left\| S^T (e_i - e_j) \right\|_1 \leq \frac{1}{2} \max_{ij} \left\| S^T (e_i - e_j) \right\|_1 .$$

Hence $\tau_1(S) \leq \frac{1}{2} \max_{ij} \|S^T (e_i - e_j)\|_1$.

To show the reverse inequality, set $y = (e_i - e_j)/2$ for some $i \neq j$. Then $y^T \mathbb{1} = 0$, $\|y\|_1 = 1$, and

$$\tau_1(S) = \max_{\substack{\|z\|_1 = 0 \\ z^T \mathbb{1} = 0}} \left\| S^T z \right\|_1 \geq \left\| S^T y \right\|_1 = \frac{1}{2} \left\| S^T (e_i - e_j) \right\|_1 .$$

Since this inequality holds for any i and j, we must have $\tau_1(S) \geq \frac{1}{2} \max_{ij} \|S^T (e_i - e_j)\|_1$.

According to the definition of the one-norm, $\|S^T (e_i - e_j)\|_1 = \sum_{k=1}^{n} |s_{ik} - s_{jk}|$, so that $\tau_1(S) = \frac{1}{2} \max_{ij} \sum_{k=1}^{n} |s_{ik} - s_{jk}|$. □

The second proof below for Theorem 3.4 is much simpler and does not require Lemma 3.3.

Proof. (Simpler proof of Theorem 3.4)

Let $y \in \mathbb{R}^n$ be a vector with $y^T \mathbb{1} = 0$, $\|y\|_1 = 1$ and $\tau_1(S) = \|S^T y\|_1$. From $y^T \mathbb{1} = 0$ follows for any vector $x \in \mathbb{R}^n$

$$\|S^T y\|_1 = \left\| \left(S - \mathbb{1} x^T \right)^T y \right\|_1 = \sum_{k=1}^{n} \left| e_k^T \left(S - \mathbb{1} x^T \right)^T y \right| \leq \sum_{k=1}^{n} \left\| \left(S - \mathbb{1} x^T \right) e_k \right\|_\infty ,$$

where the last expression follows from Hölder's inequality and $\|y\|_1 = 1$. Since $\left(S - \mathbb{1}x^T\right) e_k = Se_k - x_k \mathbb{1}$, the norm $\left\|\left(S - \mathbb{1}x^T\right) e_k\right\|_\infty = \max_{1\le l\le n} |s_{lk} - x_k|$ is minimized when x_k is the average of all elements in column Se_k, i.e.

$$x_k = \frac{1}{2}\left(s_{i_k,k} + s_{j_k,k}\right), \qquad \text{where} \quad s_{i_k,k} = \max_l s_{lk}, \quad s_{j_k,k} = \min_l s_{lk}.$$

Then

$$\max_{1\le l\le n} |s_{lk} - x| \le \frac{1}{2}(s_{i_k,k} - s_{j_k,k}) = \frac{1}{2}(e_{i_k} - e_{j_k})^T S e_k.$$

Since the x_k were chosen to minimize $\left\|\left(S - \mathbb{1}x^T\right) e_k\right\|_\infty$, we obtain

$$\begin{aligned}\sum_{k=1}^{n} \left\|\left(S - \mathbb{1}x^T\right) e_k\right\|_\infty &\le \sum_{k=1}^{n} \left|e_k^T S^T (e_{i_k} - e_{j_k})\right| \\ &\le \max_{i,j} \sum_{k=1}^{n} \left|e_k^T S^T (e_i - e_j)\right| = \max_{ij} \|S^T(e_i - e_j)\|_1.\end{aligned}$$

Hence $\|S^T y\|_1 \le \max_{ij} \|S^T(e_i - e_j)\|_1$. ☐

We can view $\tau_1(S)$ as the norm of an (oblique) projection of S, with the projection being onto range$(\mathbb{1})^\perp$.

Corollary 3.5. *If $S \in \mathbb{R}^{n\times n}$ is a stochastic matrix then*

$$\tau_1(S) = \frac{1}{2} \max_{1\le j\le n} \left\|S^T(I - e_j \mathbb{1}^T)\right\|_1.$$

Proof. In the expression for $\tau_1(S)$ from Theorem 3.4 write

$$\max_{ij} \left\|S^T(e_i - e_j)\right\|_1 = \max_{ij} \left\|S^T(I - e_j \mathbb{1}^T) e_i\right\|_1 = \max_{1\le j\le n} \left\|S^T(I - e_j \mathbb{1}^T)\right\|_1.$$

☐

Theorem 3.4, in turn, implies the second expression for $\tau_1(S)$ from Example 2.5.

Corollary 3.6 (§3.1, §4.3 in [67], pp 1733–1734 in [33]). *If $S \in \mathbb{R}^{n\times n}$ is a stochastic matrix then $\tau_1(S) = 1 - \min_{ij} \sum_{k=1}^n \min\{s_{ik}, s_{jk}\}$.*

Proof. Let l and m be indices that achieve the maximum in the second expression from Theorem 3.4,

$$\tau_1(S) = \frac{1}{2} \max_{ij} \sum_k |s_{ik} - s_{jk}| = \frac{1}{2} \sum_{k=1}^{n} |s_{mk} - s_{lk}|.$$

We remove the absolute values as in the proof of Theorem 2.6, and define the set $\mathcal{P}_m$ of all indices k with $s_{mk} \ge s_{lk}$ and the set $\mathcal{P}_l$ of all indices k with $s_{lk} > s_{mk}$. Then

$$\tau_1(S) = \frac{1}{2}\left[\sum_{k\in\mathcal{P}_m} (s_{mk} - s_{lk}) + \sum_{k\in\mathcal{P}_l} (s_{lk} - s_{mk})\right].$$

From $S\mathbb{1} = \mathbb{1}$ follows $\sum_{k\in\mathcal{P}_l} s_{lk} = 1 - \sum_{k\in\mathcal{P}_m} s_{lk}$ and $\sum_{k\in\mathcal{P}_m} s_{mk} = 1 - \sum_{k\in\mathcal{P}_l} s_{mk}$, so that

$$\tau_1(S) = 1 - \sum_{k\in\mathcal{P}_m} s_{lk} - \sum_{k\in\mathcal{P}_l} s_{mk}.$$

The elements in the sums are the smaller of the pair $\{s_{mk}, s_{lk}\}$. That is, $s_{lk} = \min\{s_{lk}, s_{mk}\}$ for $k \in \mathcal{P}_m$, and $s_{mk} = \min\{s_{lk}, s_{mk}\}$ for $k \in \mathcal{P}_l$. Thus the expression for $\tau_1(S)$ simplifies to $\tau_1(S) = 1 - \sum_{k=1}^n \min\{s_{lk}, s_{mk}\}$. The definition of l and m implies

$$\tau_1(S) = \max_{ij} \left[1 - \sum_{k=1}^n \min\{s_{ik}, s_{jk}\}\right] = 1 - \min_{ij} \sum_{k=1}^n \min\{s_{ik}, s_{jk}\}.$$

□

Several authors choose instead $1 - \tau_1(S) = \min_{ij} \sum_{k=1}^n \min\{s_{ik}, s_{jk}\}$ as a coefficient ergodicity [19, 26, 33, 38, 52, 54].

3.3. Condition Number for the Stationary Distribution. We show that $\tau_1(S)$ is a bound for the normwise condition number of the stationary distribution of S.

THEOREM 3.7 (§2 in [70]). *Let S and $S+E$ be stochastic, irreducible matrices with $\pi^T S = \pi^T$, $\widehat{\pi}^T (S+E) = \widehat{\pi}^T$ and $\|\pi\|_1 = \|\widehat{\pi}^T\|_1 = 1$. If $\tau_1(S) < 1$, then*

$$\|\widehat{\pi} - \pi\|_1 \leq \frac{\|E\|_\infty}{1 - \tau_1(S)}.$$

Proof. From $\pi^T(I-S) = 0$ and $\widehat{\pi}^T(I-S) = \widehat{\pi}^T E$ follows

$$(\widehat{\pi} - \pi)^T (I - S) = \widehat{\pi}^T E. \tag{3.2}$$

Because S is irreducible and $\tau_1(S) < 1$, the dominant eigenvalue 1 is simple and we can write $S = \mathbb{1}\pi^T + Q$, where the eigenvalues of Q are less than 1 in magnitude. Substituting this into expression (3.2) and using $\pi^T \mathbb{1} = \widehat{\pi}^T \mathbb{1} = 1$ gives

$$(\widehat{\pi} - \pi)^T (I - Q) = \widehat{\pi}^T E.$$

Because all eigenvalues of Q are less than 1 in magnitude, $I - Q$ is nonsingular and $(I-Q)^{-1} = \sum_{i=0}^\infty Q^i$ [48, p 126]. Thus

$$(\widehat{\pi} - \pi)^T = \widehat{\pi}^T E (I-Q)^{-1} = \sum_{i=0}^\infty y^T Q^i, \qquad \text{where} \quad y^T = \widehat{\pi}^T E.$$

Taking norms and applying the triangle inequality gives

$$\|\widehat{\pi} - \pi\|_1 = \left\|(\widehat{\pi} - \pi)^T\right\|_\infty = \left\|\sum_{i=0}^\infty y^T Q^i\right\|_\infty \leq \sum_{i=0}^\infty \|y^T Q^i\|_\infty = \sum_{i=0}^\infty \left\|\left(Q^i\right)^T y\right\|_1.$$

From $S\mathbb{1} = \mathbb{1}$ and $(S+E)\mathbb{1} = \mathbb{1}$ follows $E\mathbb{1} = 0$, hence $y^T\mathbb{1} = 0$. The submultiplicative property of the ergodicity coefficient from Theorem 3.2 implies

$$\sum_{i=0}^\infty \left\|\left(Q^i\right)^T y\right\|_1 \leq \sum_{i=0}^\infty \tau_1\left(Q^i\right) \|y\|_1 \leq \sum_{i=0}^\infty \left[\tau_1(Q)\right]^i \|y\|_1 \leq \frac{\|y\|_1}{1 - \tau_1(Q)}.$$

Since $y^T\mathbb{1} = 0$ implies $y^T Q = y^T S$, $\tau_1(Q) = \tau_1(S)$. Finally, use the fact that $\|\widehat{\pi}\|_1 = 1$ to bound $\|y\|_1 = \left\|E^T \widehat{\pi}\right\|_1 \leq \|E\|_\infty \|\widehat{\pi}\|_1 = \|E\|_\infty$. □

Theorem 3.7 suggests that $1/(1-\tau_1(S))$ is a bound on the condition number of π with regard to normwise absolute changes in the matrix S. From $\|\pi\|_1 = 1$ and $\|S\|_\infty = 1$ follows

$$\frac{\|\widehat{\pi}-\pi\|_1}{\|\pi\|_1} \leq \frac{1}{1-\tau_1(S)} \frac{\|E\|_\infty}{\|S\|_\infty},$$

so that $1/(1-\tau_1(S))$ is also a bound on the condition number of π with regard to normwise relative changes in S. Since Seneta's derivation [70] appeared in 1988, several tighter bounds for the condition number of π have been derived [15, §4]. Furthermore, optimal condition numbers have been established for ergodicity coefficients applied to the group inverse of $I-S$ [41].

3.4. Scrambling matrices. The bound in Theorem 3.7 for the condition number of the stationary distribution of a stochastic matrix S applies only if $\tau_1(S) < 1$. Similarly, when it comes to eigenvalue bounds, $\tau_1(S)$ is useful only if $\tau_1(S) < 1$. Unfortunately it is not easy to determine when this will happen. Knowledge of the eigenvalues of S does not seem to help.

We know from Theorem 3.2 that if S has more than one eigenvalue of magnitude 1, then $\tau_1(S) = 1$. This would suggest that primitive matrices are good candidates for which we could expect $\tau_1(S) < 1$. However, this turns out to be not always true. The primitive stochastic matrix

$$S = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

has eigenvalues 1, -0.6058, and $-0.7021 \pm 0.6383i$. Although S has only a single eigenvalue of magnitude equal to one, we nevertheless have $\tau_1(S) = 1$. Thus it is not clear how to predict the value of $\tau_1(S)$ from properties like irreducibility, primitivity, or regularity. It is easier to just give a name to matrices for which $\tau_1(S) < 1$.

DEFINITION 3.8 (§2 in [26], §3.2 in [65], p 82 in [67]). *A stochastic matrix S is scrambling if $\tau_1(S) < 1$.*

What can we say about scrambling matrices? Scrambling matrices are a subset of primitive stochastic matrices [26, 67]. Therefore matrices with a single element in each column, such as permutation matrices, cannot be scrambling. However, stochastic matrices are scrambling if no two rows are orthogonal [66, §4]; in other words any two rows share some column in which they both have a positive element [26, §2]. An extreme example is Markov matrices, which have at least one entirely positive column.

DEFINITION 3.9 (§3.2 in [65]). *A stochastic matrix S is a Markov matrix if $\max_j(\min_i s_{ij}) > 0$.*

Thus all Markov matrices are scrambling matrices. The properties of being scrambling and Markov are preserved by multiplication.

THEOREM 3.10 (Lemma 2 in [26], §4 in [66]). *If S and Q are stochastic matrices and one of them is scrambling, then SQ and QS are also scrambling.*

Proof. This follows from the submultiplicative property, $\tau_1(SQ) \leq \tau_1(S)\tau_1(Q)$ in Theorem 3.2. ∎

COROLLARY 3.11 (§3.2 in [65]). *If S and Q are stochastic matrices and one of them is a Markov matrix, then SQ and QS are also Markov matrices.*

4. Infinity-Norm Ergodicity Coefficients for Stochastic Matrices. We present properties and explicit expressions for

$$\tau_\infty(S) = \max_{\substack{\|z\|_\infty=1 \\ z^T\mathbb{1}=0}} \left\|S^T z\right\|_\infty \tag{4.1}$$

where the maximum ranges over $z \in \mathbb{R}^n$ [66, §2].

Unlike $\tau_1(S)$ which ranges between 0 and 1, $\tau_\infty(S)$ has no fixed upper bound. This means $\tau_\infty(S)$ is a coefficient of ergodicity according to Definition 2.7, but not Definition 2.2. Here is an example of a $n \times n$ stochastic matrix for which $\tau_\infty(S)$ grows proportional to n,

$$S = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \qquad \text{with } \tau_\infty(S) = \frac{n}{2}.$$

Here n is even, and the leading $n/2$ rows of S are the same, as are the trailing $n/2$ rows. Other examples of stochastic matrices for which $\tau_\infty(S) > 1$ include a 6×6 matrix in [76, p 861] and 62×62 matrix in [66, §3].

4.1. Properties and Explicit Expressions. The coefficient $\tau_\infty(S)$ has many properties in common with $\tau_1(S)$ from Section 3, because it is bounded, well-conditioned in the absolute sense, submultiplicative, and proper.

THEOREM 4.1 ((7) and (8) in [76]). *If S, S_1 and S_2 are stochastic matrices then*

1. $0 \le \tau_\infty(S) \le \|S\|_1$
2. $|\tau_\infty(S_1) - \tau_\infty(S_2)| \le \tau_\infty(S_1 - S_2)$
3. $\tau_\infty(S) = 0$ *if and only if* $\operatorname{rank}(S) = 1$
4. $|\lambda| \le \tau_\infty(S)$ *for real eigenvalues* $\lambda \ne 1$ *of* S
5. $\tau_\infty(S_1 S_2) \le \tau_\infty(S_1)\, \tau_\infty(S_2)$

Proof. The proofs are analogous to those of Theorems 3.1 and 3.2. ∎

Since $\tau_\infty(\cdot)$ is a proper ergodicity coefficient, Theorem 2.3 implies for a doubly stochastic matrix S_d that $\tau_\infty(\cdot)$ is zero at the same time for S_d and S_d^T,

$$\tau_\infty(S_d) = \tau_\infty(S_d^T) = 0 \iff \operatorname{rank}(S_d) = 1.$$

Below is an explicit expression for $\tau_\infty(S)$.

