SIAM J. MATRIX ANAL. APPL.
Vol. 32, No. 1, pp. 153–200

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 classifications. 15-02, 15A18, 15A42, 15A48, 15A51, 65F15, 65F35

DOI. 10.1137/090752948

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, $1 = \lambda_1(S) \geq |\lambda_2(S)| \geq \dots$. 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 [27, 71].

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

*Received by the editors March 17, 2009; accepted for publication (in revised form) January 18, 2011; published electronically March 9, 2011.
http://www.siam.org/journals/simax/32-1/75294.html
†Department of Mathematics, North Carolina State University, P.O. Box 8205, Raleigh, NC 27695-8205 (ipsen@ncsu.edu, http://www4.ncsu.edu/~ipsen/).
‡Georgia Tech Research Institute, Sensors and Electromagnetic Applications Laboratory, 7220 Richardson Road, Smyrna, GA 30060 (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.

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.

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

$$\tau_1(S) = \max_{\|z\|_1=1,\ z^T \mathbb{1}=0} \|S^T z\|_1, \tag{1.1}$$

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 $\mathbb{1}$. 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: $\tau_1(S_1 S_2) \leq \tau_1(S_1)\tau_1(S_2)$ for any two stochastic matrices S_1 and S_2.

Seneta [78, section 1] identifies two ways to think about ergodicity coefficients: First one can think of an ergodicity coefficient as a vector norm, maximized over a particular subspace, as in (1.1). From this point of view it is natural to extend the concept of ergodicity coefficient to any rectangular matrix A, any vector norm p, and any vector w,

$$\tau_p(w, A) = \max_{\|z\|_p=1,\ z^T w=0} \|A^T z\|_p. \tag{1.2}$$

Second, one can think of ergodicity coefficients as eigenvalue bounds expressed in terms of a deflated matrix, with the deflation approximating the dominant spectral projector,

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

Seneta's two viewpoints, maximization of vector norms and deflation by projectors, represent the guiding principle for this paper. Initially, though, it was the second point of view, the connection of ergodicity coefficients to deflated (or downdated) matrices, that sparked our interest, as we explain in the next section.

1.1. Motivation. We became interested in coefficients of ergodicity in the context of work on the Google matrix [43, 44, 45, 84, 90].

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 a 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 $\pi^T G = \pi^T$ and $\pi^T \mathbb{1} = 1$. In [43] 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, $1 = \lambda_1(G) \geq |\lambda_2(G)| \geq \cdots$. A derivation "from scratch" [25, 34, 87] shows that $|\lambda_2(G)| \leq \alpha$, but it also follows immediately from $|\lambda_2(G)| \leq \tau_1(G) = \alpha\tau_1(S) \leq \alpha$, since $\tau_1(S) \leq 1$. The asymptotic convergence of the power method on G and its relation to ergodicity has also been noted by Seneta [84, section 8.2],

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 stochastic 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 [38]. Let A be a complex square matrix with dominant eigenvalue λ and right eigenvector w, i.e., $Aw = \lambda w$. Set $H \equiv \gamma A + (1-\gamma) w x^*$, where the superscript $*$ denotes the conjugate transpose, γ is a complex scalar, and x is a complex column vector with $x^* w = 1$. Then one can show [14, Theorems 29 and 32], [38, Corollary 3.3] that $\lambda_2(H) = \gamma\, \lambda_2(A)$. With the more general ergodicity coefficient (1.2) 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.

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. Our difficulty in understanding the entirety of the existing results was the motivation for writing 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 [46], Paz [61], Paz and Reichaw [63], and Rhodius [70].

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 (section 3), infinity norm (section 4), and any p-norm (section 5). Subsequently we extend the coefficients to real matrices (section 6), and to complex matrices with maximization over arbitrary subspaces (section 7). We illustrate applications of ergodicity coefficients to estimating the sensitivity of stationary distributions (section 3.4), and determining inclusion regions for eigenvalues of several classes of matrices, which include nonnegative (section 6.5), general complex (section 7.2), and normal matrices (section 7.5). We end with a summary and a few suggestions for further research (section 8). The bibliography includes back references that point to the pages where each reference is cited.

1.3. New results. We present a self-contained proof for the explicit form of the one-norm ergodicity coefficient (Theorem 3.7).

We represent ergodicity coefficients as norms of obliquely projected matrices, in the one-norm (Corollary 3.8), and in the infinity-norm (Corollaries 4.4 and 6.14). For the two-norm, we derive explicit expressions in terms of orthogonal projections of the matrix (Theorems 6.15, 6.19, 7.6, and 7.7). We show that general ergodicity coefficients in any p-norm can be bounded by the norm of a deflated matrix (Theorem 7.2).

We illustrate that two-norm coefficients can reproduce any singular value (Corollaries 6.20 and 7.8), and that their extreme values are determined by singular values (Theorem 7.9). We apply ergodicity coefficients to determine inclusion regions for subdominant eigenvalues of general complex matrices (Theorems 7.5 and 7.11) 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 7.13).

1.4. Notation. The elements of an $m \times n$ matrix A are denoted by a_{ij}, $1 \le i \le m$, $1 \le j \le n$, and the column space is $\text{range}(A) \equiv \{b : \ b = Ax \text{ for any } x \in \mathbb{C}^n\}$. The orthogonal complement of $\text{range}(A)$ in $\mathbb{R}^m$ or $\mathbb{C}^m$ is $\text{range}(A)^\perp$. The transpose of A is A^T, and the conjugate transpose is A^*. The identity matrix is I, and its columns are the canonical vectors e_i, $i \ge 1$.

For an $n \times 1$ column vector $x = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}^T$, the one-norm and infinity norms are, respectively,

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|, \qquad \|x\|_\infty = \max_{1 \le i \le n} |x_i|.$$

The componentwise inequality $x \ge 0$ ($x > 0$) means that all elements satisfy $x_i \ge 0$ ($x_i > 0$), while $|x| > 0$ means that all elements satisfy $x_i \neq 0$. The vector x is *stochastic* if $x \ge 0$ and $x^T \mathbb{1} = 1$, where $\mathbb{1} = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix}^T$. By $x_{1:k}$ we mean the $k \times 1$ vector $x_{1:k} = \begin{pmatrix} x_1 & \dots & x_k \end{pmatrix}^T$. The $n \times n$ diagonal matrix constructed from the $n \times 1$ vector x is denoted by

$$\text{diag}(x) \equiv \begin{pmatrix} x_1 & & \\ & \ddots & \\ & & x_n \end{pmatrix}.$$

2. General ergodicity coefficients for stochastic matrices.

We present a formal definition of ergodicity, and 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 (section 4 in [53], section 1 in [75]). *Let $\{S_k\}$ be a sequence of $n \times n$ stochastic matrices, $k \ge 1$, 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 $1 \le i, j, k \le n$ and $p \ge 0$,*

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

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 [23], Dobrushin [21], Dynkin [24], Sapogov [72, 73], and Sarymsakov [74]; these papers appeared between 1920 and 1950. More recent papers by Cohn [18, 19], Dobrushin [22], Hajnal [28, 29], Kingman [48], and Paz and Reichaw [63] are written in English and are more easily accessible. Excellent summaries can be found in Seneta's technical papers

[75, section 1], [76, section 1], his biography of Markov [84, sections 7–9], and his book [85, Chapters 3–4]. This revised printing of the 1981 version of Seneta's book [85] also contains a list of more recent papers on various aspects of ergodicity coefficients. Applications of ergodicity coefficients in the context of nonhomogeneous matrix products are presented in [33].

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

DEFINITION 2.2 (page 509 in [75]). *A* coefficient of ergodicity*, or* ergodicity coefficient*, is a continuous scalar function* $\mu(\cdot)$ *defined for stochastic matrices* S *that satisfies* $0 \leq \mu(S) \leq 1$*. 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: $\hat{\mu}(S) = 1 \iff \text{rank}(S) = 1$ [15], [29, section 2], [42, p. 56], [48, section 4], [75, section 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 [61, 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*

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

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

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 (page 136 in [85]). *Let* $\{S_k\}$ *be a sequence of* $n \times n$ *stochastic matrices,* $k \geq 1$*, and let* $T^{(p,r)} = S_{p+1}S_{p+2}\cdots S_{p+r}$*. The sequence* $\{S_k\}$ *is weakly ergodic if for all* $p \geq 0$

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

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

Example 2.5. Let S be a stochastic matrix. The following are proper ergodicity coefficients [85, p. 137],

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

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

The equalities for $\tau_1(S)$ will be proved in Theorem 3.7 and Corollary 3.9.

The following ergodicity coefficient is not proper [85, p. 137]

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

This is because the $n \times n$ stochastic matrix $S = \frac{1}{n}\mathbb{1}\mathbb{1}^T$ has $\text{rank}(S) = 1$ 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 [61], [75, section 1]. Below is one of the simpler relations.

THEOREM 2.6 (page 56 in [42], pages 137–138 in [85]). *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 let $\mathcal{P}_l$ be 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 we obtain

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

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 (section 2 in [76]). *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 section 3 for more details.

However, there are situations where Definitions 2.2 and 2.7 are not consistent [55]. 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 [55, 67, 68, 71, 76]. Two popular choices are presented below.

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

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

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 vectors $x > 0$ and $y > 0$ that are not multiples of each other, is called *Birkhoff's Contraction Coefficient* [85, sections 3.1 and 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 [85, sections 3.1 and 3.4], [33, section 2.2]. Hajnal [30, (7)] presents basic properties of τ_B, and Seneta [85, sections 3.1 and 3.4] derives explicit expressions. Hartfiel [33, sections 2.2 and 5] presents an in-depth treatment of τ_B with application to nonhomogeneous matrix products. Artzrouni and others have studied τ_B in the context of more general dynamical systems [3, 4, 5].

Ergodicity coefficients defined by vector norms. Norm-based coefficients appear as early as 1956 in a paper by Dobrushin [22, sections 1.4 and 1.5]. For an operator S derived from a transition probability function and a particular norm $\|\cdot\|$, Dobrushin 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 [68, section 1] credits Seneta [76, section 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., $\|x\|_p = \|x^T\|_p$ for all vectors x [27, 31, 32, 54, 55, 66, 68, 69, 70, 71, 76, 77, 79, 84, 88, 89]. 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 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 [76, section 2], [85, section 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" [47, 62, 88]. We start with two auxiliary results in section 3.1 for vectors whose components sum to zero, like those in (3.1), and then derive properties of $\tau_1(S)$ in section 3.2. In section 3.3 we show that $\tau_1(S)$ is identical to the explicit expressions in Example 2.5. An application of $\tau_1(S)$ to condition numbers of stationary distributions is discussed in section 3.4.

3.1. Vectors whose elements sum to zero.

We present several results for real vectors whose elements sum to zero, like those that define the maximum for $\tau_1(S)$ in (3.1). These results will be instrumental in deriving explicit expressions for $\tau_1(S)$ in section 3.3, and for $\tau_\infty(S)$ in section 4.1.

First we show that vectors whose elements sum to zero can be represented as linear combinations with nonnegative coefficients of vectors $e_i - e_j$.

LEMMA 3.1 (Lemma 2.4 in [85]). *If $x \in \mathbb{R}^n$ satisfies $x \neq 0$ and $x^T \mathbb{1} = 0$, then*

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

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

If $n = 2$, assume, without loss of generality, that $x_1 > 0$. 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

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

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,

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

The following bounds from [35] apply to inner products of real vectors where one of the vectors has elements summing to zero. We extend these bounds to complex vectors.

LEMMA 3.2 (Lemma (2.3) in [35]). *If $x, y \in \mathbb{C}^n$ and $x^T \mathbb{1} = 0$, then for any scalar θ*

$$|x^T y| \leq \|x\|_p \, \|y - \theta \mathbb{1}\|_q, \qquad \frac{1}{p} + \frac{1}{q} = 1.$$

Proof. The idea is to incorporate the constraint $x^T \mathbb{1} = 0$ into the inner product, by writing $x^T y = x^T(y - \theta \mathbb{1})$, and then applying Hölder's inequality. $\square$

Below are two special cases of Lemma 3.2 for particular choices of p, q, and θ. Seneta discusses the proper attribution of (3.2) in [84, sections 7.1 and 7.3], and refers to it as *Markov's Contraction Inequality*.

LEMMA 3.3 (Corollary (2.4) in [35]). *If $x, y \in \mathbb{C}^n$ and $x^T \mathbb{1} = 0$, then*

$$|x^T y| \leq \frac{\|x\|_1}{2} \, \max_{ij} |y_i - y_j|. \tag{3.2}$$

If y is real, with elements labeled in nonincreasing order $y_{(1)} \geq \cdots \geq y_{(n)}$, then

$$|x^T y| \leq \|x\|_\infty \, \phi(y), \tag{3.3}$$

where

$$\phi(y) = \begin{cases} \sum_{i=1}^{n/2} y_{(i)} - \sum_{i=n/2+1}^{n} y_{(i)} & \text{if } n \text{ is even,} \\ \sum_{i=1}^{(n-1)/2} y_{(i)} - \sum_{i=(n+3)/2}^{n} y_{(i)} & \text{if } n \text{ is odd.} \end{cases}$$

Proof. Inequality (3.2) follows from Lemma 3.2 with $p = 1$, $q = \infty$, and θ chosen as follows. Let k and l be such that $|y_k - y_l| = \max_{ij} |y_i - y_j|$. Then all points y_i lie in a ball with diameter $|y_k - y_l|$ and center $\theta = (y_k + y_l)/2$. The distance between any y_i and the center θ is bounded by the radius $|y_k - y_l|/2$. Hence $\|y - \theta \mathbb{1}\|_\infty = \max_i |y_i - \theta| \le |y_k - y_l|/2$.