THEOREM 4.2 (§2 in [76]). *If $S \in \mathbb{R}^{n\times n}$ is a stochastic matrix then*

$$\tau_\infty(S) = \max_{1 \le j \le n} \phi(Se_j),$$

where the function ϕ is defined for $s \in \mathbb{R}^n$ with elements labeled $s_1 \ge \cdots \ge s_n$, and

$$\phi(s) = \begin{cases} \displaystyle\sum_{i=1}^{n/2} s_i - \sum_{i=n/2+1}^{n} s_i & \text{if } n \text{ is even,} \\[2ex] \displaystyle\sum_{i=1}^{(n-1)/2} s_i - \sum_{i=(n+3)/2}^{n} s_i & \text{if } n \text{ is odd.} \end{cases}$$

Proof. Let s be a column of S and x a vector that together achieve the maximum in $\tau_\infty(S)$, i.e.

$$\tau_\infty(S) = \max_{\substack{\|z\|_\infty=1 \\ z^T\mathbb{1}=0}} \left\|S^T z\right\|_\infty = \max_{\substack{\|z\|_\infty=1 \\ z^T\mathbb{1}=0}} \max_j \left|e_j^T S^T z\right| = |s^T x|.$$

The idea is to incorporate the constraint $x^T\mathbb{1} = 0$ into the matrix vector product, by writing $x^T s = x^T(s - \theta\mathbb{1})$ for any scalar θ. The Hölder inequality and $\|x\|_\infty = 1$ imply

$$|x^T s| = |x^T(s-\theta\mathbb{1})| \leq \|x\|_\infty \|s-\theta\mathbb{1}\|_1 = \|s-\theta\mathbb{1}\|_1.$$

Label the elements of s so that $s_1 \geq \ldots \geq s_n$, and choose $\theta = s_l$ where $l = n/2$ for n even, and $l = (n+1)/2$ for n odd. This gives for even n and $l = n/2$,

$$\begin{aligned}\|s-\theta\mathbb{1}\|_1 &= \sum_{i=1}^{n} |s_i - s_{n/2}| = \sum_{i=1}^{n/2}(s_i - s_{n/2}) + \sum_{i=n/2+1}^{n}(s_{n/2} - s_i) \\ &= \sum_{i=1}^{n/2} s_i - \sum_{i=n/2+1}^{n} s_i = \phi(s),\end{aligned}$$

while for odd n and $l = (n+1)/2$ we obtain

$$\|s-\theta\mathbb{1}\|_1 = \sum_{i=1}^{n}|s_i - s_l| = \sum_{i=1}^{l}(s_i - s_l) + \sum_{i=l+1}^{n}(s_l - s_i) = \sum_{i=1}^{l-1} s_i - \sum_{i=l+1}^{n} s_i = \phi(s).$$

We have shown that $\tau_\infty(S) = |x^T s| \leq \phi(s) \leq \max_{1\leq i\leq n}\phi(Se_i)$.

To show the reverse inequality, let s be a column of S so that $\max_{1\leq i\leq n}\phi(Se_i) = \phi(s)$. Let P be a permutation matrix that orders the elements of Ps in decreasing magnitude, i.e. $Ps = \begin{pmatrix} s_1 & \ldots & s_n\end{pmatrix}^T$ with $s_1 \geq \ldots \geq s_n$. Define the vector

$$y = \begin{cases} P\begin{pmatrix}\mathbb{1}_{n/2}^T & -\mathbb{1}_{n/2}^T\end{pmatrix}^T & \text{if } n \text{ even} \\ P\begin{pmatrix}\mathbb{1}_{(n-1)/2}^T & 0 & -\mathbb{1}_{(n-1)/2}^T\end{pmatrix}^T & \text{if } n \text{ odd.}\end{cases}$$

Then $\phi(s) = |y^T s|$, $y^T\mathbb{1} = 0$ and $\|y\|_\infty = 1$. Hence $\max_{1\leq i\leq n}\phi(Se_i) = \phi(s) = |y^T s| \leq \tau_\infty(S)$. ☐

Like $\tau_1(S)$ in Corollary 3.5, we can also view $\tau_\infty(S)$ as the norm of an (oblique) projection of S, with the projection being onto range$(\mathbb{1})^\perp$.

COROLLARY 4.3. *If $S \in \mathbb{R}^{n\times n}$ is a stochastic matrix then for some $1 \leq k \leq n$,*

$$\tau_\infty(S) = \left\|S^T(I - e_k\mathbb{1}^T)\right\|_\infty.$$

Proof. The proof of Theorem 4.2 implies that $\tau_\infty(S) = \max_{1\leq j\leq n}\|Se_j - \theta\mathbb{1}\|_1$, where θ is an element of Se_j, that is, $\theta = e_k^T Se_j$ for some $1 \leq k \leq n$. Then $Se_j - e_k^T Se_j\mathbb{1} = \left(I - \mathbb{1}e_k^T\right)Se_j$ and

$$\tau_\infty(S) = \max_{1\leq j\leq n}\left\|\left(I-\mathbb{1}e_k^T\right)Se_j\right\|_1 = \left\|\left(I-\mathbb{1}e_k^T\right)S\right\|_1 = \left\|S^T\left(I - e_k\mathbb{1}^T\right)\right\|_\infty.$$

$\square$

Theorem 4.2 also implies lower and upper bounds for $\tau_\infty(S)$ in terms of the coefficient $\alpha(S) = \max_j \max_{il} |s_{ij} - s_{lj}|$ from Example 2.5.

Theorem 4.4 (Proposition 4 in [76]). *If $S \in \mathbb{R}^{n\times n}$ is a stochastic matrix then*

$$\alpha(S) \le \tau_\infty(S) \le \begin{cases} \frac{n}{2}\,\alpha(S) & \text{if } n \text{ even},\\ \frac{n-1}{2}\,\alpha(S) & \text{if } n \text{ odd}.\end{cases}$$

Proof. We start with the lower bound. For every column j of S, let k_j be an index that achieves the maximum in

$$\min_{1\le k\le n} \sum_{i=1}^n |s_{ij} - s_{kj}| = \sum_{i=1}^n |s_{ij} - s_{k_j j}|.$$

Let column l of S achieve the maximum in $\alpha(S)$ so that $\alpha(S) = |s_{i_1 l} - s_{i_2 l}|$. Adding and subtracting $s_{k_l l}$ inside $\alpha(S)$ and applying the triangle inequality gives

$$\begin{aligned}\alpha(S) &= |s_{i_1 l} - s_{k_l l} + s_{k_l l} - s_{i_2 l}| \le |s_{i_1 l} - s_{k_l l}| + |s_{i_2 l} - s_{k_l l}| \le \sum_{i=1}^n |s_{il} - s_{k_l l}|\\ &\le \max_j \sum_{i=1}^n |s_{ij} - s_{k_j j}| = \tau_\infty(S),\end{aligned}$$

where the last inequality follows from the proof of Theorem 4.2. There we showed that $\tau_\infty(S) = \max_j \|Se_j - \theta\mathbb{1}\|_1 = \max_j \sum_{i=1}^k |s_{ij} - s_{k_j j}|$.

As for the upper bound, let column j assume the maximum in $\tau_\infty(S)$, and assume that the rows of S have been permuted so that $s_{1j} \ge \ldots \ge s_{nj}$. Theorem 4.2 implies for even n

$$\tau_\infty(S) = \sum_{i=1}^{n/2} (s_{ij} - s_{n/2+i,j}) \le \frac{n}{2}\,\max_{i,l} |s_{ij} - s_{lj}| \le \frac{n}{2}\,\alpha(S),$$

and for odd n,

$$\tau_\infty(S) = \sum_{i=1}^{(n-1)/2} (s_{ij} - s_{(n+1)/2+i,j}) \le \frac{n-1}{2}\,\max_{i,l} |s_{ij} - s_{lj}| \le \frac{n-1}{2}\,\alpha(S).$$

$\square$

Theorem 4.4 implies the values below for the maximum of $\tau_\infty(S)$ over all stochastic matrices S. Rhodius [55] and Lešanovský [44] attribute this result to Tan [76].

Corollary 4.5. *If $\mathcal{S}_n$ is the set of stochastic matrices in $\mathbb{R}^{n\times n}$ then*

$$\max_{S\in\mathcal{S}_n} \tau_\infty(S) = \begin{cases} \frac{n}{2} & \text{if } n \text{ even},\\ \frac{n-1}{2} & \text{if } n \text{ odd}.\end{cases}$$

Proof. Theorems 2.6 and 3.1 imply $\alpha(S) \le \tau_1(S) \le 1$. Together with Theorem 4.4 this gives $\max_{S\in\mathcal{S}_n} \tau_\infty(S) \le n/2$ for even n, and $\max_{S\in\mathcal{S}_n} \tau_\infty(S) \le (n-1)/2$ for odd n. To see that the bounds are tight, let S_1 be a $n \times n$ matrix where n is even, and S_2 a $n \times n$ where n is odd, and

$$S_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \qquad S_2 = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Then $\tau_\infty(S_1) = n/2$ and $\tau_\infty(S_2) = (n-1)/2$. ▯

Corollary 4.5 also follows from a more general result for p-norms in Theorem 5.8.

4.2. Relations between Infinity-Norm and One-Norm Coefficients. We present relations between $\tau_1(S)$ and $\tau_\infty(S)$. For stochastic matrices of very small dimension, the two coefficients are identical.

THEOREM 4.6 (§3.3 in [66]). *If S is a 2×2 or 3×3 stochastic matrix then $\tau_1(S) = \tau_\infty(S)$.*

Proof. For $n = 2$ Theorem 3.4 implies $\tau_1(S) = \frac{1}{2}\left\|S^T(e_1 - e_2)\right\|_1$, while Theorem 4.2 implies $\tau_\infty(S) = \left\|S^T(e_1 - e_2)\right\|_\infty$. Since $S^T(e_1 - e_2) = \begin{pmatrix} s & -s \end{pmatrix}^T$ for some scalar s, we obtain $\tau_1(S) = \tau_\infty(S) = |s|$.

For $n = 3$ Theorem 3.4 implies $\tau_1(S) = \frac{1}{2}\max_{1 \le i,j \le 3}\left\|S^T(e_i - e_j)\right\|_1$, while Theorem 4.2 implies $\tau_\infty(S) = \max_{1 \le i,j \le 3}\left\|S^T(e_i - e_j)\right\|_\infty$. For $i \neq j$ we have $v \equiv S^T(e_i - e_j) = \begin{pmatrix} v_1 & v_2 & -v_1 - v_2 \end{pmatrix}^T$ for some scalars v_1 and v_2. If v_1 and v_2 have the same sign, then $\|v\|_1 = 2(|v_1| + |v_2|)$ and $\|v\|_\infty = |v_1| + |v_2|$, hence $\tau_1(S) = \tau_\infty(S)$. If v_1 and v_2 have different signs, assume without loss of generality that $|v_1| \ge |v_2|$, so that $|-v_1 - v_2| = |v_1| - |v_2|$. Then $\|v\|_1 = 2|v_1|$ and $\|v\|_\infty = |v_1|$, hence $\tau_1(S) = \tau_\infty(S)$. ▯

For stochastic matrices of any order n, we can show the following relations.

THEOREM 4.7. *If $S \in \mathbb{R}^{n \times n}$ is a stochastic matrix and $n \ge 2$ then*

$$\frac{2}{n}\tau_1(S) \le \tau_\infty(S) \le \begin{cases} \frac{n}{2}\,\tau_1(S) & \text{if } n \text{ even}, \\ \frac{n-1}{2}\,\tau_1(S) & \text{if } n \text{ odd}. \end{cases}$$

Proof. We start with the lower bound. In Theorem 3.4, let $i \neq j$ be indices so that $2\tau_1(S) = \left\|S^T(e_i - e_j)\right\|_1 \le n\|S^T(e_i - e_j)\|_\infty$. From $(e_i - e_j)^T\mathbb{1} = 0$ and $\|e_i - e_j\|_\infty = 1$ follows $\|S^T(e_i - e_j)\|_\infty \le \tau_\infty(S)$.

As for the upper bound, Theorem 4.4 implies $\tau_\infty(S) \le \frac{n}{2}\alpha(S)$ for even n and $\tau_\infty(S) \le \frac{n-1}{2}\alpha(S)$ for odd n. Now combine this with the bound $\alpha(S) \le \tau_1(S)$ from Theorem 2.6. ▯

The upper bound in Theorem 4.7 is tight; for the matrices S_1 and S_2 in the proof of Corollary 4.5 we have $\tau_\infty(S_1) = \frac{n}{2}\tau_1(S_1)$ and $\tau_\infty(S_2) = \frac{n-1}{2}\tau_1(S_2)$.

Doubly Stochastic Matrices. Theorem 4.7 relates $\tau_1(S)$ and $\tau_\infty(S)$ for stochastic matrices S. In the special case of doubly stochastic matrices S_d, the transpose S_d^T is also stochastic, so that we can try to relate $\tau_1(S_d)$ and $\tau_\infty(S_d^T)$. The motivation is as follows. We know that $\|S\|_1 = \|S^T\|_\infty$ in general. However, $\tau_1(\cdot)$ and $\tau_\infty(\cdot)$ are norms of matrices that restricted to subspaces orthogonal to the dominant right eigenvector $\mathbb{1}$. Since S_d and S_d^T have the same right dominant eigenvector $\mathbb{1}$, then is there anything we can say about $\tau_1(S_d)$ and $\tau_\infty(S_d^T)$?

THEOREM 4.8 (p 344 in [68]). *If $S_d \in \mathbb{R}^{n\times n}$ is a doubly stochastic matrix then*

$$\tau_1(S_d) = 1 \implies \tau_\infty\left(S_d^T\right) = 1.$$

Proof. The explicit expression $\tau_1(S_d) = 1 - \min_{ij} \sum_{k=1}^n \min\{s_{ik}, s_{jk}\}$ from Corollary 3.6 implies: If $\tau_1(S_d) = 1$ then there exist indices i and j so that $s_{ik} = 0$ or $s_{jk} = 0$ for every k. First consider the case n even. Then one row of S_d contains at least $n/2$ zeros, implying that a column c of S_d^T contain at least $n/2$ zeros. If we label the elements of c so that $c_1 \geq \ldots \geq c_n$ then $c_i = 0$, $n/2+1 \leq i \leq n$. Theorem 4.2 implies $\phi(c) = \sum_{i=1}^{n/2} c_i - \sum_{i=n/2+1}^n c_i = 1$ so that $\tau_\infty(S_d^T) = 1$.

If n is odd, an analogous argument implies that one row of S_d contains at least $(n+1)/2$ zeros and $\phi(c) = \sum_{i=1}^{(n-1)/2} c_i - \sum_{i=(n+3)/2}^n c_i = 1$. ◻

The converse of Theorem 4.8 is not true [68, p 344]. The symmetric matrix

$$S_d = \begin{pmatrix} 1/3 & 0 & 0 & 2/3 \\ 0 & 5/6 & 0 & 1/6 \\ 0 & 0 & 5/6 & 1/6 \\ 2/3 & 1/6 & 1/6 & 0 \end{pmatrix}$$

has $\tau_\infty\left(S_d^T\right) = 1$ but $\tau_1(S_d) = 5/6 < 1$.

We can make a stronger statement for a particular class of doubly stochastic matrices, where all rows contain the same elements, but not necessarily in the same order. Such matrices have been studied in [53].

THEOREM 4.9 (p 345 in [68]). *If $S_d \in \mathbb{R}^{n\times n}$ is a doubly stochastic matrix in which all rows contain the same elements, then $\tau_1(S_d) \leq \tau_\infty\left(S_d^T\right)$.*

Proof. First consider the case when n is even. Since all rows of S_d contain the same elements, $s_1 \geq \ldots \geq s_n$, and $\sum_{i=1}^n s_i = 1$, Theorem 4.2 implies

$$\tau_\infty(S_d^T) = \sum_{i=1}^{n/2} s_i - \sum_{i=n/2+1}^{n} s_i = 1 - 2\sum_{i=n/2+1}^{n} s_i.$$

From $\sum_{k=1}^n \min\{s_{il}, s_{jk}\} \geq 2\sum_{i=n/2+1}^n s_i$ and Corollary 3.6 follows

$$\tau_1(S) = 1 - \min_{ij} \sum_{k=1}^n \min\{s_{ik}, s_{jk}\} \leq 1 - 2\sum_{i=n/2+1}^{n} s_i = \tau_\infty(S_d^T).$$

If n is odd, an analogous argument implies

$$\tau_\infty(S_d^T) = \sum_{i=1}^{(n-1)/2} s_i - \sum_{i=(n+3)/2}^{n} s_i = 1 - 2\sum_{i=(n+3)/2}^{n} s_i - s_{(n+1)/2},$$

and

$$\tau_1(S_d) = 1 - \min_{ij} \sum_{k=1}^n \min\{s_{ik}, s_{jk}\} \leq 1 - 2\sum_{i=(n+3)/2}^{n} s_i - s_{(n+1)/2} = \tau_\infty(S_d^T).$$

□

Corollary 4.10 (p 345 in [68]). *If $S_d \in \mathbb{R}^{n\times n}$ is a symmetric stochastic matrix in which all rows contain the same elements then $\tau_1(S_d) \leq \tau_\infty(S_d)$.*

For the special class of symmetric matrices in Corollary 4.10, one can show that equality holds, i.e. $\tau_1(S_d) = \tau_\infty(S_d)$, for $n = 2, 3, 4$ [68, p 345-346].

5. p-Norm Ergodicity Coefficients for Stochastic Matrices. For any integer $p \geq 1$, the p-norm ergodicity coefficient of a stochastic matrix S is [66, §2]

$$\tau_p(S) = \max_{\substack{\|z\|_p=1 \\ z^T\mathbb{1}=0}} \left\|S^T z\right\|_p \tag{5.1}$$

where the maximum ranges over $z \in \mathbb{R}^n$. We present basic properties of $\tau_p(S)$, and derive the maximal value of $\tau_P(S)$ over all stochastic matrices S.

The coefficient $\tau_p(S)$ has the same basic properties as $\tau_\infty(S)$ in Theorem 4.1; it is bounded, well-conditioned in the absolute sense, proper, and submultiplicative.