Inequality (3.3) follows from Lemma 3.2 with $p = \infty$, $q = 1$, and $\theta = y_{(l)}$, where $l = n/2$ for n even, and $l = (n+1)/2$ for n odd. This gives, for even n and $l = n/2$,

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

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

$$\begin{aligned} \|y - \theta \mathbb{1}\|_1 &= \sum_{i=1}^{n} |y_{(i)} - y_{(l)}| = \sum_{i=1}^{l} (y_{(i)} - y_{(l)}) + \sum_{i=l+1}^{n} (y_{(l)} - y_{(i)}) \\ &= \sum_{i=1}^{l-1} y_{(i)} - \sum_{i=l+1}^{n} y_{(i)} = \phi(y). \qquad \square \end{aligned}$$

3.2. Properties. We show that $\tau_1(S)$ in (3.1) is a proper coefficient of ergodicity in the sense of Definition 2.2, that it is submultiplicative, and that it represents 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, $S\mathbb{1} = \lambda \mathbb{1} = \mathbb{1}$, 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 3.4 (section 4.3 in [85]). *If S, S_1, and S_2 are stochastic matrices, then*

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

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

$$0 \le \tau_1(S) = \max_{\|z\|_1=1, z^T \mathbb{1}=0} \left\|S^T z\right\|_1 \le \max_{\|z\|_1=1} \|S^T z\|_1 = \|S^T\|_1 = \|S\|_\infty = 1.$$

2. Let $\tau_1(S_1) \ge \tau_1(S_2)$, 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

$$0 \le \tau_1(S_1) - \tau_2(S_2) = \|S_1^T y\|_1 - \max_{\substack{\|z\|_1=1 \\ z^T \mathbb{1}=0}} \left\|S_2^T z\right\|_1 \le \left\|S_1^T y\right\|_1 - \left\|S_2^T y\right\|_1.$$

   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 $i \ne j$. 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. ∎

Remark 3.5. For stochastic matrices S and $S + E$, the proof Theorem 3.4 implies

$$|\tau_1(S + E) - \tau_1(S)| \le \|E\|_\infty.$$

This means, small changes in S produce only small changes in $\tau_1(S)$. In other words, $\tau_1(S)$ is well-conditioned in the absolute sense with respect to changes in S.

Since τ_1 is a proper ergodicity coefficient, Theorem 2.3 implies for a doubly stochastic matrix S_d that τ_1 is zero at the same time for S_d and S_d^T [61, Property 1],

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

Now we 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 3.6 (section 5.2 in [6], section 3.2 in [76], section 4 in [78]). *If S, S_1, and S_2 are stochastic matrices, then*

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

Proof.

1. If $\lambda \ne 1$ is a real eigenvalue of S, then there is a real left eigenvector v with $S^T v = \lambda v$ 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

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

If $\lambda \neq 1$ is a complex eigenvalue, then its eigenvectors v are complex, too. In this case the following more general, but less intuitive, proof applies [76, section 3.2], [78, section 4]. For any complex vector y define the function

$$f(y) \equiv \frac{1}{2} \max_{ij} |y_i - y_j| = \frac{1}{2} \max_{i,j} |(e_i - e_j)^T y|.$$

Applying f to the vector Sy gives

$$f(Sy) = \frac{1}{2} \max_{ij} |(e_i - e_j)^T Sy| = \frac{1}{2} |(e_k - e_l)^T Sy|$$

for some k and l. Since $(e_k - e_l)^T S\mathbb{1} = 0$, (3.2) implies

$$\frac{1}{2} |(e_k - e_l)^T Sy| \leq \frac{\|S^T(e_k - e_l)\|_1}{2} \frac{\max_{ij} |y_i - y_j|}{2} \leq \tau_1(S)\, f(y),$$

where the last inequality follows from the expression for τ_1 in (3.1) and the fact that $(e_k - e_l)^T \mathbb{1} = 0$. Therefore, for any real or complex vector y we have

$$f(Sy) \leq \tau_1(S)\, f(y). \tag{3.4}$$

Now let v be a right eigenvector associated with an eigenvalue $\lambda \neq 1$ of S, so that $Sv = \lambda v$. Then

$$f(v) = \frac{1}{2} \max_{ij} |(e_i - e_j)^T v| = \frac{1}{2} |(e_k - e_l)^T v|$$

for some k and l, and

$$\begin{aligned} f(Sv) &= \frac{1}{2} \max_{ij} |(e_i - e_j)^T Sv| \geq \frac{1}{2} |(e_k - e_l)^T Sv| = \frac{|\lambda|}{2} |(e_k - e_l)^T v| \\ &= |\lambda|\, f(v). \end{aligned}$$

Hence $|\lambda|\, f(v) \leq f(Sv)$. Combining this with (3.4) gives

$$|\lambda|\, f(v) \leq \tau_1(S)\, f(v).$$

Since v is a right eigenvector of S associated with an eigenvalue $\lambda \neq 1$, v cannot be a multiple of $\mathbb{1}$, hence $f(v) \neq 0$. Dividing the above inequality by $f(v)$ gives $|\lambda| \leq \tau_1(S)$. An alternative proof in [6, section 5.2] is based on constructing a seminorm on $\mathbb{C}^n$ that is equal to $\tau_1(S)$.

2. Let y be a vector with $\tau_1(S_1S_2) = \| (S_1S_2)^T y\|_1$, $\|y\|_1 = 1$, 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

$$\begin{aligned} \tau_1(S_1S_2) &= \left\|(S_1S_2)^T y\right\|_1 = \left\|S_2^T S_1^T y\right\|_1 = \left\|S_1^T y\right\|_1 \left\|S_2^T x\right\|_1 \\ &\leq \tau_1(S_1)\, \tau_1(S_2). \qquad \square \end{aligned}$$

The interesting feature of Theorem 3.6 is that the magnitude of complex eigenvalues can be bounded by an expression that is maximized over real vectors. Seneta [81, p. 191] credits the submultiplicative property to Dobrushin [22]; it also appears in a 1975 paper by Kingman [48, (4.10)].

Kirkland and Neumann [50] characterize classes of irreducible stochastic matrices S that have subdominant eigenvalues $\lambda \neq 1$ for which equality holds in Theorem 3.6. In their work, as in section 1.1, we find a connection to the Google matrix. Kirkland and Neumann transform the matrices S so that they have constant row sums, and they do this by adding to S a suitable matrix of rank one. The resulting matrix has the form of a Google matrix.

3.3. Explicit expressions. We show that $\tau_1(S)$ 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 [57, pp. 358-359], [56, 58]. 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 [83, section 5], [84, section 7]. On the occasion of Markov's 150th birthday in 2006, Seneta offers a fascinating account of Markov's life and legacy [84, sections 1–6].

In the context of the Central Limit Theorem for Markov chains [22, 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" [84, p. 10]. Other authors, such as Kirkland and Neumann [50, section 1], credit Deutsch and Zenger [20] with this bound. Paz [62] refers to the expression $1 - \min_{ij} \sum_k \min\{s_{ik}, s_{jk}\}$ as the δ-coefficient, which was later adopted by other authors, including Tan [89] and Rhodius [66, 69, 70].

We show that the first expression for $\tau_1(S)$ in Example 2.5 is identical to the expression in (3.1), i.e.,

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

THEOREM 3.7 (sections 3.1 and 4.3 in [85]). *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 3.1 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

$$\tau_1(S) = \left\|S^T x\right\|_1 \leq \sum_{i\neq j} \frac{y_{ij}}{2} \left\|S^T (e_i - e_j)\right\|_1 \leq \frac{1}{2} \max_{ij} \left\|S^T (e_i - e_j)\right\|_1. \tag{3.5}$$

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

Theorem 3.7 has several consequences. First 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 3.8. *If $S \in \mathbb{R}^{n\times n}$ is a stochastic matrix, then*

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

Proof. In the expression for $\tau_1(S)$ from Theorem 3.7 write

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

Another consequence of Theorem 3.7 is the second expression for $\tau_1(S)$ from Example 2.5.

COROLLARY 3.9 (sections 3.1 and 4.3 in [85], pages 1733–1734 in [41]). *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.[3] Any two real numbers s and t satisfy

$$\min\{s,t\} = \frac{1}{2}\left(|s+t| - |s-t|\right).$$

Applying this to the expression in Theorem 3.7 yields

$$\frac{1}{2}\sum_{k=1}^n |s_{ik} - s_{jk}| = \frac{1}{2}\sum_{k=1}^n (s_{ik} + s_{jk}) - \sum_{k=1}^n \min\{s_{ik}, s_{jk}\} = 1 - \sum_{k=1}^n \min\{s_{ik}, s_{jk}\},$$

where the last equality follows because the elements in each row of S sum to one. Taking the maximum over i and j gives the desired expression. ∎

In place of the above expression, several authors choose as a coefficient ergodicity instead the negative version $1 - \tau_1(S) = \min_{ij}\sum_{k=1}^n \min\{s_{ik}, s_{jk}\}$ [22, 29, 41, 46, 63, 65].

In many situations, $\tau_1(S)$ is only of interest when it is less than 1. Eigenvalue bounds, and bounds on the condition number of the stationary distribution as in Theorem 3.14 in the next section, are just two examples. Corollary 3.9 implies that $\tau_1(S) < 1$ is possible if and only if any two rows of S "intersect" by having a positive element in a corresponding position. Such matrices have their own name.

DEFINITION 3.10 (section 2 in [29], section 3.2 in [75], page 82 in [85]). *A stochastic matrix S is scrambling if any two rows share some column in which they both have a positive element.*

What can we say about scrambling matrices? Scrambling matrices are a subset of primitive stochastic matrices [29, 85]. Therefore matrices with a single element in each column, such as permutation matrices, cannot be scrambling. An extreme example is Markov matrices, which have at least one entirely positive column.

DEFINITION 3.11 (section 3.2 in [75]). *A stochastic matrix S is a Markov matrix if* $\max_j\left(\min_i s_{ij}\right) > 0$.

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

---

[3] We thank one of the reviewers for simplifying this proof.

THEOREM 3.12 (Lemma 2 in [29], section 4 in [76]). *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, $\tau_1(SQ) \le \tau_1(S)\tau_1(Q)$ in Theorem 3.6. ∎

THEOREM 3.13 (section 3.2 in [75]). *If S and Q are stochastic matrices and one of them is a Markov matrix, then SQ and QS are also Markov matrices.*

More examples and properties of scrambling matrices can be found in [33, section 4.1].

Next we derive a perturbation bound for the stationary distribution of scrambling matrices.

3.4. Condition number for the stationary distribution. We show that $\tau_1(S)$ yields a bound for the normwise condition number of the stationary distribution of a scrambling matrix.

THEOREM 3.14 (section 2 in [79]). *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 S is scrambling so that $\tau_1(S) < 1$, then*

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

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

$$(\widehat{\pi} - \pi)^T (I - S) = \widehat{\pi}^T E. \tag{3.6}$$

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 (3.6) 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 $(I - Q)^{-1} = \sum_{i=0}^{\infty} Q^i$ [59, p. 126]. Thus

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

Taking norms and applying the triangle inequality gives

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

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 τ_1 in Theorem 3.6 implies

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

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

Theorem 3.14 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 [79] appeared in 1988, several tighter bounds for the condition number of π have been derived [17, section 4], [49], as have optimal condition numbers in terms of ergodicity coefficients applied to the group inverse of $I - S$ [52]. In [84, section 9] Seneta surveys more recent perturbation results, including those by Cho and Meyer [16, 17], Kirkland, Neumann, and Shader [51], Haviv and Van der Heyden [35], Hunter [39, 40], and himself [80, 82].

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 $z \in \mathbb{R}^n$ [76, section 2]. We present properties and explicit expressions for $\tau_\infty(S)$ in section 4.1, and exhibit relations between $\tau_\infty(S)$ and $\tau_1(S)$ in section 4.2.

Unlike $\tau_1(S)$ which is bounded above by 1, $\tau_\infty(S)$ has no fixed upper bound that is independent of the dimension of the matrix S. 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 an $n \times n$ 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 [88, p. 861] and a 62×62 matrix in [76, section 3].

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

THEOREM 4.1 ((7) and (8) in [88]). *If S, S_1, and S_2 are stochastic matrices, then*

1. $0 \leq \tau_\infty(S) \leq \|S\|_1$,
2. $|\tau_\infty(S_1) - \tau_\infty(S_2)| \leq \tau_\infty(S_1 - S_2)$,
3. $\tau_\infty(S) = 0$ *if and only if* $\operatorname{rank}(S) = 1$,
4. $\tau_\infty(S_1 S_2) \leq \tau_\infty(S_1)\, \tau_\infty(S_2)$.

Proof. The proofs are analogous to those of Theorems 3.4 and 3.6, as shown in [88]. □

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,

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

We derive an explicit expression for $\tau_\infty(S)$ first, before we show it is a bound on the subdominant eigenvalues.