Theorem 5.1 (§1 in [77], [55]). *If S, S_1 and S_2 are stochastic matrices then*

1. $0 \leq \tau_p(S) \leq \left\|S^T\right\|_p$
2. $|\tau_p(S_1) - \tau_p(S_2)| \leq \tau_p(S_1 - S_2)$
3. $\tau_p(S) = 0$ *if and only if* $\operatorname{rank}(S) = 1$
4. $|\lambda| \leq \tau_p(S)$ *for real eigenvalues* $\lambda \neq 1$ *of* S
5. $\tau_p(S_1 S_2) \leq \tau_p(S_1)\,\tau_p(S_2)$.

Proof. The proofs are analogous to those of Theorems 3.1 and 3.2. □

For 2×2 stochastic matrices, all coefficients $\tau_p(S)$ are the same and identical to the magnitude of the subdominant eigenvalue of S.

Theorem 5.2 (p 585 in [66]). *If S is a 2×2 stochastic matrix with eigenvalues 1 and λ, then $\tau_p(S) = |\lambda|$ for all integers $p \geq 1$.*

Proof. Let $S = \begin{pmatrix} s_1 & 1-s_1 \\ s_2 & 1-s_2 \end{pmatrix}$ with $0 \leq s_1, s_2 \leq 1$. Then $\lambda = s_1 - s_2$. If $\lambda = 1$ then $S = I$, and $\tau_p(S) = 1$ for all p. If $\lambda < 1$ then vectors z with $z^T\mathbb{1} = 0$ and $\|z\|_p = 1$ satisfy $S^T z = \lambda z$. Hence $\tau_p(S) = |\lambda|$ for all p. □

5.1. The Maximal Value Over All Stochastic Matrices. In 1988 Rhodius [55] determined, for any p-norm, the maximal values of $\tau_p(S)$ over all stochastic matrices S. To this end he showed that $\max_S \tau_p(S)$ is achieved by an extreme point, which is a stochastic matrix Q that has a single one in each row. Then he exploited the particular structure of $Q^T z$ to determine $\max_z \|Q^T z\|_1$ as a function of p and the matrix dimension n. We illustrate this development.

To start with, we present two compactness results, for which detailed proofs can be found in [64, §2.5.4]

Lemma 5.3. *The set $\mathcal{S}_n$ of $n\times n$ stochastic matrices is convex and compact. The set of extreme points $\operatorname{Extr}(\mathcal{S}_n)$ consists of stochastic matrices that have a single one in each row.*

Lemma 5.4. *The set $\mathcal{H}_n = \left\{x \,:\, x \in \mathbb{R}^n,\ x^T\mathbb{1} = 0 \text{ and } \|x\|_p = 1\right\}$ is compact.*

First we show that for the maximum over all stochastic matrices it suffices to look at the extreme points of the set of stochastic matrices.

Theorem 5.5 (Theorem 1(a) in [55]). *For integers $p \geq 1$*

$$\max_{S\in\mathcal{S}_n} \tau_p(S) = \max_{z\in\mathcal{H}_n} \max_{Q\in\operatorname{Extr}(\mathcal{S}_n)} \|Q^T z\|_p.$$

Proof. Since $\mathcal{S}_n$ and $\mathcal{H}_n$ are compact sets, see Lemmas 5.3 and 5.4, and since the p-norm is a continuous real-valued function, we can switch the two maxima below,

$$\max_{S\in\mathcal{S}_n} \tau_p(S) = \max_{S\in\mathcal{S}_n} \max_{z\in\mathcal{H}_n} \|S^T z\|_p = \max_{z\in\mathcal{H}_n} \max_{S\in\mathcal{S}_n} \|S^T z\|_p.$$

Fix $z \in \mathcal{H}_n$. Then the convex function $f(S) = \|S^T z\|_p$ on the convex compact set $\mathcal{S}_n$ attains its maximum at an extreme point [31, p 535]. Thus $\max_{S\in\mathcal{S}_n} f(S) = \max_{Q\in\mathrm{Extr}(\mathcal{S}_n)} f(Q)$. □

For matrices $Q \in \mathrm{Extr}(\mathcal{S}_n)$ one can write vectors $Q^T z$ in terms of sets that record the position of ones in each column of Q.

Remark 5.1 (p 142 in [55]). *If $Q \in \mathrm{Extr}(\mathcal{S}_n)$ and $z \in \mathbb{R}^n$ then*

$$Q^T z = \begin{pmatrix} \sum_{i\in D_1} z_i \\ \vdots \\ \sum_{i\in D_n} z_i \end{pmatrix},$$

where the set D_j contains the indices of all rows that have a 1 in position j, that is, $i \in D_j$ if $q_{ij} = 1$.

We illustrate this for the case $n = 3$. Let

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad z = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}.$$

The stochastic matrix Q is in $\mathrm{Extr}(\mathcal{S}_3)$, and

$$Q^T z = \begin{pmatrix} z_1 + z_2 \\ 0 \\ z_3 \end{pmatrix} = \begin{pmatrix} \sum_{i\in D_1} z_i \\ \sum_{i\in D_2} z_i \\ \sum_{i\in D_3} z_i \end{pmatrix}.$$

From $q_{11} = q_{21} = 1$ follows $D_1 = \{1, 2\}$. Since $q_{i2} = 0$ for all i we have $D_2 = \emptyset$, while $q_{33} = 1$ implies $D_3 = \{3\}$.

Now we show that $\max_S \tau_p(S)$ can be obtained by summing particular elements in a vector that achieves the maximum.

Corollary 5.6 (p 144 in [55]). *For integers $p \geq 1$*

$$\max_{S\in\mathcal{S}_n} \tau_p(S) = \max_{1\leq l\leq n-1} \max_{z\in\mathcal{H}_n\cap\mathcal{A}_l} 2^{1/p}(z_1 + \cdots + z_l),$$

where $\mathcal{A}_l = \{z \in \mathbb{R}^n : \; z_1, \ldots, z_l > 0, \; z_{l+1}, \ldots, z_n \leq 0\}$.

Proof. We only give a sketch of the main argument. Remark 5.1 implies

$$\max_{S\in\mathcal{S}_n} \tau_p(S) = \max_{z\in\mathcal{H}_n} \max_{D_1,\ldots,D_n} \left\| \begin{pmatrix} \sum_{i\in D_1} z_i \\ \vdots \\ \sum_{i\in D_n} z_i \end{pmatrix} \right\|_p,$$

where the maximum ranges over all sets D_j that satisfy $D_1 \cup \ldots \cup D_n = \{1, \ldots, n\}$, and $D_i \cap D_j = \emptyset$ for $i \neq j$. Rhodius [55] does not give a proof that this expression is equal to the desired expression $\max_{1\leq l\leq n-1} \max_{z\in\mathcal{H}_n\cap\mathcal{A}_l} 2^{1/p}(z_1 + \cdots + z_l)$. However, an inequality is easy to see.

A vector $z \in \mathcal{H}_n \cap \mathcal{A}_l$ satisfies

$$0 = z^T \mathbb{1} = z_{1:l}^T \mathbb{1} + z_{l+1:n}^T \mathbb{1} = \|z_{1:l}\|_1 - \|z_{l+1:n}\|_1.$$

Hence $\|z_{l+1:n}\|_1 = \|z_{1:l}\|_1$. Let the maximum of the desired expression be achieved by an index k and a vector $x \in \mathcal{H}_n \cap \mathcal{A}_k$, i.e.

$$\max_{1 \le l \le n-1} \max_{z \in \mathcal{H}_n \cap \mathcal{A}_l} (z_1 + \cdots + z_l) = x_1 + \cdots + x_k = \|x_{1:k}\|_1.$$

Then

$$\max_{z \in \mathcal{H}_n} \max_{D_1, \ldots, D_n} \left\| \begin{pmatrix} \sum_{i \in D_1} z_i \\ \vdots \\ \sum_{i \in D_n} z_i \end{pmatrix} \right\|_p \ge \left\| \begin{pmatrix} x_1 + \cdots + x_k \\ x_{k+1} + \cdots + x_n \end{pmatrix} \right\|_p$$
$$= (\|x_{1:k}\|_1^p + \|x_{k+1:n}\|_1^p)^{1/p} = 2^{1/p} \|x_{1:k}\|_1.$$

□

Next we characterize vectors z that achieve the maximum in Corollary 5.6 and show that their elements $z_1, \ldots, z_l$ can be chosen to be all the same.

Theorem 5.7 (Theorem 2 in [55]). *For integers $p \ge 1$, the function*

$$f(z, l) = 2^{1/p} (z_1 + \cdots + z_l)$$

achieves its maximum over $\mathcal{H}_n \cap \mathcal{A}_l$ at vectors z with $z_1 = \cdots = z_l$ and $z_{l+1} = \cdots = z_n$.

Proof. Let $x \in \mathcal{H}_n \cap \mathcal{A}_l$ be a vector where $f(z, l)$ achieves the maximum, i.e. $f(x, l) = \max_{z \in \mathcal{H}_n \cap \mathcal{A}_l} f(z, l)$. We construct a vector y of the desired form by setting

$$y_1 = \cdots = y_l = \frac{1}{l} (x_1 + \cdots + x_l), \qquad y_{l+1} = \cdots = y_n = \frac{1}{n-l} (x_{l+1} + \cdots + x_n).$$

Then $y^T \mathbb{1} = 0$, $y \in \mathcal{A}_l$, and $f(y, l) = f(x, l)$. We will show that $f(y/\|y\|_p, l) = f(x, l)$. Since x achieves the maximum of $f(z, l)$, we must have $f(y/\|y\|_p, l) \le f(x, l)$. Suppose we can show that $\|y\|_p \le 1$. Combined with $f(x, l) = f(y, l)$ this implies

$$f\left(\frac{y}{\|y\|_p}, l\right) = \frac{1}{\|y\|_p} f(y, l) = \frac{1}{\|y\|_p} f(x, l) \ge f(x, l).$$

Therefore $f(y/\|y\|_p, l) = f(x, l)$.

We still need to show that $\|y\|_p \le 1$. Since the leading l elements of y are the same, and so are the trailing $n-l$ elements, we get $\|y\|_p^p = l|y_1|^p + (n-l)|y_{l+1}|^p$. Write $y_1 = x_{1:l}^T \mathbb{1}_l / l$, where $x_{1:l} = \begin{pmatrix} x_1 & \ldots & x_l \end{pmatrix}$. The Hölder inequality with $1/p + 1/q = 1$ gives

$$|y_1| \le \frac{1}{l} \|x_{1:l}\|_p \|\mathbb{1}_l\|_q = \frac{1}{l} \|x_{1:l}\|_p l^{1/q} = l^{-1/p} \|x_{1:l}\|_p.$$

Similarly, $|y_{1+l}| \le (n-l)^{-1/p} \|x_{l+1:n}\|_p$, where $x_{l+1:n} = \begin{pmatrix} x_{l+1} & \ldots & x_n \end{pmatrix}$. Substituting the bounds for $|y_1|$ and $|y_{l+1}|$ into the above expression for $\|y\|_p^p$ implies $\|y\|_p \le \|x\|_p = 1$. □

At last, the characterization of the vectors in Theorem 5.7 makes it possible to determine explicit values for the maximum in Corollary 5.6.

THEOREM 5.8 (Theorem 3 in [55]). *For integers $p \geq 1$*

$$\max_{S \in \mathcal{S}_n} \tau_p(S) = \begin{cases} \left(\frac{n}{2}\right)^{1-1/p} & \text{if } n \text{ is even,} \\ \left(\frac{1}{2}\right)^{1-1/p} \left(\frac{2}{(n+1)^{1-p}+(n-1)^{1-p}}\right)^{1/p} & \text{if } n \text{ is odd.} \end{cases}$$

Proof. Corollary 5.6 and Theorem 5.7 imply

$$\max_{S \in \mathcal{S}_n} \tau_p(S) = \max_{1 \leq l \leq n-1} \max_{z \in \mathcal{H}_n \cap \mathcal{A}_l} f(z,l) = 2^{1/p} \max_{1 \leq l \leq n-1} \max_{z \in \mathcal{H}_n \cap \mathcal{A}_l} l z_1.$$

We need to determine $l z_1$ so that $\max_{S \in \mathcal{S}_n} \tau_p(S)$ does not depend on l or z. From $z^T \mathbb{1} = 0$ follows $l z_1 + (n-l) z_{l+1} = 0$, hence $z_{l+1} = -l z_1/(n-l)$. Substituting this into $\|z\|_p^p = 1$ gives $l z_1 = \left(l^{1-p} + (n-l)^{1-p}\right)^{-1/p}$. Hence

$$2^{1/p} \max_{1 \leq l \leq n-1} \max_{z \in \mathcal{H}_n \cap \mathcal{A}_l} l z_1 = \left(\frac{2}{l^{1-p} + (n-l)^{1-p}}\right)^{1/p}.$$

This expression is maximized if $l = n/2$ for even n, and $l = (n \pm 1)/2$ for odd n. □

In the special case $p = \infty$, Theorem 5.8 reduces to Corollary 4.5.

6. Ergodicity Coefficients Based on the Stationary Distribution. In 1983 Tan [77] defined the ergodicity coefficient

$$\tau_p(\pi, S) = \max_{\substack{\|z\|_p = 1 \\ z^T \pi = 0}} \left\|S^T z\right\|_p$$

where $\pi \neq 0$ is the stationary distribution of the stochastic matrix S, i.e. $\pi^T S = \pi^T$, and the maximum ranges over $z \in \mathbb{R}^n$. We present properties of these coefficients, and explicit expression for the coefficients in the one-, infinity- and two-norms.

Like the previous coefficients based on the vector $\mathbb{1}$, see Theorem 5.1, the coefficient $\tau_p(\pi, S)$ is bounded and proper. However, because different stochastic matrices have different stationary distributions, $\tau_p(\pi, S)$ is submultiplicative only with regard to powers of the same matrix. Continuity is not clear either because π depends on S.

THEOREM 6.1 (p 279 in [77]). *If S is a stochastic matrix, $\pi^T S = \pi^T$ and $\pi \neq 0$, then*

1. $0 \leq \tau_p(\pi, S) \leq \left\|S^T\right\|_p$.
2. $\tau_p(\pi, S) = 0$ *if and only if* $\operatorname{rank}(S) = 0$.
3. $|\lambda| \leq \tau_p(\pi, S)$ *for real eigenvalues* $\lambda \neq 1$.
4. $\tau_p\left(\pi, S^{l+k}\right) \leq \tau_p\left(\pi, S^l\right)\, \tau_p\left(\pi, S^k\right)$ *for integers* $l, k \geq 1$.

Proof. The proofs are analogous to those of Theorems 3.1 and 3.2. □

With regard to explicit expressions for the one-norm and infinity-norm coefficients, the idea is to view π as a diagonal scaling of $\mathbb{1}$, i.e. $\pi = D\mathbb{1}$ where $D = \operatorname{diag}(\pi)$.

For the one-norm coefficient $\tau_1(\pi, S) = \max_{\|z\|_1 = 1, z^T \pi = 0} \left\|S^T z\right\|_1$ we start with the easier case when all elements of π are non zero. For example, if the stochastic matrix S is irreducible, then all elements of its stationary distribution are positive.

THEOREM 6.2 (Remark p 284 in [77]). *If S is an irreducible stochastic matrix, $\pi^T S = \pi^T$ where $\pi > 0$, then*

$$\tau_1(\pi, S) = \max_{i \neq j} \frac{\left\|S^T D^{-1}(e_i - e_j)\right\|_1}{\left\|D^{-1}(e_i - e_j)\right\|_1},$$

where $D = \text{diag}(\pi)$.

Proof. This follows from a more general result for real matrices in Theorem 7.8. $\square$

Corollary 6.3. *If* $S \in \mathbb{R}^{n\times n}$ *is a stochastic matrix then*

$$\tau_1(S) = \frac{1}{2} \max_{1\le j\le n} \left\| S^T (I - e_j \mathbb{1}^T) \right\|_1 .$$

If the stochastic matrix S is reducible, then there exists a permutation matrix P so that

$$PSP^T = \begin{pmatrix} A & 0 \\ C & B \end{pmatrix} \tag{6.1}$$

where the $l \times l$ matrix A is irreducible, B is square, and each row of C has a positive entry [7, p 223]. As a consequence, the stationary distribution π can be permuted to $P\pi = \begin{pmatrix} \pi_{1:l}^T & 0 \end{pmatrix}^T$, where $\pi_{1:l}$ is a $l \times 1$ vector with $\pi_{1:l} > 0$. In this case the ergodicity coefficient of S can be expressed in terms of the ergodicity coefficient of the irreducible submatrix A.

Theorem 6.4 (Theorem 3 in [77]). *Let* S *be a stochastic reducible matrix of the form (6.1) with stationary distribution* $P\pi = \begin{pmatrix} \pi_{1:l}^T & 0 \end{pmatrix}^T$ *where* $\pi_{1:l} > 0$. *Then*

$$\tau_1(\pi, S) = \max\left\{ \tau_1(\pi_{1:l}, A), \|B\|_\infty \right\}.$$

Proof. This follows from a more general result for real matrices in Theorem 7.9. $\square$

The explicit expression below for $\tau_\infty(\pi, S) = \max_{\|z\|_\infty = 1, z^T\pi = 0} \left\| S^T z \right\|_\infty$ holds for reducible as well as irreducible matrices.