THEOREM 4.2 (section 2 in [88]). *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 in Lemma 3.3.

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 inequality (3.3) implies $|x^T s| \le \phi(s) \le \max_{1\le i\le n} \phi(Se_i)$.

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

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

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

With the help of Theorem 4.2 we can now show that $\tau_\infty(S)$ is an eigenvalue bound.

THEOREM 4.3 (section 3 in [78]). *If S is a stochastic matrix, then $|\lambda| \le \tau_\infty(S)$ for all eigenvalues $\lambda \ne 1$ of S.*

Proof. The idea is to show that maximizing over complex vectors in $\tau_\infty(S)$ gives the same value as maximizing over real vectors.

Let z be any complex vector with $z^T\mathbb{1} = 0$ and $\|z\|_\infty = 1$. Since the infinity norm is the maximal row sum, $\|S^T z\|_\infty = |e_j^T S^T z|$ for some j. From (3.3) follows $|(Se_j)^T z| \le \|z\|_\infty \phi(Se_j) = \phi(Se_j)$, where the right-hand side is independent of z. Theorem 4.2 implies

$$\max_{z\in\mathbb{C}^n, \|z\|_\infty=1, z^T\mathbb{1}=0} \|S^T z\|_\infty \le \phi(Se_j) \le \tau_\infty(S).$$

If v is a possibly complex left eigenvector associated with an eigenvalue $\lambda \ne 1$, so that $S^T v = \lambda v$ and $\|v\|_\infty = 1$, then $v^T\mathbb{1} = 0$ and

$$|\lambda| = |\lambda|\ \|v\|_\infty = \|S^T v\|_\infty \le \max_{z\in\mathbb{C}^n, \|z\|_\infty=1, z^T\mathbb{1}=0} \|S^T z\|_\infty \le \tau_\infty(S). \qquad ☐$$

Like $\tau_1(S)$ in Corollary 3.8, we can also view $\tau_\infty(S)$ as the norm of an (oblique) projection of S, with the projection being onto $\operatorname{range}(\mathbb{1})^\perp$.

COROLLARY 4.4. *If $S \in \mathbb{R}^{n\times n}$ is a stochastic matrix, then for some $1 \le k \le n$,*

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

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

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

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 4.5 (Proposition 4 in [88]). *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 on $\tau_\infty(S)$. 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 $\alpha(S) = |s_{i_1 l} - s_{i_2 l}|$ for some i_1 and i_2. Adding and subtracting $s_{k_l l}$ inside $\alpha(S)$ and applying the triangle inequality gives

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

where the last inequality follows from the proof of Theorem 4.2, since we showed there 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 on $\tau_\infty(S)$, let column j assume the maximum in $\tau_\infty(S)$, and assume that the rows of S have been permuted so that $s_{1j} \ge \cdots \ge s_{nj}$. 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,

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

Theorem 4.5 implies the value in Corollary 4.6 for the maximum of $\tau_\infty(S)$ over all stochastic matrices S. Rhodius [66] and Lešanovský [55] attribute this result to Tan [88]. In addition, we characterize the class of all stochastic matrices that achieve this maximal value.

COROLLARY 4.6. *If $\mathcal{S}_n$ is the set of stochastic matrices in $\mathbb{R}^{n\times n}$, then*

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

The matrices that achieve this bound must be row or column permutations of the two matrices below for even and odd n, respectively,

$$S_e = \begin{pmatrix} \mathbb{1}_{n/2} & 0 \\ 0 & \hat{S}_e \end{pmatrix}, \qquad S_o = \begin{pmatrix} \mathbb{1}_{(n-1)/2} & 0 \\ 0 & \hat{S}_o \end{pmatrix},$$

where $\hat{S}_e$ and $\hat{S}_o$ are stochastic[4] *matrices of dimension $(n/2) \times (n-1)$ and $((n+1)/2) \times (n-1)$, respectively.*

Proof. Theorems 2.6 and 3.4 imply $\alpha(S) \leq \tau_1(S) \leq 1$. Together with Theorem 4.5 this gives $\max_{S\in\mathcal{S}_n} \tau_\infty(S) \leq n/2$ for even n, and $\max_{S\in\mathcal{S}_n} \tau_\infty(S) \leq (n-1)/2$ for odd n.

Applying the explicit expression for τ_∞ from Theorem 4.2 to their leading column shows that S_e and S_o achieve these bounds; and so do row or column permutations of S_e and S_o, because τ_∞ is invariant under permutations.

To show that all matrices must have this form in order to achieve the maximal value for τ_∞, consider the case of even n, and the explicit expression for τ_∞ from Theorem 4.2 with the notation from Lemma 3.3. A matrix S with $\tau_\infty(S) = n/2$ must have have at least one column s with $\sum_{i=1}^{n/2} s_{(i)} - \sum_{i=n/2+1}^{n} s_{(i)} = n/2$. Hence $\sum_{i=1}^{n/2} s_{(i)} \geq n/2$. But the elements of S are bounded above by 1, so that $\sum_{i=1}^{n/2} s_{(i)} \leq n/2$. The two inequalities imply $\sum_{i=1}^{n/2} s_{(i)} = n/2$ and $\sum_{i=n/2+1}^{n} s_{(i)} = 0$. Hence the $n/2$ largest elements of s must be equal to 1, while the remaining $n/2$ elements must be equal to 0. Since the elements in each row of S sum to one, the $n/2$ rows of S containing the nonzero elements of s must have zero elements everywhere else. The argument for odd n is analogous. ∎

The maximal value for $\tau_\infty(S)$ in Corollary 4.6 also follows from a more general result for p-norms in Theorem 5.9.

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

THEOREM 4.7 (section 3.3 in [76]). *If S is a 2×2 or 3×3 stochastic matrix, then $\tau_1(S) = \tau_\infty(S)$.*

Proof. For $n = 2$ Theorem 3.7 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 3.7 implies $\tau_1(S) = \frac{1}{2}\max_{1\leq i,j\leq 3}\left\|S^T(e_i - e_j)\right\|_1$, while Theorem 4.2 implies $\tau_\infty(S) = \max_{1\leq i,j\leq 3}\left\|S^T(e_i - e_j)\right\|_\infty$. 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 $|v_1| \geq |v_2|$, so that $|-v_1 - v_2| = |v_1| - |v_2|$. Then $\|v\|_1 = 2|v_1|$ and $\|v\|_\infty = |v_1|$, hence $\tau_1(S) = \tau_\infty(S)$. ∎

[4] We can extend the definition of stochasticity to a rectangular matrix Z, by requiring that Z have all elements in $[0, 1]$ and satisfy $Z\mathbb{1} = \mathbb{1}$.

For stochastic matrices of larger dimensions, the following relations hold.

THEOREM 4.8. *If $S \in \mathbb{R}^{n\times n}$ is a stochastic matrix and $n \geq 2$, then*

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

Proof. We start with the lower bound on $\tau_\infty(S)$. In Theorem 3.7, let $i \neq j$ be indices so that $2\tau_1(S) = \left\|S^T(e_i - e_j)\right\|_1 \leq n\|S^T(e_i - e_j)\|_\infty$. From $(e_i - e_j)^T\mathbb{1} = 0$ and $\|e_i - e_j\|_\infty = 1$ follows $\|S^T(e_i - e_j)\|_\infty \leq \tau_\infty(S)$.

As for the upper bound on $\tau_\infty(S)$, Theorem 4.5 implies $\tau_\infty(S) \leq \frac{n}{2}\alpha(S)$ for even n and $\tau_\infty(S) \leq \frac{n-1}{2}\alpha(S)$ for odd n. Now combine this with the bound $\alpha(S) \leq \tau_1(S)$ from Theorem 2.6. ∎

The upper bound in Theorem 4.8 is tight for the matrices S_e and S_o from Corollary 4.6.

Doubly stochastic matrices. Theorem 4.8 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 are 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 4.9 (page 344 in [77]). *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 3.9 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 contains at least $n/2$ zeros. If we label the elements of c so that $c_1 \geq \cdots \geq c_n$, 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 4.9 is not true [77, 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 [64].

THEOREM 4.10 (page 345 in [77]). *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, $s_1 \geq \cdots \geq s_n$, 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 3.9 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

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

Corollary 4.11 (page 345 in [77]). *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 4.11, one can show that equality holds, i.e., $\tau_1(S_d) = \tau_\infty(S_d)$, for $n = 2, 3, 4$ [77, 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 [76, section 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 (section 1 in [89], [66]). *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. $\tau_p(S_1 S_2) \leq \tau_p(S_1)\, \tau_p(S_2)$.
5. *If S is irreducible and 1 is the only eigenvalue of modulus 1, then $|\lambda| \leq \tau_p(S)$ for all eigenvalues $\lambda \neq 1$.*

Proof. The proofs for items 1–4 are analogous to those of Theorems 3.4 and 3.6. The bound for item 5 follows from Theorem 6.21. $\square$

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 (page 585 in [76]). *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 $\lambda = s_1 - s_2$. All 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. ☐

In 1988 Rhodius [66] 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 $\max_z \|Q^T z\|_p$ as a function of p and the matrix dimension n. We illustrate this development.

To start with, we present two compactness results.

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

Proof. If S_1 and S_2 are stochastic matrices, then so is $\gamma S_1 + (1-\gamma) S_2$ for any $0 \leq \gamma \leq 1$. This is because the elements of $\gamma S_1 + (1-\gamma) S_2$ are in $[0,1]$ and

$$(\gamma S_1 + (1-\gamma) S_2)\, \mathbb{1} = \gamma S_1 \mathbb{1} + (1-\gamma) S_2 \mathbb{1} = \gamma \mathbb{1} + (1-\gamma) \mathbb{1} = \mathbb{1}.$$

Hence the set $\mathcal{S}_n$ is convex.

To show compactness of $\mathcal{S}_n$, we establish that it is bounded and closed. The matrix Hölder inequality [36, (6.19)] implies for a matrix $S \in \mathcal{S}_n$ that

$$\|S\|_p \leq \|S\|_1^{1/p}\, \|S\|_\infty^{1-1/p} = \|S\|_1^{1/p} \leq n^{1/p},$$

since $\|S\|_\infty = 1$, and the elements of S are bounded by 1. Hence $\mathcal{S}_n$ is bounded.

To show that $\mathcal{S}_n$ is closed, let $\{S_k\}$ be a sequence of stochastic matrices that converges to a matrix Z, i.e., $\|S_k - Z\|_p \to 0$ as $k \to \infty$. This implies componentwise convergence $(S_k)_{ij} \to Z_{ij}$. Since the elements $(S_k)_{ij}$ are in the closed interval $[0,1]$, Z_{ij} must be in $[0,1]$. It remains to assert that the elements in each row of Z sum to 1. The convergence of $\{S_k\}$ implies that for every $\epsilon > 0$ there exists a k so that

$$\epsilon > \|S_k - Z\|_p \geq \frac{\|(S-Z)\mathbb{1}\|_p}{\|\mathbb{1}\|_p} = n^{-p} \|\mathbb{1} - Z\mathbb{1}\|_p.$$

Hence $Z\mathbb{1} = \mathbb{1}$. We have shown that the limit of a converging sequence of stochastic matrices is also stochastic, so that $\mathcal{S}_n$ is closed. The compactness of $\mathcal{S}_n$ follows because a closed and bounded set in a finite dimensional vector space is compact.

An $n \times 1$ canonical vector e_j has $n-1$ zero elements, so that it cannot be expressed as a linear combination of other stochastic vectors. Therefore matrices in $\mathcal{S}_n$ whose rows are (transposes of) canonical vectors are extreme points of $\mathcal{S}_n$. ☐

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.*

Proof. We show that $\mathcal{H}_n$ is closed and bounded. Compactness of $\mathcal{H}_n$ then follows because $\mathcal{H}_n$ is a subset of a finite dimensional vector space.

The set $\mathcal{H}_n$ is bounded because $\|x\|_p = 1$. To show that $\mathcal{H}_n$ is closed, let $\{x_k\}$ be a sequence of vectors in $\mathcal{H}_n$ that converges to some vector z, i.e., $\|x_k - z\|_p \to 0$ as $k \to \infty$. Thus, for every $\epsilon > 0$ there exists a k so that

$$\epsilon > \|x_k - z\|_p = \left\|(x_k - z)^T\right\|_q \geq \frac{\left|(x_k - z)^T \mathbb{1}\right|}{\|\mathbb{1}\|_q} = \frac{|z^T \mathbb{1}|}{\|\mathbb{1}\|_q} = n^{-q} |z^T \mathbb{1}|,$$

where $\frac{1}{p} + \frac{1}{q} = 1$. Thus $z^T \mathbb{1} = 0$ and $z \in \mathcal{H}$. This means, the limit of a converging sequence of vectors in $\mathcal{H}_n$ is also in $\mathcal{H}_n$. Therefore $\mathcal{H}_n$ is closed. ☐

Now we are ready for the important results. First we establish that in order to determine the maximal τ_p for all stochastic matrices, it suffices to look at the extreme points of $\mathcal{S}_n$.

THEOREM 5.5 (Theorem 1(a) in [66]). *For integers* $p \geq 1$

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

Proof. Since $\mathcal{S}_n$ and $\mathcal{H}_n$ are compact sets, as was established in 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 [37, 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 5.6 (page 142 in [66]). 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 $\max_{S\in\mathcal{S}_n} \tau_p(S)$ can be obtained by summing only the positive elements in those vectors that achieve the maximum.