Theorem 6.5 (Theorem 1 in [77]). *Let* $S \in \mathbb{R}^{n\times n}$ *be a stochastic reducible matrix of the form (6.1) with stationary distribution* $P\pi = \begin{pmatrix} \pi_{1:k}^T & 0 \end{pmatrix}^T$ *where* $\pi_{1:k}^T > 0$. *Then*

$$\tau_\infty(\pi, S) = \max_{1\le i\le n} \phi(Se_i),$$

where the function ϕ *is defined for* $s \in \mathbb{R}^n$ *with elements labeled* $s_1/\pi_1 \ge \ldots \ge s_k/\pi_k$, l *is the smallest integer such that* $\sum_{j=1}^{l} \pi_j \ge \frac{1}{2}$, *and*

$$\phi(s) = \sum_{i=1}^{l-1} s_i + \left(\sum_{j=l+1}^{k} \pi_j - \sum_{j=1}^{l-1} \pi_j \right) \frac{s_l}{|\pi_l|} - \sum_{i=l+1}^{k} s_i + \sum_{i=k+1}^{n} s_i .$$

Proof. This follows from a more general result for real matrices in Theorem 7.10. $\square$

We can view $\tau_\infty(\pi, S)$ as the norm of an (oblique) projection of S, with the projection being onto $\text{range}(\pi)^\perp$.

Corollary 6.6. *Let the assumptions of Theorem 6.5 hold. Then for some* $1 \le k \le n$

$$\tau_\infty(\pi, S) = \left\| S^T \left(I - D^{-1} e_k \pi^T \right) \right\|_\infty .$$

where $D = P^T \text{diag}\left(\text{diag}(\pi_{1:k}), I_{n-k}\right) P$.

Proof. The proof is analogous to that of Corollary 4.3. □

With regard to an explicit expression for the two-norm coefficient $\tau_2(\pi, S) = \max_{\|z\|_2=1, z^T\pi=0} \left\|S^T z\right\|_2$, the idea is to incorporate the constraint $z^T\pi = 0$ into the matrix. Then the constrained maximization problem of order n can be reduced to an unconstrained problem of order $n-1$.

Let $S \in \mathbb{R}^{n\times n}$ be a stochastic matrix of the form (6.1) and distinguish the leading row of the permuted matrix

$$PS = \begin{pmatrix} s_1^T \\ S_{2:n}^T \end{pmatrix}.$$

In the expression for $\tau_2(\pi, S)$ below we replace the $n \times n$ matrix S by a $n \times (n-1)$ matrix.

THEOREM 6.7 (Theorem 2 in [77]). *Let $S \in \mathbb{R}^{n\times n}$ be a stochastic matrix of the form (6.1) with stationary distribution $P\pi = \begin{pmatrix}\pi_{1:m}^T & 0\end{pmatrix}^T$ where $\pi_{1:m} > 0$. Then*

$$\tau_2(\pi, S) = \left\|RL^{-1}\right\|_2,$$

where $R = S_{2:n} - s_1\pi_{2:n}^T/\pi_1$ and $L^TL = I_{n-1} + \pi_{2:n}\pi_{2:n}^T/\pi_1^2$.

Proof. Because the two-norm corresponds to a quadratic form, we can use the constraint $z^T\pi = 0$ to eliminate an element from z and reduce the dimension of the maximization problem $\tau_2(\pi, S) = \max_{z^T\pi=0, \|z\|_2=1} \|S^Tz\|_2$.

Let $z \in \mathbb{R}^n$ be a vector with $z^T\pi = 0$ and $\|z\|_2 = 1$, and partition $Pz = \begin{pmatrix}z_1 & z_{2:n}^T\end{pmatrix}^T$. From $0 = z^T\pi = (Pz)^T(P\pi) = z_1\pi_1 + z_{2:n}^T\pi_{2:n}$ and $\pi_1 > 0$ follows

$$z_1 = -z_{2:n}^T\pi_{2:n}/\pi_1.$$

Substituting this expression for z_1 into S^Tz gives $PS^Tz = s_1z_1 + S_{2:n}z_{2:n} = Rz_{2:n}$, hence $\|S^Tz\|_2 = \|Rz_{2:n}\|_2$. Furthermore, substituting the expression for z_1 into the other constraint $1 = \|z\|_2^2 = (Pz)^T(Pz)$ gives $z_{2:n}^TQz_{2:n} = 1$ where $Q = I_{n-1} + \pi_{2:n}\pi_{2:n}^T/\pi_1^2$. The matrix Q is real symmetric positive definite, and has a Cholesky factorization $Q = L^TL$, so that $\|Lz_{2:n}\|_2 = 1$.

Thus the problem of maximizing $\|S^Tz\|_2$ for $z \in \mathbb{R}^n$ subject to $z^T\pi = 0$ and $\|z\|_2 = 1$ is equivalent to maximizing $\|Rx\|_2$ for $x \in \mathbb{R}^{n-1}$ subject to $\|Lx\|_2 = 1$. At last, since L is nonsingular we can set $y = Lx$, so that the maximization problem becomes

$$\tau_2(\pi, S) = \max_{\|y\|_2=1} \|RL^{-1}y\|_2 = \|RL^{-1}\|_2.$$

□

One can further simplify the expression in Theorem 6.7, and show that $\tau_2(\pi, S)$ is simply the norm of an orthogonal projection of S onto $\text{range}(\pi)^\perp$. The resulting expression below is a special case of Theorems 7.15 and 8.6.

COROLLARY 6.8. *If $S \in \mathbb{R}^{n\times n}$ is a stochastic matrix with $\pi^TS = \pi^T$ and $\pi \neq 0$ then*

$$\tau_2(\pi, S) = \left\|\left(I - \frac{\pi\pi^T}{\|\pi\|_2^2}\right)S\right\|_2.$$

Proof. Theorem 6.7 implies $\|RL^{-1}\|_2^2 = \left\| R\left(I + \pi_{2:n}\pi_{2:n}^T/\pi_1^2\right)^{-1} R^T \right\|_2$. From $\left(I + \pi_{2:n}\pi_{2:n}^T/\pi_1^2\right)^{-1} = I - \pi_{2:n}\pi_{2:n}^T/\|\pi\|_2^2$ follows

$$\|RL^{-1}\|_2^2 = \left\| S^T S - \frac{\pi\pi^T}{\|\pi\|_2^2} \right\|_2 = \left\| S^T \left(I - \frac{\pi\pi^T}{\|\pi\|_2^2} \right) S \right\|_2 = \left\| \left(I - \frac{\pi\pi^T}{\|\pi\|_2^2} \right) S \right\|_2^2,$$

where the last equality follows from the fact that $I - \pi\pi^T/\|\pi\|_2^2$ is Hermitian and idempotent. □

7. Ergodicity Coefficients For Real Matrices.

In 1984 Seneta [69, (1)] extended the coefficient of ergodicity from stochastic matrices to rectangular matrices $A \in \mathbb{R}^{m\times n}$ and arbitrary vectors $w \in \mathbb{R}^m$,

$$\tau_p(w, A) = \max_{\substack{\|z\|_p=1 \\ z^T w=0}} \left\| A^T z \right\|_p,$$

where the maximum ranges over $z \in \mathbb{R}^m$.

We present properties and explicit expressions, as well as applications to eigenvalue bounds for nonnegative matrices.

7.1. Properties and Bounds Common to All p-Norm Coefficients.

We present properties of $\tau_p(w, A)$ for rectangular matrices A. The coefficients $\tau_p(w, A)$ are bounded, well-conditioned in the second argument, and only very weakly submultiplicative because w is generally not an eigenvector of A.

THEOREM 7.1. *If $A, A_1, A_2 \in \mathbb{R}^{m\times n}$ and $w \in \mathbb{R}^m$ then*
1. $0 \le \tau_p(w, A) \le \|A^T\|_p$
2. $|\tau_p(w, A_1) - \tau_p(w, A_2)| \le \tau_p(w, A_1 - A_2)$
3. $\tau_p(w, AB) \le \|B^T\|_p \, \tau_p(w, A)$ *for* $B \in \mathbb{R}^{n\times k}$.

Proof.
1. This follows from $\max_{\substack{\|z\|_p=1 \\ z^T w=0}} \left\| A^T z \right\|_p \le \max_{\|z\|_p=1} \left\| A^T z \right\|_p = \left\| A^T \right\|_p$.
2. Let $\tau_p(w, A_1) \ge \tau_p(w, A_2)$, and $\tau_p(w, A_1) = \left\| A_1^T y \right\|_p$ for some vector $y \in \mathbb{R}^m$ with $\|y\|_p = 1$ and $y^T w = 0$. Then

$$0 \le \tau_p(w, A_1) - \tau_p(w, A_2) = \left\| A_1^T y \right\|_p - \max_{\substack{\|z\|_p=1 \\ z^T w=0}} \left\| A_2^T z \right\|_p \le \left\| A_1^T y \right\|_p - \left\| A_2^T y \right\|_p.$$

The triangle inequality implies

$$\begin{aligned} \left\| A_1^T y \right\|_p - \left\| A_2^T y \right\|_p &\le \left\| (A_1 - A_2)^T y \right\|_p \le \max_{\substack{\|z\|_p=1 \\ z^T w=0}} \left\| (A_1 - A_2)^T z \right\|_p \\ &= \tau_p(w, A_1 - A_2). \end{aligned}$$

3. Let $y \in \mathbb{R}^m$ be a vector with $\tau_p(w, BA) = \left\| (BA)^T y \right\|_p$, $y^T w = 0$ and $\|y\|_p = 1$. The submultiplicative property of the p-norms implies

$$\begin{aligned} \tau_p(w, BA) &= \|A^T B^T y\|_p \le \|A^T\|_p \|B^T y\|_p \\ &\le \|A^T\|_p \max_{\substack{\|z\|_p=1 \\ z^T w=0}} \|B^T z\|_p = \|A^T\|_p \, \tau_p(w, B). \end{aligned}$$

□

If w happens to be a real eigenvector of A then a submultiplicative property holds for powers of A, similar to the one for stochastic matrices in Theorem 6.1.

THEOREM 7.2. *Let $A \in \mathbb{R}^{n\times n}$ and $w \in \mathbb{R}^n$ be a right eigenvector of A. Then for $l, m \geq 1$*

$$\tau_p\left(w, A^{l+m}\right) \leq \tau_p\left(w, A^l\right) \tau_p\left(w, A^m\right).$$

Proof. Let $y \in \mathbb{R}^n$ be a vector with $\tau_p\left(w, A^{l+m}\right) = \left\|\left(A^T\right)^{l+m} y\right\|_p$, $y^T w = 0$ and $\|y\|_p = 1$. Since $Aw = \lambda w$ for some real number λ, we have

$$\left[(A^T)^m y\right]^T w = y^T A^m w = \lambda^m y^T w = 0.$$

Hence the vector $x = \left(A^T\right)^m y / \left\|\left(A^T\right)^m y\right\|_p$ satisfies $\|x\|_p = 1$ and $x^T w = 0$, so that

$$\tau_p\left(w, A^{l+m}\right) = \left\|\left(A^T\right)^{l+m} y\right\|_p = \left\|\left(A^T\right)^l x\right\|_p \left\|\left(A^T\right)^m y\right\|_p \leq \tau_p\left(w, A^l\right) \tau_p\left(w, A^m\right).$$

□

Bounds. We present two upper bounds for $\tau_p(w, A)$ that could possibly improve the bound in Theorem 7.1. They also furnish eigenvalue bounds for irreducible non-negative matrices in §7.5.

The first bound expresses the constraint $z^T w = 0$ in terms of a rank-one downdate of the matrix.

THEOREM 7.3 (Theorem 5.5 in [60]). *If $A \in \mathbb{R}^{m\times n}$ and $w \in \mathbb{R}^m$ with $w \neq 0$ then for all vectors $x \in \mathbb{R}^n$*

$$\tau_p(w, A) \leq \left\|\left(A - wx^T\right)^T\right\|_p.$$

In particular for $p = 2$

$$\tau_2(w, A) \leq \left\|A - wx^T\right\|_2 \leq \left\|A - wx^T\right\|_F,$$

where $\min_x \|A - wx^T\|_F = \left\|\left(I - \frac{ww^T}{\|w\|_2^2}\right) A\right\|_F$.

Proof. Let $z \in \mathbb{R}^m$ be a vector with $z^T w = 0$ and $\|z\|_p = 1$. Then

$$\left(A - wx^T\right)^T z = A^T z - x\, w^T z = A^T z$$

implies

$$\tau_p(w, A) = \max_{\substack{\|z\|_p=1 \\ z^T w=0}} \left\|A^T z\right\|_p = \max_{\substack{\|z\|_p=1 \\ z^T w=0}} \left\|\left(A - wx^T\right)^T z\right\|_p \leq \left\|\left(A - wx^T\right)^T\right\|_p,$$

where the last inequality follows from Theorem 7.1.

From $\|A^T\|_2 = \|A\|_2$ and $\|A\|_2 \leq \|A\|_F$ for any matrix A follows

$$\left\|\left(A - wx^T\right)^T\right\|_2 = \left\|A - wx^T\right\|_2 \leq \left\|A - wx^T\right\|_F.$$

To find a vector x that minimizes $\left\|A - wx^T\right\|_F$, write the Frobenius norm as a sum of two-norms,

$$\left\|A - wx^T\right\|_F^2 = \sum_{i=1}^{n} \left\|\left(A - wx^T\right) e_i\right\|_2^2 = \sum_{i=1}^{n} \|Ae_i - wx_i\|_2^2 .$$

Thus $\min_x \left\|A - wx^T\right\|_F^2$ consists of n independent minimization problems $\|Ae_i - wx_i\|_2$. Each minimization problem $\|wx_i - Ae_i\|_2$ is a least squares problem with $m \times 1$ coefficient matrix w of full column rank and right hand side Ae_i. The unique solution is $x_i = \left(w^T w\right)^{-1} w^T Ae_i = e_i^T A^T w/\|w\|_2^2$. Thus $x = A^T w/\|w\|_2^2$, and

$$A - wx^T = A - \frac{ww^T A}{\|w\|_2^2} = \left(I - \frac{ww^T}{\|w\|_2^2}\right) A.$$

□

The second bound relates coefficients based on different vectors. We will use it to show continuity of $\tau_p(w, A)$ with regard to w.

Theorem 7.4 (Lemma 4.1 in [24]). *Let $B \in \mathbb{R}^{m\times n}$ and $f \in \mathbb{R}^m$. Then for every pair of vectors $v, w \in \mathbb{R}^m$ with $f^T v = f^T w = 1$*

$$\tau_p(w, A) \le \tau_p(v, A) + 2\left\|A^T\right\|_p \|f\|_p\|v - w\|_q,$$

where $1/p + 1/q = 1$.

Proof. If $\tau_p(w, A) = 0$ the inequality holds trivially. Now assume that $y \in \mathbb{R}^m$ is a vector such that $y^T w = 0$, $\|y\|_p = 1$, and $\tau_p(w, A) = \left\|A^T y\right\|_p > 0$. We will bound $\tau_p(v, A)$ from below in terms of a projection of y, namely the vector $(I_m - fv^T)y = y - v^T y\, f$.

If $y - v^T y\, f = 0$, then $y = v^T yf$. This, together with $w^T f = 1$ implies $w^T y = v^T y\, w^T f = v^T y$. Hence $0 = w^T y = v^T y$. From $v^T y = 0$ and $\|y\|_p = 1$ follows

$$\tau_p(v, A) = \max_{\substack{\|z\|_p=1\\ z^T v=0}} \left\|A^T z\right\|_p \ge \|A^T y\|_p = \tau(w, A),$$

so that the desired inequality holds.

If $y - v^T y\, f \ne 0$ we can define the vector

$$z = \frac{y - v^T y\, f}{\|y - v^T y\, f\|_p} = \frac{y - (v - w)^T y\, f}{\|y - (v - w)^T y\, f\|_p},$$

which satisfies $\|z\|_p = 1$ and $z^T v = 0$, so that

$$\tau_p(v, A) \ge \left\|A^T z\right\|_p = \frac{\left\|A^T y - (v - w)^T\, y\, A^T f\right\|_p}{\left\|y - (v - w)^T\, y\, f\right\|_p}.$$

The triangle and Hölder inequalities imply

$$\|A^T z\|_p \ge \frac{\left\|A^T y\right\|_p - \|v - w\|_q \|y\|_p \left\|A^T\right\|_p \|f\|_p}{\|y\|_p + \|v - w\|_q \|y\|_p\|f\|_p}.$$

From $\tau_p(w,A) = \left\|A^T y\right\|_p$ and $\|y\|_p = 1$ follows

$$\tau_p(v,A) \geq \frac{\tau_p(w,A) - \|v-w\|_q \left\|A^T\right\|_p \|f\|_p}{1 + \|v-w\|_q \|f\|_p}.$$

Rearranging gives

$$\tau_p(w,A) \leq \tau_p(v,A)\left[1 + \|v-w\|_q \|f\|_p\right] + \|v-w\|_q \left\|A^T\right\|_p \|f\|_p .$$

The desired inequality now follows from $\tau_p(v,A) \leq \left\|A^T\right\|_p$ in Theorem 7.1. ☐

Theorem 7.4 implies that $\tau_p(w,A)$ is a continuous function of the first argument on the set $\left\{x \in \mathbb{R}^n : \ x^T f = 1\right\}$.