COROLLARY 5.7 (page 144 in [66]). *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. Theorem 5.5 and Remark 5.6 imply

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

where the maximum ranges over all sets D_j that satisfy $D_1 \cup \cdots \cup D_n = \{1, \ldots, n\}$, and $D_i \cap D_j = \emptyset$ for $i \neq j$. Define

$$R \equiv \max_{1 \le l \le n-1} \max_{z \in \mathcal{H}_n \cap \mathcal{A}_l} 2^{1/p}(z_1 + \cdots + z_l).$$

Since Rhodius [66] did not provide a proof for $L = R$, we give one below.

First we show that $L \geq R$. The vectors appearing in R are vectors $z \in \mathcal{H}_n \cap \mathcal{A}_l$ that satisfy

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

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

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

From $\|x_{1:k}\|_1 = \|x_{k+1:n}\|_1$ follows

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

Now we show that $L \leq R$. Let $x \in \mathcal{H}_n$ be a vector that attains the maximum in L, and let k be the number of positive elements in x. Then we can label the elements of x so that $x_1, \ldots, x_k > 0$ and $x_{k+1}, \ldots, x_n \leq 0$, which means $x \in \mathcal{A}_k$. As above, $x \in \mathcal{H}_n$ implies that $\|x_{1:k}\|_1 = \|x_{k+1:n}\|_1$. Consider the elements of the vector in L, and split the sums into their positive and negative parts,

$$\sum_{i \in D_j} x_i = p_j - n_j, \qquad \text{where} \quad p_j \equiv \sum_{i \in D_j, x_i > 0} x_i, \quad n_j \equiv -\sum_{i \in D_j, x_i \le 0} x_i,$$

and $p_j = 0$ or $n_j = 0$ if the corresponding set is empty. Then $\|x_{1:k}\|_1 = \sum_{j=1}^n p_j = \sum_{j=1}^n n_j$. From $p_j \geq 0$ and $n_j \geq 0$ follows $|p_j - n_j| \leq \max\{p_j, n_j\}$, hence

$$|p_j - n_j|^p \le (\max\{p_j, n_j\})^p \le p_j^p + n_j^p.$$

Applying these inequalities to the pth power of L gives

$$\begin{aligned} L^p &= \sum_{j=1}^n \left| \sum_{i \in D_j} x_i \right|^p = \sum_{j=1}^n |p_j - n_j|^p \le \sum_{j=1}^n p_j^p + \sum_{j=1}^n n_j^p \\ &\le \left(\sum_{j=1}^n p_j \right)^p + \left(\sum_{j=1}^n n_j \right)^p = 2\|x_{1:k}\|_1^p \le R^p. \qquad \square \end{aligned}$$

Next we characterize vectors z that achieve the maximum in Corollary 5.7 and show that their elements $z_1, \ldots, z_l$ can be chosen to be all the same.

Theorem 5.8 (Theorem 2 in [66]). *For integers $p \geq 1$, 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

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

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 $f(y/\|y\|_p, l) \leq f(x,l)$. Suppose we can show that $\|y\|_p \leq 1$. Combined with $f(x,l) = f(y,l)$ this implies

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

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

We still need to show that $\|y\|_p \leq 1$. 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

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

Similarly, $|y_{1+l}| \leq (n-l)^{-1/p}\|x_{l+1:n}\|_p$, 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 that $\|y\|_p \leq \|x\|_p = 1$. ☐

At last, the characterization of the vectors in Theorem 5.8 makes it possible to determine explicit values for the maximum in Corollary 5.7.

THEOREM 5.9 (Theorem 3 in [66]). *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 5.7 and Theorem 5.8 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 lz_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 $lz_1 + (n-l)z_{l+1} = 0$, hence $z_{l+1} = -lz_1/(n-l)$. Substituting this into $\|z\|_p^p = 1$ gives $lz_1 = \left(l^{1-p} + (n-l)^{1-p}\right)^{-1/p}$. 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 5.9 reduces to Corollary 4.6.

6. Ergodicity coefficients for real matrices. In 1984 Seneta [78, (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 general properties and bounds in section 6.1, and explicit expressions for coefficients defined by different norms: for the one-norm in section 6.2, for the infinity-norm in section 6.3, and for the two-norm in section 6.4. Eigenvalue bounds for nonnegative matrices are discussed in section 6.5.

6.1. 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 6.1. *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 $\max_{\substack{\|z\|_p=1\\ z^Tw=0}} \left\|A^Tz\right\|_p \le \max_{\|z\|_p=1} \left\|A^Tz\right\|_p = \left\|A^T\right\|_p$.
2. Let $\tau_p(w,A_1) \ge \tau_p(w,A_2)$, and $\tau_p(w,A_1) = \left\|A_1^Ty\right\|_p$ for some vector $y \in \mathbb{R}^m$ with $\|y\|_p = 1$ and $y^Tw = 0$. Then
$$0 \le \tau_p(w,A_1) - \tau_p(w,A_2) = \left\|A_1^Ty\right\|_p - \max_{\substack{\|z\|_p=1\\ z^Tw=0}} \left\|A_2^Tz\right\|_p \le \left\|A_1^Ty\right\|_p - \left\|A_2^Ty\right\|_p.$$
The triangle inequality implies
$$\begin{aligned}\left\|A_1^Ty\right\|_p - \left\|A_2^Ty\right\|_p &\le \left\|(A_1-A_2)^T y\right\|_p \le \max_{\substack{\|z\|_p=1\\ z^Tw=0}} \left\|(A_1-A_2)^T z\right\|_p \\ &= \tau_p(w,A_1-A_2).\end{aligned}$$
3. Let $y \in \mathbb{R}^m$ be a vector with $\tau_p(w,BA) = \|(BA)^T y\|_p$, $y^Tw = 0$, and $\|y\|_p = 1$. The submultiplicative property of the p-norms implies
$$\begin{aligned}\tau_p(w,BA) &= \|A^TB^Ty\|_p \le \|A^T\|_p\|B^Ty\|_p \\ &\le \|A^T\|_p \max_{\substack{\|z\|_p=1\\ z^Tw=0}} \|B^Tz\|_p = \|A^T\|_p\, \tau_p(w,B). \qquad \square\end{aligned}$$

If w happens to be a real eigenvector of A, then a submultiplicative property holds for powers of A.