Corollary 7.5 (Corollary 4.2 in [24]). *Let $A \in \mathbb{R}^{m\times n}$ and $f \in \mathbb{R}^m$. Then for every pair of vectors $v, w \in \mathbb{R}^m$ with $f^T v = f^T w = 1$*

$$|\tau_p(w,A) - \tau_p(v,A)| \leq 2\left\|A^T\right\|_p \|f\|_p \|v-w\|_q,$$

where $1/p + 1/q = 1$.

7.2. Explicit Expressions for One-Norm Coefficients. We derive an explicit expression for $\tau_1(w,A)$ for real rectangular matrices A.

We start with an easy case, that of square matrices with constant row sum and $w = \mathbb{1}$, since this scenario is similar to that of stochastic matrices.

Theorem 7.6 ([1, 54], p 584 in [66], pp 189 - 191 in [71]). *If $A \in \mathbb{R}^{n\times n}$ with $A\mathbb{1} = a\mathbb{1}$ then*

$$\tau_1(\mathbb{1},A) = \frac{1}{2}\max_{ij}\sum_{k=1}^{n}|a_{ik} - a_{jk}| = a - \min_{ij}\sum_{k=1}^{n}\min\left\{a_{ik}, a_{jk}\right\}.$$

Proof. This is an extension of the explicit expression for $\tau_1(S)$ for stochastic matrices S in Theorem 3.4 and Corollary 3.6, and a special case of Theorem 7.8 below. ☐

For general, real matrices A, we begin with the case where all elements of w are nonzero, i.e. $|w| > 0$. We view w as a diagonal scaling of the vector $\mathbb{1}$, and derive an expression for vectors x that satisfy $x^T w = 0$. This is done in the extension below of Lemma 3.3.

Lemma 7.7 (p 192 in [69]). *If $x, w \in \mathbb{R}^n$ with $x \neq 0$, $|w| > 0$ and $x^T w = 0$, then*

$$x = \sum_{i\neq j} y_{ij}\,\frac{D^{-1}(e_i - e_j)}{\left\|D^{-1}(e_i - e_j)\right\|_1}, \qquad \text{where} \quad y_{ij} \geq 0, \quad \sum_{i\neq j} y_{ij} = \|x\|_1,$$

and $D = \mathrm{diag}(w)$.

Proof. The proof proceeds by induction over the dimension n of x.

If $n = 2$ assume $w_1 x_1 > 0$. Then, $0 = x^T w = x^T(D\mathbb{1}) = (Dx)^T\,\mathbb{1}$. As in Lemma 3.3 this implies $Dx = w_1 x_1 \begin{pmatrix}1 & -1\end{pmatrix}^T$. Hence $x = w_1 x_1\, D^{-1}(e_1 - e_2)$. From $D^{-1}(e_1 - e_2) = \begin{pmatrix}1/w_1 & 1/w_2\end{pmatrix}^T$ and $\left\|D^{-1}(e_1 - e_2)\right\|_1 = |1/w_1| + |1/w_2|$ follows

$$x = y_{12}\,\frac{D^{-1}(e_1 - e_2)}{\left\|D^{-1}(e_1 - e_2)\right\|_1}, \qquad \text{where} \qquad y_{12} = w_1 x_1\left(\frac{1}{|w_1|} + \frac{1}{|w_2|}\right) > 0.$$

Finally $|w_1 x_1| = |w_2 x_2|$ implies

$$y_{12} = \left|\frac{w_1 x_1}{w_1}\right| + \left|\frac{w_1 x_1}{w_2}\right| = |x_1| + \left|\frac{w_2 x_2}{w_2}\right| = \|x\|_1.$$

Now assume the lemma holds for $n \geq 2$ and we will show it holds for $n+1$. Let $x \neq 0$ be a vector of dimension $n+1$ with $x^T w = 0$. Assume Dx has been permuted so that $w_n x_n > 0$, $w_{n+1} x_{n+1} < 0$ and $w_n x_n = \max_{1 \leq i \leq n+1} |w_i x_i|$. Then

$$Dx = \begin{pmatrix} (Dx)_{1:n-1} \\ w_n x_n \\ w_{n+1} x_{n+1} \end{pmatrix} = \begin{pmatrix} (Dx)_{1:n-1} \\ w_n x_n + w_{n+1} x_{n+1} \\ 0 \end{pmatrix} - w_{n+1} x_{n+1} \begin{pmatrix} 0_{1:n-1} \\ 1 \\ -1 \end{pmatrix}.$$

Multiplying by D^{-1} gives

$$x = \begin{pmatrix} x_{1:n-1} \\ x_n + \frac{w_{n+1} x_{n+1}}{w_n} \\ 0 \end{pmatrix} - w_{n+1} x_{n+1}\, D^{-1}(e_n - e_{n+1}).$$

Define the vector $\hat{x} = \begin{pmatrix} x_{1:n-1}^T & x_n + \frac{w_{n+1} x_{n+1}}{w_n} \end{pmatrix}^T$ of dimension n, which satisfies $\hat{x}^T w_{1:n} = x^T w = 0$. If $\hat{x} = 0$ then the conclusion follows as in the case $n = 2$. If $\hat{x} \neq 0$ then we apply the induction hypothesis to $\hat{x}$ and obtain

$$\hat{x} = \sum_{i \neq j} y_{ij} \, \frac{D_{1:n}^{-1}(e_i - e_j)}{\left\|D_{1:n}^{-1}(e_i - e_j)\right\|_1}, \qquad \text{where} \qquad y_{ij} \geq 0, \quad \sum_{i \neq j} y_{ij} = \|\hat{x}\|_1,$$

and $D_{1:n}$ is the leading principal submatrix of order n of D.

Applying the definition of $\hat{x}$ and setting $y_{n,n+1} = -w_{n+1} x_{n+1} \, \|D^{-1}(e_n - e_{n+1})\|_1$ gives

$$\begin{aligned} \sum_{i \neq j} y_{ij} + y_{n,n+1} &= \|\hat{x}\|_1 - w_{n+1} x_{n+1} \, \|D^{-1}(e_n - e_{n+1})\|_1 \\ &= \|x_{1:n-1}\|_1 + \left| x_n + \frac{w_{n+1} x_{n+1}}{w_n} \right| - w_{n+1} x_{n+1} \left(\frac{1}{|w_n|} + \frac{1}{|w_{n+1}|} \right). \end{aligned}$$

From $w_n x_n = \max_{1 \leq i \leq n+1} |w_i x_i| > 0$ and $w_{n+1} x_{n+1} < 0$ follows

$$\left| x_n + \frac{w_{n+1} x_{n+1}}{w_n} \right| = \frac{1}{|w_n|} \left(w_n x_n + w_{n+1} x_{n+1} \right).$$

Hence

$$\sum_{i \neq j} y_{ij} + y_{n,n+1} = \|x_{1:n-1}\|_1 + \frac{w_n x_n}{|w_n|} - \frac{w_{n+1} x_{n+1}}{|w_{n+1}|}.$$

From $w_n x_n > 0$ and $w_{n+1} x_{n+1} < 0$ follows

$$\frac{w_n x_n}{|w_n|} - \frac{w_{n+1} x_{n+1}}{|w_{n+1}|} = \frac{|w_n x_n|}{|w_n|} + \frac{|w_{n+1} x_{n+1}|}{|w_{n+1}|} = |x_n| + |x_{n+1}|.$$

Therefore $\sum_{i \neq j} y_{ij} + y_{n,n+1} = \|x\|_1$. At last, the definition of $y_{n,n+1}$ yields the desired expression for x,

$$\begin{aligned} x &= \begin{pmatrix} \hat{x} \\ 0 \end{pmatrix} - w_{n+1} x_{n+1} \, D^{-1}(e_n - e_{n+1}) \\ &= \sum_{i \neq j} y_{ij} \, \frac{D^{-1}(e_i - e_j)}{\|D^{-1}(e_i - e_j)\|_1} + y_{n,n+1} \, \frac{D^{-1}(e_n - e_{n+1})}{\|D^{-1}(e_n - e_{n+1})\|_1}. \end{aligned}$$

□

When $w = \mathbb{1}$, then $D = I$ and $\|D^{-1}(e_i - e_j)\|_1 = 2$, so that Lemma 7.7 reduces to Lemma 3.3.

As in Theorem 3.4, we make use of Lemma 7.7 to determine an explicit expression for $\tau_1(w, A)$ for real matrices A and real vectors w. We distinguish the two cases when all elements of w are nonzero, and when some elements of w can be zero.

The expression below, for w with all elements nonzero, extends Theorems 3.4 and 6.2 from stochastic to real matrices.

THEOREM 7.8 (p 193 in [69]). *If $A \in \mathbb{R}^{m\times n}$, $w \in \mathbb{R}^m$ and $|w| > 0$, then*

$$\tau_1(w, A) = \max_{ij} \frac{\left\|A^T D^{-1}(e_i - e_j)\right\|_1}{\|D^{-1}(e_i - e_j)\|_1}.$$

Proof. Let $\tau_1(w, A) = \|A^T x\|_1$ where $x^T w = 0$ and $\|x\|_1 = 1$. Applying Lemma 7.7 to x gives $x = \sum_{i\neq j} y_{ij}\, D^{-1}(e_i - e_j)/\left\|D^{-1}(e_i - e_j)\right\|_1$, where $\sum_{i\neq j} y_{ij} = \|x\|_1 = 1$ and $D = \mathrm{diag}(w)$. The triangle inequality applied to $\|A^T x\|_1$ yields

$$\tau_1(w, A) = \left\|A^T x\right\|_1 \leq \sum_{i\neq j} y_{ij} \frac{\|A^T D^{-1}(e_i - e_j)\|_1}{\|D^{-1}(e_i - e_j)\|_1} \leq \max_{ij} \frac{\|A^T D^{-1}(e_i - e_j)\|_1}{\|D^{-1}(e_i - e_j)\|_1}.$$

To show the reverse inequality, set $y = D^{-1}(e_i - e_j)/\left\|D^{-1}(e_i - e_j)\right\|_1$ for some $i \neq j$. Then $y^T w = y^T D\mathbb{1} = 0$, $\|y\|_1 = 1$, and

$$\tau_1(w, A) = \max_{\substack{\|z\|_1 = 0 \\ z^T w = 0}} \|A^T z\|_1 \geq \|A^T y\|_1 = \frac{\left\|A^T D^{-1}(e_i - e_j)\right\|_1}{\|D^{-1}(e_i - e_j)\|_1}.$$

Since this inequality holds for any i and j,

$$\tau_1(w, A) \geq \max_{ij} \frac{\|A^T D^{-1}(e_i - e_j)\|_1}{\|D^{-1}(e_i - e_j)\|_1}.$$

□

If A is a stochastic matrix and $w = \mathbb{1}$, then Theorem 7.8 reduces to Theorem 3.4. If A is also irreducible with stationary distribution π and if $w = \pi$, then Theorem 7.8 reduces to Theorem 6.2.

Now we consider the more general situation when w can have zero elements. We choose a permutation matrix P to isolate the nonzero elements in w, and permute the rows of A correspondingly,

$$Pw = \begin{pmatrix} w_{1:k} \\ 0 \end{pmatrix}, \qquad PA = \begin{pmatrix} A_k \\ A_{m-k} \end{pmatrix},$$

where $w_{1:k}$ is a $k \times 1$ vector with $|w_{1:k}| > 0$, and A_k has k rows. The following expression extends Theorem 6.4 from stochastic to real matrices.

THEOREM 7.9 (p 194 in [69]). *Let $A \in \mathbb{R}^{m\times n}$, $w \in \mathbb{R}^m$, and P a permutation matrix so that $|w_{1:k}| > 0$. Then*

$$\tau_1(w, A) = \max\left\{\tau_1(w_{1:k}, A_k),\ \|A_{m-k}\|_\infty\right\}.$$

Proof. Let $x \in \mathbb{R}^m$ be a vector with $\tau_1(w, A) = \|A^T x\|_1$, $\|x\|_1 = 1$, and $x^T w = 0$. Partitioning $Px = \begin{pmatrix} x_k^T & x_{m-k}^T \end{pmatrix}^T$ with x_k being $k \times 1$ gives

$$A^T x = A^T P^T P x = A_k^T x_k + A_{m-k}^T x_{m-k}.$$

We distinguish the cases $x_k = 0$ and $x_k \neq 0$.

If $x_k = 0$, then $\left\|A^T x\right\|_1 \leq \left\|A_{m-k}^T\right\|_1 \|x_{m-k}\|_1$ and $\|x_{m-k}\|_1 = \|x\|_1 = 1$. Hence $\tau_1(w, A) = \|A^T x\|_1 \leq \left\|A_{m-k}^T\right\|_1 = \|A_{m-k}\|_\infty$. To show the reverse inequality, choose x such that $Px = e_{k+i}$ for some $1 \leq i \leq m-k$ and $\|A_{m-k}^T\|_1 = \|A_{m-k}^T e_i\|_1 = \|A^T x\|_1$. Since the trailing $m-k$ elements of Pw are zero, $x^T w = 0$. This, together with $\|x\|_1 = 1$ implies $\|A_{m-k}^T\|_1 = \|A^T x\|_1 \leq \tau_1(w, A)$.

If $x_k \neq 0$ then $0 = x^T w = x^T P^T P w = x_k^T w_{1:k}$, and we can apply Lemma 7.7 to obtain

$$x_k = \sum_{i \neq j} y_{ij} \, \frac{D_k^{-1}(e_i - e_j)}{\left\|D_k^{-1}(e_i - e_j)\right\|_1}, \qquad \text{where} \quad y_{ij} \geq 0, \quad \sum_{i \neq j} y_{ij} = \|x_k\|_1,$$

and $D_k = \operatorname{diag}(w_{1:k})$. Substituting this into the above expression for $\|A^T x\|_1$ gives

$$\begin{aligned}
\left\|A^T x\right\|_1 &\leq \|x_k\|_1 \, \max_{ij} \frac{\left\|A_k^T D_k^{-1}(e_i - e_j)\right\|_1}{\left\|D_k^{-1}(e_i - e_j)\right\|_1} + \left\|A_{m-k}^T\right\|_1 \|x_{m-k}\|_1 \\
&\leq \left(\|x_k\|_1 + \|x_{m-k}\|_1\right) \, \max\left\{ \max_{ij} \frac{\left\|A_k^T D_k^{-1}(e_i - e_j)\right\|_1}{\left\|D_k^{-1}(e_i - e_j)\right\|_1}, \; \left\|A_{m-k}^T\right\|_1 \right\}.
\end{aligned}$$

Now Theorem 7.8 and $\|x_k\|_1 + \|x_{m-k}\|_1 = \|x\|_1 = 1$ imply

$$\tau_1(w, A) \leq \max\left\{\tau_1(w_{1:k}, A_k), \; \|A_{m-k}\|_\infty\right\}.$$

The reverse inequality follows, as in the proof of Theorem 7.8, by picking a vector y whose leading k elements are $P^T D_k^{-1}(e_i - e_j)/\left\|D_k^{-1}(e_i - e_j)\right\|_1$ for some $1 \leq i, j \leq k$, and whose trailing $n-k$ elements are zero. ☐

If A is a stochastic reducible matrix with stationary distribution π then Theorem 7.9 with $w = \pi$ reduces to Theorem 6.4.

7.3. Explicit Expressions for Infinity-Norm Coefficients. We extend Theorems 4.2 and 6.5 from stochastic matrices to real matrices.

THEOREM 7.10 (§3 in [69]). *If $A \in \mathbb{R}^{m \times n}$ and $w \in \mathbb{R}^m$, and P a permutation matrix so that $Pw = \begin{pmatrix} w_{1:k}^T & 0 \end{pmatrix}^T$ with $|w_{1:k}| > 0$, then*

$$\tau_\infty(w, A) = \max_{1 \leq i \leq n} \phi(Ae_i),$$

where the function ϕ is defined for $a \in \mathbb{R}^n$ with elements labeled $a_1/w_1 \geq \ldots \geq a_k/w_k$, l is the smallest integer such that $\sum_{j=1}^{l} |w_j| \geq \sum_{j=l+1}^{k} |w_j|$, and

$$\phi(a) = \sum_{i=1}^{l-1} \frac{|w_i|}{w_i} a_i + \left(\sum_{j=l+1}^{k} |w_j| - \sum_{j=1}^{l-1} |w_j| \right) \frac{a_l}{|w_l|} - \sum_{i=l+1}^{k} \frac{|w_i|}{w_i} a_i + \sum_{i=k+1}^{n} |a_i|.$$

Proof. We start as in the proof of Theorem 4.2. Let a be a column of A and x a vector that together achieve the maximum in $\tau_\infty(w,A)$, i.e.

$$\tau_\infty(w,A) = \max_{\substack{\|z\|_\infty=1\\ z^Tw=0}} \left\|A^Tz\right\|_\infty = \max_{\substack{\|z\|_\infty=1\\ z^Tw=0}} \max_j \left|e_j^TA^Tz\right| = |a^Tx|.$$

The idea is to incorporate the constraint $x^Tw=0$ into the matrix vector product, by writing $x^Ta = x^T(a-\theta w)$ for any scalar θ. If P is a permutation matrix then the Hölder inequality and $\|x\|_\infty = 1$ imply

$$|x^Ta| = |(Px)^T(Pa-\theta Pw)| \le \|Px\|_\infty \|Pa-\theta Pw\|_1 = \|Pa-\theta Pw\|_1.$$

Choose a permutation matrix P so that $Pw = \begin{pmatrix} w_{1:k}^T & 0\end{pmatrix}^T$ with $|w_{1:k}|>0$, and $Pa = \begin{pmatrix} a_1 & \dots & a_n\end{pmatrix}^T$ with $a_1/w_1 \ge \ldots \ge a_k/w_k$. Then

$$\|Pa-\theta Pw\|_1 = \sum_{i=1}^k |w_i| \left|\frac{a_i}{w_i}-\theta\right| + \sum_{i=k+1}^n |a_i|.$$

Set $\theta = a_l/w_l$, split the sum and remove absolute values,

$$\left\|Pa - \frac{a_l}{w_l}\,Pw\right\|_1 = \sum_{i=1}^l |w_i| \left(\frac{a_i}{w_i}-\frac{a_l}{w_l}\right) + \sum_{i=l+1}^k |w_i| \left(\frac{a_l}{w_l}-\frac{a_i}{w_i}\right) + \sum_{i=k+1}^n |a_i| = \phi(a).$$

We have shown that $\tau_\infty(w,A) = |x^Ta| \le \phi(a) \le \max_{1\le i\le n}\phi(Ae_i)$.

To show the reverse inequality, let a be a column of A so that $\max_{1\le i\le n}\phi(Ae_i) = \phi(a)$. Define the vector y with elements

$$y_i = \begin{cases} \frac{-|w_i|}{w_i} & 1\le i\le l-1\\ \frac{1}{w_l}\left(\sum_{i=1}^{l-1}|w_i| - \sum_{i=l+1}^k |w_i|\right) & i=l\\ \frac{|w_i|}{w_i} & l+1\le i\le k\\ \operatorname{sign}(a_i) & k+1\le i\le n.\end{cases}$$

Then $\phi(s) = |y^Ts|$ and $y^T\mathbb{1} = 0$. Clearly, $|y_i|=1$ for $i\ne l$. Since l is the smallest integer such that $\sum_{j=1}^l |w_j| \ge \sum_{j=l+1}^k |w_j|$, this implies $\sum_{j=1}^{l-1}|w_j| \le \sum_{j=l}^k |w_j|$. From these two inequalities follows $|y_l|\le 1$. Thus $\|y\|_\infty = 1$ so that $\max_{1\le i\le n}\phi(Ae_i) = \phi(a) = |y^Ts| \le \tau_\infty(w,A)$. □

If A is a stochastic matrix and $w=\mathbb{1}$, then Theorem 7.10 reduces to Theorem 4.2. If A has stationary π and $w=\pi$ then Theorem 7.10 reduces to Theorem 6.5.

We can view $\tau_\infty(w,A)$ as the norm of an (oblique) projection of A, with the projection being onto $\operatorname{range}(w)^\perp$. This is an extension of Corollaries 4.3 and 6.6 from stochastic matrices to real matrices.

Corollary 7.11. *Let the assumptions of Theorem 7.10 hold. Then for some* $1\le k\le n$

$$\tau_\infty(w,A) = \left\|A^T\left(I - D^{-1}e_kw^T\right)\right\|_\infty.$$

where $D = P^T\operatorname{diag}\left(\operatorname{diag}(w_{1:k}), I_{n-k}\right)P$.

Proof. The proof is analogous to that of Corollary 4.3. □

7.4. Explicit Expressions for Two-Norm Coefficients. We derive four different expressions for $\tau_2(w, A)$ and extend results in [66, 74, 77] for stochastic matrices to real rectangular matrices.

In the first expression we represent $\tau_2(w, A)$ as the norm of a matrix with one row less. The expression is similar in spirit to the one for stochastic matrices in Theorem 6.7.

THEOREM 7.12 (First Expression). *Let $A \in \mathbb{R}^{m\times n}$, and $w \in \mathbb{R}^m$ with $w \neq 0$. Let $Q \in \mathbb{R}^{m\times m}$ be an orthogonal matrix with leading column $Qe_1 = w/\|w\|_2$, and partition $A^T Q = \begin{pmatrix} a & A_{m-1}^T \end{pmatrix}$, where A_{m-1} has $m-1$ rows. Then $\tau_2(w, A) = \|A_{m-1}\|_2$.*

Proof. Let $z \in \mathbb{R}^m$ be a vector with $\|z\|_2 = 1$ and $z^T w = 0$. Because the first column of Q is a multiple of w, $Q^T z = \begin{pmatrix} 0 & \hat{z}^T \end{pmatrix}^T$, where $\hat{z}$ has $m-1$ elements, $\|\hat{z}\|_2 = 1$, and

$$A^T z = A^T Q Q^T z = \begin{pmatrix} a & A_{m-1}^T \end{pmatrix} \begin{pmatrix} 0 \\ \hat{z} \end{pmatrix} = A_{m-1}^T \hat{z}.$$

To obtain the expression for $\tau_2(w, A)$, we take the maximum,

$$\tau_2(w, A) = \max_{\substack{\|z\|_2=1 \\ z^T w=0}} \|A^T z\|_2 = \max_{\|\hat{z}\|_2=1} \|A_{m-1}^T \hat{z}\|_2 = \|A_{m-1}^T\|_2 = \|A_{m-1}\|_2.$$

□

With the help of Theorem 7.12 we represent a vector y that achieves the maximum for $\tau_2(w, A)$ as the solution of a linear system with right-hand side w.

THEOREM 7.13 (Second Expression). *In addition to the conditions of Theorem 7.12, let $\tau \equiv \tau_2(w, A) = \|A^T y\|_2$ where $\|y\|_2 = 1$ and $y^T w = 0$. Then*

$$(AA^T - \tau^2 I) y = \gamma\, w, \qquad \text{where} \quad \gamma \equiv \frac{w^T A A^T y}{\|w\|_2^2}.$$

Proof. We represent the singular value problem $\|A_{m-1}\|_2$ from the proof of Theorem 7.12 as an eigenvalue problem,

$$Q^T A A^T Q = \begin{pmatrix} a^T a & a^T A_{m-1}^T \\ A_{m-1} a & B \end{pmatrix}, \qquad \text{where} \quad B \equiv A_{m-1} A_{m-1}^T.$$

Since $Q^T A A^T Q$ is real symmetric positive semi-definite, so is its leading principal submatrix B. Thus, $\tau^2 = \|A_{m-1}\|_2^2 = \|B\|_2^2$ is a dominant eigenvalue of B, and

$$\tau^2 = y^T A A^T y = y^T Q Q^T A A^T Q Q^T y = \hat{y}^T B \hat{y}, \qquad \text{where} \quad Q^T y = \begin{pmatrix} 0 \\ \hat{y} \end{pmatrix}.$$

This means that the $(m-1)\times 1$ vector $\hat{y}$ is a unit-norm eigenvector of B for the dominant eigenvalue τ^2, i.e. $(B - \tau^2 I)\hat{y} = 0$. Therefore

$$\begin{aligned}(AA^T - \tau^2 I) y &= Q \left[Q^T A A^T Q - \tau^2 I \right] Q^T y = Q \begin{pmatrix} a^T a - \tau^2 & a^T A_{m-1}^T \\ A_{m-1} a & B - \tau^2 I \end{pmatrix} \begin{pmatrix} 0 \\ \hat{y} \end{pmatrix} \\ &= Q \begin{pmatrix} a^T A_{m-1}^T \hat{y} \\ (B - \tau^2 I)\hat{y} \end{pmatrix} = Q \begin{pmatrix} a^T A_{m-1}^T \hat{y} \\ 0 \end{pmatrix} = a^T A_{m-1} \hat{y}\, Q e_1.\end{aligned}$$

It remains to express the last quantity in terms of A, y, and w. From $A^TQ = \begin{pmatrix} a & A_{m-1}^T \end{pmatrix}$, $Q^Ty = \begin{pmatrix} 0 \\ \hat{y} \end{pmatrix}$, and $Qe_1 = w/\|w\|_2$ follows

$$\begin{aligned} a^T A_{m-1}^T \hat{y} &= \begin{pmatrix} a^Ta & a^TA_{m-1}^T \end{pmatrix} \begin{pmatrix} 0 \\ \hat{y} \end{pmatrix} = a^T \begin{pmatrix} a & A_{m-1}^T \end{pmatrix} \begin{pmatrix} 0 \\ \hat{y} \end{pmatrix} \\ &= e_1^T Q^T AA^T QQ^T y = e_1^T Q^T AA^T y = w^T AA^T y/\|w\|_2. \end{aligned}$$

Hence

$$a^T A_{m-1}^T \hat{y}\, Qe_1 = \frac{w^T AA^T y}{\|w\|_2^2} w.$$

□

Theorem 7.13 in turn leads to a third characterization which was introduced by Seneta and Tan [74, Proposition 1]. We replace the Lagrange multiplier based proof by one that is matrix based. The characterization involves the adjugate, which is defined by $A\,\mathrm{adj}[A] = \det(A)I$ for square matrices A.

THEOREM 7.14 (Third Expression). *Let $A \in \mathbb{R}^{m\times n}$, and $w \in \mathbb{R}^m$ with $w \neq 0$. Then $(\tau_2(w, A))^2$ is the largest root of the polynomial $w^T\mathrm{adj}[AA^T - \lambda I]w$ of degree $m-1$ in λ.*

Proof. As in Theorem 7.13, let $\tau \equiv \tau_2(w, A) = \|A^Ty\|_2$ where $\|y\|_2 = 1$ and $y^Tw = 0$. We distinguish two cases, depending on whether τ^2 is an eigenvalue of AA^T or not.

- If τ^2 is not an eigenvalue of AA^T then $AA^T - \tau^2 I$ is nonsingular. From Theorem 7.13, and the relation between inverse and adjugate follows

  $$y = \gamma\,(AA^T - \tau^2 I)^{-1} w = \gamma\, \frac{\mathrm{adj}[AA^T - \tau^2 I]w}{\det(AA^T - \tau^2 I)}.$$

  Multiplying by w^T yields $0 = w^Ty = w^T\mathrm{adj}[AA^T - \tau^2 I]w$. Hence τ^2 is a root of $w^T\mathrm{adj}[AA^T - \tau^2 I]w$.
- If τ^2 is an eigenvalue of AA^T then one can choose y to be an eigenvector associated with τ^2. Let $AA^T = V\Omega V^T$ be an eigenvalue decomposition, where V is real orthogonal, and Ω is diagonal. For any scalar λ, $\mathrm{adj}[AA^T - \lambda I] = V\mathrm{adj}[\Omega - \lambda I]V^T$, because $\mathrm{adj}(XY) = \mathrm{adj}(Y)\mathrm{adj}(X)$ for square matrices X and Y, and $\mathrm{adj}(V) = \det(V)V^T$ for a real orthogonal matrix V. Thus

  $$w^T\mathrm{adj}[AA^T - \lambda I]w = w^TV\mathrm{adj}[\Omega - \lambda I]V^Tw.$$

  Assume that the diagonal elements $\omega_1, \dots, \omega_m$ of Ω are ordered so that $\omega_1 = \tau^2$ and $y = Ve_1$. From $0 = y^Tw = e_1^TV^Tw$ follows that the leading element of V^Tw is zero, i.e. $V^Tw = \begin{pmatrix} 0 & w_2 & \dots & w_m \end{pmatrix}^T$. The matrix $\Omega - \lambda I$ is a diagonal matrix, and so is its adjugate,

  $$\mathrm{adj}[\Omega - \lambda I] = \begin{pmatrix} \prod_{j\neq 1}(\omega_j - \lambda) & & \\ & \ddots & \\ & & \prod_{j\neq m}(\omega_m - \lambda) \end{pmatrix}.$$

  Substituting the expressions for V^Tw and $\mathrm{adj}[\Omega - \lambda I]$ into $w^T\mathrm{adj}[AA^T - \lambda I]w$

gives

$$w^T \operatorname{adj}[AA^T - \lambda I] w = \sum_{i=2}^{m} |w_i|^2 \prod_{j=1, j\neq i}^{m} (\omega_j - \lambda)$$
$$= \sum_{i=2}^{m} |w_i|^2 (\omega_1 - \lambda) \prod_{j=2, j\neq i}^{m} (\omega_j - \lambda).$$

Therefore $\tau^2 = \omega_1$ is a root of $w^T \operatorname{adj}[AA^T - \lambda I] w$.

Since each product $(\omega_1 - \lambda) \prod_{j=2, j\neq i}^{m} (\omega_j - \lambda)$ in $w^T \operatorname{adj}[AA^T - \lambda I] w$ consists of $m-1$ factors, the quantity $w^T \operatorname{adj}[AA^T - \lambda I] w$ is a polynomial of degree $m-1$ in λ. We still need to argue that $\omega_1 = \tau^2$ is the largest root of this polynomial. Applying the partitioning in Theorem 7.12 to the adjugate gives

$$\operatorname{adj}[AA^T - \lambda I] = Q \operatorname{adj}\left[\begin{pmatrix} a^T a - \lambda & a^T A_{m-1}^T \\ A_{m-1} a & B - \lambda I \end{pmatrix}\right] Q^T.$$

From $Qe_1 = w/\|w\|_2$ follows

$$w^T \operatorname{adj}[AA^* - \lambda I] w = \|w\|_2^2 e_1^T \operatorname{adj}\left[\begin{pmatrix} a^T a - \lambda & a^T A_{m-1}^T \\ A_{m-1} a & B - \lambda I \end{pmatrix}\right] e_1 = \|w\|_2^2 \det(B - \lambda I).$$

Since τ^2 is the largest root of $\det(B - \lambda I)$, it must also be the largest root of $w^T \operatorname{adj}[AA^T - \lambda I] w$. □

The expression below extends Corollary 6.8 from stochastic matrices to real rectangular matrices and arbitrary vectors. It suggests that $\tau_2(w, A)$ is the norm of an orthogonal projection of A projected onto the subspace $\operatorname{range}(w)^\perp$. In this sense, the expression below resembles the one in Theorem 7.12.

THEOREM 7.15 (Fourth Expression). *Let $A \in \mathbb{R}^{m\times n}$ and $w \in \mathbb{R}^m$ with $w \neq 0$. Then*

$$\tau_2(w, A) = \left\| \left(I - \frac{ww^T}{\|w\|_2^2} \right) A \right\|_2.$$

Proof. The proof is similar to the proofs of Theorem 6.7 and Corollary 6.8. The constraint $z^T w = 0$ is incorporated into $A^T z$ and into $z^T z = 1$. Also, a more general result is proved Theorem 8.6. □

Note that the two-norm expression for $\tau_2(w, A)$ in Theorem 7.15 represents an equality, while the Frobenius norm expression in Theorem 7.3 is only a bound.

In the special case when w is a dominant singular vector of A the expression in Theorem 7.15 reduces to the second largest singular value of A.

COROLLARY 7.16. *Let $A \in \mathbb{R}^{m\times n}$ have singular values $\sigma_1(A) \geq \sigma_2(A) \geq \ldots$ and dominant singular vectors v and u so that $Av = \sigma_1(A)\, u$ and $\|u\|_2 = \|v\|_2 = 1$. Then*

$$\tau_2(u, A) = \tau_2(v, A^T) = \sigma_2(A).$$

Proof. From $u^T A = \sigma_1(A) v^T$ and Theorem 7.15 follows

$$\tau_2(u, A) = \left\| \left(I - uu^T \right) A \right\|_2 = \|A - \sigma_1(A) uv^T\|_2.$$

Let $A = U\Sigma V^T$ be a singular value decomposition where $\Sigma_{11} = \sigma_1(A)$, and U and V are real orthogonal. Then $Ue_1 = u$, $Ve_1 = v$ and

$$\|A - \sigma_1(A)uv^T\|_2 = \|\Sigma - \sigma_1(A)e_1e_1^T\|_2 = \sigma_2(A).$$

Hence $\tau_2(u, A) = \sigma_2(A)$. The proof for $\tau_2(v, A^T)$ is analogous. □

Corollary 7.16 is an extension of the expression

$$\tau_2(\mathbb{1}, S_D) = \tau_2\left(\mathbb{1}, S_D^T\right) = \sigma_2(S_D)$$

for doubly stochastic matrices S_D in [74, p 3].

7.5. Eigenvalue Bounds for Nonnegative Matrices. We present bounds on inclusion regions for subdominant eigenvalues of nonnegative irreducible matrices.

Let $A \in \mathbb{R}^{n\times n}$ be a nonnegative irreducible matrix with eigenvalues λ_j and Perron vector $u > 0$ so that

$$Au = \lambda_1 u, \qquad \text{where} \quad \lambda_1 \geq |\lambda_2| \geq \cdots \geq |\lambda_n|. \tag{7.1}$$

Ergodicity coefficients $\tau_p(u, A)$ that are based on the Perron vector u bound the modulus of all subdominant eigenvalues, real as well as complex. The bound below extends the one-norm bound for stochastic matrices in Theorem 3.2 to p-norm bounds for nonnegative matrices.

THEOREM 7.17 (Theorem 3.1 in [60]). *If $A \in \mathbb{R}^{n\times n}$ is a nonnegative irreducible matrix as in (7.1) then*

$$|\lambda_i| \leq \tau_p(u, A), \qquad 2 \leq i \leq n.$$

Proof. The proof consists of constructing a norm $\|\cdot\|_c$ on $\mathbb{C}^n$ whose restriction to $\mathbb{R}^n$ is $\|\cdot\|_p$ and so that $\tau_p(u, A) = \max_{\substack{\|z\|_c=1 \\ z^T u=0,\, z\in\mathbb{C}^n}} \left\|A^T z\right\|_c$. □

An alternative option for eigenvalue bounds is to convert the matrix into one with constant row sums.

LEMMA 7.18 (p 293 in [5]). *If $A \in \mathbb{R}^{n\times n}$ is a nonnegative irreducible matrix as in (7.1) and $D_u = \operatorname{diag}(u)$ then $(D_u^{-1}AD_u)\,\mathbb{1} = \lambda_1 \mathbb{1}$.*

Since similarity transformations preserve the eigenvalues, one can bound the eigenvalues of A in terms of an ergodicity coefficient based on $D_u^{-1}AD_u$ and its Perron vector $\mathbb{1}$.

THEOREM 7.19 (p 63 in [60]). *If $A \in \mathbb{R}^{n\times n}$ is a nonnegative irreducible matrix as in (7.1) and $D_u = \operatorname{diag}(u)$ then*

$$|\lambda_i| \leq \tau_p\left(\mathbb{1}, D_u^{-1}AD_u\right), \qquad 2 \leq i \leq n.$$

Remark 7.1 (p 346 in [68]). *The bound in Theorem 7.19, when applied to stochastic matrices can be tighter than the bound in Theorem 7.17.*

Consider the stochastic matrix S with stationary distribution π, where

$$S = \begin{pmatrix} 1/2 & 5/16 & 3/32 & 3/32 \\ 1/2 & 5/16 & 3/32 & 3/32 \\ 0 & 5/8 & 3/16 & 3/16 \\ 0 & 5/8 & 3/16 & 3/16 \end{pmatrix} \qquad \text{and} \qquad \pi = \begin{pmatrix} 5/13 \\ 5/13 \\ 3/26 \\ 3/26 \end{pmatrix}.$$

With $D_\pi = \mathrm{diag}(\pi)$ *we obtain*

$$D_\pi^{-1} S D_\pi = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 5/16 & 5/16 & 3/16 & 3/16 \\ 5/16 & 5/16 & 3/16 & 3/16 \\ 5/16 & 5/16 & 3/16 & 3/16 \end{pmatrix}.$$

The eigenvalues of S *are* 1, 3/16 *and* 0, *while the ergodicity coefficients are*

$$\tau_1(S) = \frac{1}{2}, \quad \tau_1\left(S^T\right) = \frac{13}{16}, \quad \tau_\infty(S) = 1, \quad \tau_\infty\left(S^T\right) = \frac{5}{8},$$

and

$$\tau_1\left(D_u^{-1} S D_u\right) = \frac{3}{8}, \qquad \tau_\infty\left(D_u^{-1} S D_u\right) = \frac{3}{16}.$$

Thus $\tau_1\left(D_u^{-1} S D_u\right)$ *and* $\tau_\infty\left(D_u^{-1} S D_u\right)$ *are the tightest upper bounds on the subdominant eigenvalues.*

Below we bound $\tau_p(u, A)$ by the norm of A deflated by its dominant spectral projector.

Corollary 7.20. *If* $A \in \mathbb{R}^{n\times n}$ *is a nonnegative irreducible matrix as in (7.1) and* $v^T A = \lambda_1 v^T$ *then*

$$|\lambda_i| \le \tau_p(u, A) \le \|(A - \lambda_1 u v^T)^T\|_p, \qquad 2 \le i \le n.$$

Proof. This follows from Theorem 7.17, and from Theorem 7.3 with $x = \lambda_1 v$. □

In particular, Corollary 7.20 implies an inclusion interval for $\tau_p(u, A)$ in terms of A deflated by its dominant spectral projector,

$$\rho\left(A - \lambda_1 u v^T\right) \le \tau_p(u, A) \le \|(A - \lambda_1 u v^T)^T\|_p,$$

where $\rho(A) = \max_{1\le i\le n} |\lambda_i|$.

8. Complex Matrices and General Subspaces.

The most general form of ergodicity coefficient [60, §7], [29, §3], is defined for complex matrices $A \in \mathbb{C}^{m\times n}$ and the maximization takes place over subspaces of arbitrary dimension,

$$\tau_p(W, A) \equiv \max_{\substack{\|z\|_p = 1 \\ z^* W = 0}} \|A^* z\|_p,$$

where the maximum ranges over $z \in \mathbb{C}^m$.

We discuss properties of these coefficients and their application to inclusion regions for subdominant eigenvalues. Then we focus on the two-norm coefficient, for which derive explicit expressions and establish its relation to singular values. For normal matrices, we show that the two-norm coefficient is a Lehmann bound.

8.1. Properties Common to All p-Norm Coefficients.

Like for real matrices, the coefficients $\tau_p(W, A)$ are bounded, well-conditioned in the second argument, and weakly submultiplicative in the second argument.

Theorem 8.1. *If* $A, A_1, A_2 \in \mathbb{C}^{m\times n}$ *and* $W \in \mathbb{C}^{m\times k}$ *then*

1. $0 \le \tau_p(W, A) \le \|A^*\|_p$
2. $|\tau_p(W, A_1) - \tau_p(W, A_2)| \le \tau_p(W, A_1 - A_2)$

3. $\tau_p(W, BA) \leq \|A^*\|_p \, \tau_p(W, B)$.

Proof. The proof is analogous to that for real matrices in Theorem 7.1. □

As in the case of real matrices in Theorem 7.3,

the upper bound on $\tau_p(W, A)$ can possibly be improved by representing $\tau_p(W, A)$ as the norm of a downdated matrix.

THEOREM 8.2. *If $A \in \mathbb{C}^{m\times n}$ and $W \in \mathbb{C}^{m\times k}$ then for all $X \in \mathbb{C}^{k\times n}$*

$$\tau_p(W, A) \leq \|A^* - XW^*\|_p .$$

Proof. Let $z \in \mathbb{C}^n$ be a vector with $z^*W = 0$ and $\|z\|_p = 1$. Then

$$(A - WX^*)^* \, z = A^*z - XW^*z = A^*z$$

implies

$$\tau_p(W, A) = \max_{\substack{\|z\|_p=1 \\ z^*W=0}} \left\|(A - WX^*)^* \, z\right\|_p = \tau_p\left(W, A - WX^*\right) \leq \left\|(A - WX^*)^*\right\|_p ,$$

where the last inequality follows from Theorem 8.1. □

Invariant Subspaces. If the columns of W happen to span an invariant subspace of A, then a submultiplicative property holds for powers of A. We also show that the ergodicity coefficients can determine eigenvalue inclusion regions.

THEOREM 8.3. *Let $A \in \mathbb{C}^{n\times n}$ and $W \in \mathbb{C}^{n\times k}$ with $AW = WC$ for some $C \in \mathbb{C}^{k\times k}$. Then for $l, m \geq 1$*

$$\tau_p\left(W, A^{l+m}\right) \leq \tau_p\left(W, A^l\right) \tau_p\left(W, A^m\right).$$

Proof. The proof is analogous to that for real matrices in Theorem 7.2. □

If W spans a right invariant subspace associated with dominant eigenvalues, then an eigenvalue bound holds that is similar to the one for stochastic matrices in Theorem 3.2 and for irreducible nonnegative matrices in Theorem 7.17.

THEOREM 8.4. *Let $A \in \mathbb{C}^{n\times n}$ have eigenvalues λ_j, labeled $|\lambda_1| \geq \ldots \geq |\lambda_n|$. If $\lambda_k \neq \lambda_{k+1}$ for some $1 \leq k < n$, and if the columns of $W \in \mathbb{C}^{n\times k}$ span a right invariant subspace of A associated with $\lambda_1, \ldots, \lambda_k$ then*

$$|\lambda_i| \leq \tau_p(W, A), \qquad k+1 \leq i \leq n.$$

Proof. Let $v \in \mathbb{C}^n$ be a left eigenvector of A for an eigenvalue λ_i, $k+1 \leq i \leq n$, so that $v^*A = \lambda_i v^*$ and $\|v\|_p = 1$. Since the columns of W represent a right invariant subspace associated with eigenvalues different from λ_i, we have $v^*W = 0$. Hence

$$|\lambda_i| = \left|\overline{\lambda_i} v\right| = \|A^*v\|_p \leq \max_{\substack{\|z\|_p=1 \\ z^*W=0}} \|A^*z\|_p = \tau_p(W, A), \qquad k+1 \leq i \leq n.$$

□

If the columns of W are actually Jordan vectors, then we can say something about the tightness of $\tau_2(W, A)$ as an eigenvalue bound. Let $A = X\Lambda X^{-1}$ where

$$\Lambda = \begin{pmatrix} \Lambda_1 & \\ & \Lambda_2 \end{pmatrix}, \qquad X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}, \qquad X^{-1} = \begin{pmatrix} Y_1^* \\ Y_2^* \end{pmatrix},$$

and

$$\Lambda_1 = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{pmatrix}, \qquad \Lambda_2 = \begin{pmatrix} \lambda_{k+1} & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

THEOREM 8.5. *Let $A \in \mathbb{C}^{n\times n}$ have eigenvalues λ_j, labeled $|\lambda_1| \geq \ldots \geq |\lambda_n|$. If $\lambda_k \neq \lambda_{k+1}$ for some $1 \leq k < n$, then*

$$|\lambda_{k+1}| \leq \tau_p(X_1, A) \leq |\lambda_{k+1}| \, \|X_2^*\|_p \, \|Y_2^*\|_p \, .$$

Proof. The lower bound follows from Theorem 8.4.

With regard to the upper bound, let y be a vector with $\tau_p(W, A) = \|A^*y\|_p$, $y^*W = 0$ and $\|y\|_p = 1$. Then $A^*y = Y_2\Lambda_2^*X_2^*\, y$ and

$$\tau_p(W, A) = \|A^*y\|_p \leq \|X_2^*\|_p \, \|\Lambda_2^*\|_p \, \|Y_2^*\|_p \, .$$

Since Λ_2 is a diagonal matrix, $\|\Lambda_2\|_p = |\lambda_{k+1}|$. ◻

Theorem 8.5 suggests that the tightness of $\tau_2(W, A)$ as an eigenvalue bound for the subdominant eigenvalues depends on their Jordan vectors. An analogous bound based on the Schur decomposition will be presented in Theorem 8.11.

In the remaining sections we concentrate on two-norm coefficients.

8.2. Explicit Expressions for Two-Norm Coefficients. The four explicit expressions for two-norm coefficients of real matrices in Section 7.4 also hold for complex matrices. In particular, we extend Theorems 7.12 and 7.15 to complex matrices and general subspaces. The following two bounds demonstrate that $\tau_2(W, A)$ is the norm of an orthogonally projected matrix, where the projection is onto range$(W)^{\perp}$.

THEOREM 8.6. *Let $A \in \mathbb{C}^{m\times n}$, and let $W \in \mathbb{C}^{m\times k}$ have orthonormal columns. Let $Q \in \mathbb{C}^{m\times m}$ be a unitary matrix with $Q = \begin{pmatrix} W & Q_2 \end{pmatrix}$, and partition $A^*Q = \begin{pmatrix} A_k & A_{m-k} \end{pmatrix}$, where A_{m-k} has $m-k$ columns. Then*

$$\tau_2(W, A) = \|A_{m-k}\|_2.$$

Proof. The proof is analogous to that of Theorem 7.12 for real matrices. ◻

In contrast to the matrix A_{m-k} in Theorem 8.6, which has fewer rows than A, the projected matrix below has the same dimension as A.

THEOREM 8.7. *Let $A \in \mathbb{C}^{m\times n}$ and $W \in \mathbb{C}^{m\times k}$ with* rank$(W) = k$. *Then*

$$\tau_2(W, A) = \left\| \left(I - W(W^*W)^{-1}W^*\right) A \right\|_2 .$$

Proof. Let z be a vector with $z^*W = 0$ and $\|z\|_2 = 1$; and let P be permutation matrix so that

$$PW = \begin{pmatrix} W_k \\ W_{m-k} \end{pmatrix}, \qquad Pz = \begin{pmatrix} z_k \\ z_{m-k} \end{pmatrix},$$

where the $k \times k$ matrix W_k is nonsingular and z_k has k elements. Then $z^*W = 0$ implies $z_k^* = -z_{m-k}^*X$, where $X = W_{m-k}W_k^{-1}$. We can incorporate the constraint $z^*W = 0$ into z by writing

$$Pz = \begin{pmatrix} -X^* \\ I_{m-k} \end{pmatrix} z_{m-k}.$$

As a consequence,

$$A^*z = A^*P^* \begin{pmatrix} -X^* \\ I_{m-k} \end{pmatrix}, \qquad 1 = z^*z = z_{m-k}^*(I_{m-k} + XX^*)z_{m-k}.$$

The matrix $I_{m-k}+XX^*$ is Hermitian positive definite, hence has a Hermitian positive-definite square root $(I_{m-k} + XX^*)^{1/2}$. Setting $y = (I_{m-k} + XX^*)^{1/2}z_{m-k}$ gives $y^*y = 1$ and

$$A^*z = A^*By, \qquad \text{where} \quad B = P^* \begin{pmatrix} -X^* \\ I_{m-k} \end{pmatrix} (I_{m-k} + XX^*)^{-1/2}.$$

Therefore, maximizing $\|A^*z\|_2$ over $z \in \mathbb{C}^m$ subject to $z^*W = 0$ and $\|z\|_2 = 1$ is equivalent to maximizing $\|A^*By\|_2$ over $y \in \mathbb{C}^{m-k}$ subject to $\|y\|_2 = 1$. This means $\tau_2(W, A) = \max_{\|y\|_2=1} \|A^*By\|_2 = \|A^*B\|_2$.

The $m \times (m-k)$ matrix B has orthonormal columns, thus $\|A^*B\|_2 = \|B^*A\|_2 = \|BB^*A\|_2$, and BB^* is the orthogonal projector onto the space

$$\text{range}(B) = \text{range}\left(P^* \begin{pmatrix} -X^* \\ I_{m-k} \end{pmatrix}\right) = \text{range}\left(P^* \begin{pmatrix} I_k \\ X \end{pmatrix}\right)^{\perp}.$$

Since

$$P^* \begin{pmatrix} I_k \\ X \end{pmatrix} = P^* \begin{pmatrix} W_k \\ W_{m-k} \end{pmatrix} W_k^{-1} = W\, W_k^{-1},$$

we obtain $\text{range}(B) = \text{range}(W)^{\perp}$. The uniqueness of orthogonal projectors implies $BB^* = I - W(W^*W)^{-1}W^*$. □

8.3. Two-Norm Coefficients and Singular Values. We show that two-norm ergodicity coefficients are closely related to singular values.

The result below It implies that two-norm ergodicity coefficients based on dominant singular vectors can reproduce any singular value. This is in contrast to eigenvalues, where ergodicity coefficients yield only bounds, see Theorem 8.5. The result below extends Corollary 7.16 from real matrices to complex matrices and arbitrary subspaces.

Corollary 8.8. *Let $A \in \mathbb{C}^{m\times n}$ have singular values $\sigma_1 \geq \sigma_2 \geq \ldots$, and let the columns of U_j and V_j consist of the respective left and right singular vectors associated with $\sigma_1, \ldots, \sigma_j$. If $1 \leq k < \min\{m, n\}$ then*

$$\tau_2(U_k, A) = \tau_2(V_k, A^*) = \sigma_{k+1}.$$

Proof. This follows from Theorem 8.7, and the proof is analogous to that of Corollary 7.16. □

More generally, for all matrices W with k columns, singular values σ_{k+1} and σ_1 represent the extreme values for $\tau_2(W, A)$.

Theorem 8.9. *Let $A \in \mathbb{C}^{m\times n}$ have singular values $\sigma_1 \geq \sigma_2 \geq \ldots$. If $1 \leq k \leq \min\{m, n\}$ then*

$$\min_{W\in\mathbb{C}^{n\times k}} \tau_2(W, A) = \sigma_{k+1}, \qquad \max_{W\in\mathbb{C}^{n\times k}} \tau_2(W, A) = \sigma_1.$$

Proof. The variational characterization of singular values [31, Theorem 7.3.10] implies

$$\sigma_{k+1} = \min_{X\in\mathbb{C}^{n\times k}} \max_{\substack{z\in\mathbb{C}^n, \|z\|_2=1\\ z^*X=0}} \|A^*z\|_2 \leq \max_{\substack{\|z\|_2=1\\ z^*W=0}} \|A^*z\|_2 = \tau_2(W,A).$$

Corollary 8.8 shows that the minimum is attained if the columns of W are the k left singular vectors associated with the k largest singular values σ_j, $1 \leq j \leq k$.

The upper bound follows from Theorem 8.1. The maximum is attained if the columns of W are k left singular vectors associated with singular values σ_j for $j > 1$. □

8.4. Two-Norm Coefficients and Eigenvalues. For a normal matrix, Theorems 8.5 and 8.9 readily imply that the two-norm ergodicity coefficients based on dominant eigenvectors can reproduce the magnitude of any eigenvalue.

THEOREM 8.10. *Let $A \in \mathbb{C}^{n\times n}$ be a normal matrix with eigenvalues λ_j, labelled so that $|\lambda_1| \geq \ldots \geq |\lambda_n|$, and $\lambda_k \neq \lambda_{k+1}$. If $W \in \mathbb{C}^{n\times k}$ has orthonormal columns that span an invariant subspace associated with $\lambda_1, \ldots, \lambda_k$ then*

$$|\lambda_{k+1}| = \tau_2(W,A).$$

Proof. Since A is normal, it has singular values $|\lambda_j|$, and Theorem 8.9 implies $|\lambda_{k+1}| \leq \tau_2(W,A)$. The Jordan matrices X are unitary, so that the Jordan vectors X_2 and Y_2 in Theorem 8.5 have orthonormal columns and $\|X_2\|_2 = \|Y_2\|_2 = 1$. From Theorem 8.5 follows $\tau_2(W,A) \leq |\lambda_{k+1}|$. □

More generally, one can try to use $\tau_2(W,A)$ as an inclusion region for subdominant eigenvalues of nonnormal matrices A. In Theorem 8.5 this was done by choosing dominant Jordan vectors for W. Below we derive an analogous bound when W consists of dominant Schur vectors.

Let $A = Q(\Lambda + N)Q^*$ be a Schurdecomposition, where Q is unitary, Λ is diagonal, and N is strictly upper triangular.

THEOREM 8.11. *Let $A \in \mathbb{C}^{n\times n}$ have eigenvalues λ_j, ordered so that $|\lambda_1| \geq \ldots \geq |\lambda_k|$ and $\lambda_k \neq \lambda_{k+1}$. Let $W \in \mathbb{C}^{n\times k}$ have orthonormal columns that span an invariant subspace associated with $\lambda_1, \ldots, \lambda_k$. Then*

$$|\lambda_{k+1}| \leq \tau_2(W,A) \leq |\lambda_{k+1}| + \|N\|_2$$

Proof. The matrix W can be chosen as the leading k columns of Q, see [23, Lemma 7.1.2]. The remaining proof is similar to the proofs Theorem 8.5 and 8.10. □

Theorem 8.11 implies that $\tau_2(W,A)$ based on dominant Schur vectors provides good inclusion regions for subdominant eigenvalues if A is close to normal, that is, if the departure of A from normality, $\|N\|_2$, is small.

Connection to Lehmann Bounds. We illustrate that two-norm ergodicity coefficients for normal matrices are special cases of Lehmann bounds.

So-called "Lehmann bounds" are a particular type of eigenvalue inclusion regions. They are expressed in terms of singular values of the matrix restricted to a subspace, see [49, §10.5] for Hermitian matrices and [6] for general matrices. Theorem 8.12 below presents Lehmann bounds for normal matrices. We use $\sigma_i(B)$ to denote the ith largest singular value of the matrix B.

THEOREM 8.12 (Corollary 2.3 in [6]). *Let $A \in \mathbb{C}^{n\times n}$ be a normal matrix, $X \in \mathbb{C}^{n\times m}$ have orthonormal columns, and γ be a complex scalar. Then each disk*

$$\{\lambda : |\lambda - \gamma| \le \sigma_i\left((A-\gamma I)X\right)\}, \qquad 1 \le i \le m$$

contains at least $m-i+1$ eigenvalues of A.

It turns out that for any full column rank matrix W, the ergodicity coefficient $\tau_2(W,A)$ is a Lehmann bound for a normal matrix A.

THEOREM 8.13. *Let $A \in \mathbb{C}^{n\times n}$ be a normal matrix, let $W \in \mathbb{C}^{n\times k}$ have linearly independent columns, and let γ be a complex scalar. Then the disk*

$$\{\lambda : |\lambda - \gamma| \le \tau_2\left(W, A-\gamma I\right)\}$$

contains at least $n-k$ eigenvalues of A. That is, $\tau_2\left(W,(A-\gamma I)\right)$ is a Lehman bound.

Proof. Let $X \in \mathbb{C}^{n\times(n-k)}$ have orthonormal columns with $\mathrm{range}(X) = \mathrm{range}(W)^{\perp}$. Then

$$\begin{aligned}\sigma_1\left((A-\gamma I)^* X\right) &= \max_{\|y\|_2=1} \left\|(A-\gamma I)^* Xy\right\|_2 = \max_{\|Xy\|_2=1} \left\|(A-\gamma I)^* Xy\right\|_2 \\ &= \max_{\substack{\|z\|_2=1 \\ z^*W=0}} \left\|(A-\gamma I)^* z\right\|_2 = \tau_2\left(W, A-\gamma I\right).\end{aligned}$$

Now apply Theorem 8.12 with $m = n-k$ and $i = 1$. □

Acknowledgments. We thank Carl Meyer, Steve Campbell, and Steve Kirkland for their encouragement and many helpful discussions.

REFERENCES

[1] Y. A. Al'pin and N. Z. Gabassov, *A remark on the localization of the characteristic values of real matrices*, Izvestiya VUZ. Matematika, 20 (1976), pp. 98–100.
[2] M. Artzrouni, *The local coefficient of ergodicity of a nonnegative matrix*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 507–516.
[3] M. Artzrouni and O. Gavart, *Nonlinear matrix iterative processes and generalized coefficients of ergodicity*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1343–1353.
[4] M. Artzrouni and X. Li, *A note on the coefficient of ergodicity of a column-allowable nonnegative matrix*, Linear Algebra Appl., 214 (1995), pp. 93–101.
[5] F. L. Bauer, E. Deutsch, and J. Stoer, *Abschätzungen für Eigenwerte positiver linearer Operatoren*, Linear Algebra Appl., 2 (1969), pp. 275–301.
[6] C. Beattie and I. C. Ipsen, *Inclusion regions for matrix eigenvalues*, Linear Algebra Appl., 358 (2003), pp. 281–291.
[7] A. Berman and R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, Academic Press, New York, 1979.
[8] S. N. Bernštein, *Sur le théorème limite du calcul des probabilités*, Math. Ann., 85 (1922), pp. 237–241.
[9] ———, *Sur les sommes de quantités dépendantes*, Izv. Akad. Nauk SSSR, 15-17 (1926), pp. 1459–1478.
[10] ———, *Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes*, Math. Ann., 97 (1926), pp. 1–59. (Russian translation in Uspekhi Mat. Nauk, 10 (1944), pp. 65–144).
[11] ———, *Sur les sommes de quantités dépendantes*, Dokl. Akad. Nauk SSSR., A, 4 (1928), pp. 55–60.
[12] ———, *Teoriya Veroiatnostei*, Gostehizdat, Moscow-Leningrad, fourth ed., 1946. Relevant portions are republished in [13].
[13] ———, *Classification of Markov chains and their matrices*, Sobranie Sochineniy: Tom IV, Teoriya Veroiatnostei i Matematicheskaia Statistika [1911-1946], (1964), pp. 455–483.
[14] S. Chatterjee and E. Seneta, *Towards consensus: some convergence theorems on repeated averaging*, J. Appl. Probability, 14 (1977), pp. 89–97.

[15] G. E. Cho and C. D. Meyer, *Comparison of perturbation bounds for the stationary distribution of a Markov chain*, Linear Algebra Appl., 335 (2001), pp. 137–150.
[16] H. Cohn, *Finite non-homogeneous Markov chains: Asymptotic behaviour*, Adv. Appl. Probab., 8 (1976), pp. 502–516.
[17] ———, *On a paper by Doeblin on non-homogeneous Markov chains*, Adv. Appl. Probab., 13 (1981), pp. 388–401.
[18] R. L. Dobrushin, *Central limit theorem for non-stationary Markov chains*, Dokl. Akad. Nauk SSSR., 102: 1 (1955), pp. 5–8.
[19] ———, *Central limit theorem for non-stationary Markov chains, I, II*, Theory Prob. Appl., 1 (1956), pp. 65–80.
[20] W. Doeblin, *Le cas discontinu des probabilités en chaîne*, Publ. Fac. Sci. Univ. Masaryk (Brno), 236 (1937), pp. 3–13.
[21] E. B. Dynkin, *On certain limit theorems for Markov chains*, Ukrain. Mat. Zh., 6: 1 (1954), pp. 21–27.
[22] L. Elden, *The eigenvalues of the Google matrix*, Tech. Rep. LiTH-MAT-R-04-01, Department of Mathematics, Linköping University, Sweden, December 2003.
[23] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, third ed., 1996.
[24] O. Gross and U. G. Rothblum, *Approximations of the spectral radius, corresponding eigenvector, and second largest modulus of an eigenvalue for square, nonnegative, irreducible matrices*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 15–32.
[25] J. Hajnal, *The ergodic properties of non-homogeneous finite Markov chains*, Proc. Cambridge Philos. Soc., 52 (1956), pp. 67–77.
[26] ———, *Weak ergodicity in nonhomogeneous Markov chains*, Proc. Cambridge Philos. Soc., 54 (1958), pp. 233–246.
[27] ———, *On products of non-negative matrices*, Math. Proc. Cambridge Philos. Soc., 79 (1976), pp. 521–530.
[28] D. Hartfiel, *Coefficients of ergodicity for imprimitive matrices*, Commun. Stat. – Stochastic Models, 15 (1999), pp. 81–88.
[29] D. Hartfiel and U. G. Rothblum, *Convergence of inhomogeneous products of matrices and coefficients of ergodicity*, Linear Algebra Appl., 277 (1998), pp. 1–9.
[30] T. H. Haveliwala and S. D. Kamvar, *The second eigenvalue of the Google matrix*, Tech. Rep. 2003-20, Computer Science Department, Stanford University, Palo Alto, CA, 2003.
[31] R. A. Horn and C. A. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
[32] R. A. Horn and S. Serra-Capizzano, *A general setting for the parametric Google matrix*, Internet Math., 3 (2006), pp. 385–411.
[33] M. Iosifescu, *On two recent papers on ergodicity in non-homogeneous Markov chains*, Ann. Math. Statist., 43 (1972), pp. 1732–1736.
[34] ———, *Finite Markov Processes and their Applications*, John Wiley, 1980.
[35] I. C. F. Ipsen and S. Kirkland, *Convergence analysis of a PageRank updating algorithm by Langville and Meyer*, SIAM J. Matrix Anal. Appl., 27 (2006), pp. 952–967.
[36] I. C. F. Ipsen and T. M. Selee, *Pagerank computation, with special attention to dangling nodes*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 1281–1296.
[37] I. C. F. Ipsen and R. S. Wills, *Mathematical properties and analysis of Google's PageRank*, Bol. Soc. Esp. Mat. Apl., 34 (2006), pp. 191–196.
[38] D. Isaacson and R. Madsen, *Positive columns for stochastic matrices*, J. Appl. Prob., 11 (1974), pp. 829–835.
[39] D. L. Isaacson and R. Madsen, *Markov Chains*, Wiley, 1976.
[40] J. F. C. Kingman, *Geometrical aspects of the theory of non-homogeneous Markov chains*, Math. Proc. Cambridge Philos. Soc., 77 (1975), pp. 171–183.
[41] S. J. Kirkland, M. Neumann, and N.-S. Sze, *On optimal condition numbers for Markov chains*, Numer. Math., 110 (2008), pp. 521–537.
[42] A. N. Kolmogorov, *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung*, Math. Ann., 104 (1931), pp. 415–458.
[43] T. S. Leóng, *A note on upper bounds on the maximum modulus of subdominant eigenvalues of nonnegative matrices*, Linear Algebra Appl., 106 (1988), pp. 1–4.
[44] A. Lešanovský, *Coefficient of ergodicity generated by non-symmetrical vector norms*, Czechoslovak Mathematical Journal, 40 (1990), pp. 284–294.
[45] A. A. Markov, *Extension of the law of large numbers to dependent quantities (in Russian)*, Izvestiia Fiz.-Matem. Obsch. Kazan Univ., 2 (1906), pp. 135–156. English translation [47].
[46] ———, *Izbrannie Trudy (Selected Works)*, A.N.S.S.S.R., Leningrad, 1951.

[47] ———, *The extension of the law of large numbers onto quantities depending on each other*, in Probability and Statistics. Russian Papers selected and translated by Oscar Sheynin, O. B. Sheynin, ed., NG Verlag, Berlin, 2004, pp. 143–158. Translation of [45] into English.

[48] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, Society for Industrial and Applied Mathematics, Philadelphia, 2000.

[49] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice Hall, Englewood Cliffs, 1980.

[50] A. Paz, *Ergodic theorems for infinite probabilistic tables*, Ann. Math. Statist., 41 (1970), pp. 539–550.

[51] ———, *Introduction to Probabilistic Automata*, Academic Press, 1971.

[52] A. Paz and M. Reichaw, *Ergodic theorems for sequences of infinite stochastic matrices*, Proc. Cambridge Philos. Soc., 63 (1967), pp. 777–784.

[53] H. Perfect and L. Mirsky, *Spectral properties of doubly-stochastic matrices*, Monatshefte für Mathematik, 69 (1965), pp. 35 – 57.

[54] Y. A. Pykh, *Application of the coefficient of ergodicity in the estimation of the spectral radius of real matrices*, Math. Notes, 23 (1978), pp. 74–76.

[55] A. Rhodius, *The maximal value for coefficients of ergodicity*, Stochastic Process. Appl., 29 (1988), pp. 141–145.

[56] ———, *Über explizite Formeln für Ergodizitätskoeffizienten*, Wiss. Z. Pädagog. Hochsch. "Karl Friedrich Wilhelm Wander", 23 (1989), pp. 69–74.

[57] ———, *On explicit forms for ergodicity coefficients*, Linear Algebra Appl., 194 (1993), pp. 71–83.

[58] ———, *On the maximum of ergodicity coefficients, the Dobrushin ergodicity coefficient, and products of stochastic matrices*, Linear Algebra Appl., 253 (1997), pp. 141–154.

[59] ———, *On ergodicity coefficients of infinite stochastic matrices*, Z. Anal. Anwendungen, 19 (2000), pp. 873–887.

[60] U. G. Rothblum and C. P. Tan, *Upper bounds on the maximum modulus of subdominant eigenvalues of nonnegative matrices*, Linear Algebra Appl., 66 (1985), pp. 45–86.

[61] N. A. Sapogov, *The Laplace-Lyapunov limit theorem for singular Markov chains*, Dokl. Akad. Nauk SSSR., 58: 9 (1947), pp. 1905–1908.

[62] ———, *On singular Markov chains*, Dokl. Akad. Nauk SSSR., 58: 2 (1947), pp. 193–196.

[63] T. A. Sarymsakov, *On inhomogeneous Markov chains*, Dokl. Akad. Nauk S.S.S.R., 120 (1958), pp. 465–467.

[64] T. M. Selee, *Stochastic Matrices: Ergodicity Coefficients and Applications to Ranking*, PhD thesis, North Carolina State University, 2008.

[65] E. Seneta, *On the historical development of the theory of finite inhomogeneous Markov chains*, Proc. Cambridge Philos. Soc., 74 (1973), pp. 507–513.

[66] ———, *Coefficients of ergodicity: structure and applications*, Adv. in Appl. Probab., 11 (1979), pp. 576–590.

[67] ———, *Non-Negative Matrices and Markov Chains*, Springer-Verlag, New York, second ed., 1981.

[68] ———, *Spectrum localization by ergodicity coefficients for stochastic matrices*, Linear and Multilinear Algebra, 14 (1983), pp. 343–347.

[69] ———, *Explicit forms for ergodicity coefficients and spectrum localization*, Linear Algebra Appl., 60 (1984), pp. 187–197.

[70] ———, *Perturbations of the stationary distribution measured by ergodicity coefficients*, Adv. Appl. Probab., 20 (1988), pp. 228–230.

[71] ———, *Applications of ergodicity coefficients to homogeneous Markov chains*, in Doeblin and modern probability, H. Cohn, ed., vol. 149 of Contemp. Math., Providence, RI, 1993, Amer. Math. Soc., pp. 189 – 199.

[72] ———, *Markov and the birth of chain dependence theory*, International Statistical Review, 64 (1996), pp. 255–263.

[73] ———, *Markov and the creation of Markov chains*, in MAM 2006: Markov Anniversary Meeting, A. N. Langville and W. J. Stewart, eds., Boson Books, Raleigh, North Carolina, USA, 2006, pp. 1–20.

[74] E. Seneta and C. P. Tan, *The Euclidean and Frobenius ergodicity coefficients and spectrum localization*, Bull. Malaysian Math. Soc., 7 (1984), pp. 1–7.

[75] S. Serra-Capizzano, *Jordan canonical form of the Google matrix: A potential contribution to the PageRank computation*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 305–312.

[76] C. P. Tan, *A functional form for a particular coefficient of ergodicity*, J. Appl. Prob., 19 (1982), pp. 858–863.

[77] ———, *Coefficients of ergodicity with respect to vector norms*, J. Appl. Prob., 20 (1983), pp. 277–287.

[78] R. S. Wills and I. C. F. Ipsen, *Ordinal ranking for Google's PageRank*, SIAM J. Matrix Anal. Appl., 30 (2009), pp. 1677–1696.