THEOREM 6.2. *Let $A \in \mathbb{R}^{n\times n}$ and $w \in \mathbb{R}^n$ be a right eigenvector of A. Then for $l, m \ge 1$*
$$\tau_p\left(w, A^{l+m}\right) \le \tau_p\left(w, A^l\right) \tau_p\left(w, A^m\right).$$

Proof. Let $y \in \mathbb{R}^n$ be a vector with $\tau_p\left(w, A^{l+m}\right) = \|\left(A^T\right)^{l+m} y\|_p$, $y^Tw = 0$, and $\|y\|_p = 1$. Since $Aw = \lambda w$ for some real number λ, we have
$$\left[(A^T)^m y\right]^T w = y^TA^mw = \lambda^m y^Tw = 0.$$
Hence the vector $x = \left(A^T\right)^m y / \left\|\left(A^T\right)^m y\right\|_p$ satisfies $\|x\|_p = 1$ and $x^Tw = 0$, so that
$$\tau_p\left(w, A^{l+m}\right) = \left\|\left(A^T\right)^{l+m} y\right\|_p = \left\|\left(A^T\right)^{l} x\right\|_p \left\|\left(A^T\right)^{m} y\right\|_p \le \tau_p\left(w, A^l\right) \tau_p\left(w, A^m\right). \qquad \square$$

The special case of Theorems 6.1 and 6.2 for stochastic matrices and their stationary distributions was shown in [89, p. 279].

Bounds. We present two upper bounds for $\tau_p(w,A)$ that could possibly improve the bound in Theorem 6.1. They also furnish eigenvalue bounds for irreducible nonnegative matrices in section 6.5.

The first bound expresses the constraint $z^T w = 0$ in terms of a rank-one downdate of the matrix. The bounds for τ_2 below involve the Frobenius norm, which is defined as $\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}$.

THEOREM 6.3 (Theorem 5.5 in [71]). *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 $p = 2$ we have

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

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

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

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

This proves the first bound.

For the case $p = 2$, use $\|A^T\|_2 = \|A\|_2$ and $\|A\|_2 \leq \|A\|_F$ to conclude that

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

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 $\hat{x}_i = \left(w^T w\right)^{-1} w^T A e_i = e_i^T A^T w / \|w\|_2^2$. Thus $\hat{x} \equiv A^T w / \|w\|_2^2$, and

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

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 6.4 (Lemma 4.1 in [27]). *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 $v^T f = w^T f = 1$*

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

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 $(I_m - f v^T) y = y - v^T y\, f$.

If $y - v^T y\, f = 0$, then $y = v^T y f$. 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

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

so that the desired inequality holds.

If $y - v^T y\, f \neq 0$ 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

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

The triangle and Hölder inequalities imply

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

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 6.1. ∎

Theorem 6.4 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 6.5 (Corollary 4.2 in [27]). *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* $v^T f = w^T f = 1$

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

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

6.2. 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 6.6 ([2, 65], page 584 in [76], pages 189–191 in [81]). *If* $A \in \mathbb{R}^{n\times n}$ *with* $A\mathbb{1} = a\mathbb{1}$, *then*

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

Proof. This is an extension of the explicit expression for $\tau_1(S)$ for stochastic matrices S in Theorem 3.7 and Corollary 3.9, and a special case of Theorem 6.8 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 3.1.

LEMMA 6.7 (page 192 in [78]). *If* $x, w \in \mathbb{R}^n$ *with* $x \neq 0$, $|w| > 0$, *and* $x^T w = 0$, *then*

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

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

Proof. The idea is to view the inner product $x^T w$ as an inner product involving the vector $\mathbb{1}$, and then to apply Lemma 3.1. To this end write

$$0 = x^T w = x^T(D\mathbb{1}) = \tilde{x}^T\,\mathbb{1}, \qquad \text{where} \quad \tilde{x} \equiv Dx.$$

Applying Lemma 3.1 to $\tilde{x}$ gives

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

Multiplying by D^{-1} gives the desired linear combination

$$x = \sum_{i\neq j} y_{ij}\,\frac{D^{-1}(e_i - e_j)}{\|D^{-1}(e_i - e_j)\|_1}, \qquad \text{where} \quad y_{ij} \equiv \frac{\tilde{y}_{ij}}{2}\,\|D^{-1}(e_i - e_j)\|_1.$$

It remains to show that $\sum_{i\neq j} y_{ij} = \|x\|_1$. This is done by induction over n.

For $n = 2$ the proof of Lemma 3.1 implies $\tilde{y}_{12} = 2\tilde{x}_1$, where $\tilde{x}_1 = w_1 x_1 > 0$. From $|w_1 x_1| = |w_2 x_2|$ follows

$$y_{12} = \frac{\tilde{y}_{12}}{2}\left(\frac{1}{|w_1|} + \frac{1}{|w_2|}\right) = |x_1| + \left|\frac{w_2 x_2}{w_2}\right| = \|x\|_1.$$

For the induction step assume that $\tilde{x} = Dx$ has been permuted so that $w_n x_n > 0$, $w_{n+1}x_{n+1} < 0$, and $w_n x_n = \max_{1\leq i\leq n+1}|w_i x_i|$. We write

$$x = \hat{x} - w_{n+1}x_{n+1}\,D^{-1}(e_n - e_{n+1}), \qquad \text{where} \quad \hat{x} \equiv \begin{pmatrix} x_{1:n-1} \\ x_n + \frac{w_{n+1}x_{n+1}}{w_n} \\ 0 \end{pmatrix}$$

and $\hat{x}^T w = x^T w = 0$. If $\hat{x} = 0$, then the conclusion follows as in the case $n = 2$. If $\hat{x} \neq 0$, apply the induction hypothesis to the leading n elements of $\hat{x}$,

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

Set $y_{n,n+1} \equiv -w_{n+1} x_{n+1} \, \|D^{-1}(e_n - e_{n+1})\|_1$ and use the definition of $\hat{x}$ to obtain

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

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 $\sum_{i \neq j} y_{ij} + y_{n,n+1} = \|x\|_1$. ▯

When $w = \mathbb{1}$, then $D = I$ and $\|D^{-1}(e_i - e_j)\|_1 = 2$, so that Lemma 6.7 reduces to Lemma 3.1.

As in Theorem 3.7, we make use of Lemma 6.7 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 vectors w with all nonzero elements, extends Theorem 3.7. from stochastic to real matrices.

Theorem 6.8 (page 193 in [78]). *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 6.7 to x gives $x = \sum_{i \neq j} y_{ij} \, D^{-1} \left(e_i - e_j \right) / \left\| D^{-1} \left(e_i - e_j \right) \right\|_1$, where $\sum_{i \neq j} y_{ij} = \|x\|_1 = 1$ and $D = \operatorname{diag}(w)$. 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,

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

If A is a stochastic matrix and $w = \mathbb{1}$, then Theorem 6.8 reduces to Theorem 3.7. If A is also irreducible, then all elements of its stationary distribution π are nonzero, and we obtain the bound below for $w = \pi$.

THEOREM 6.9 (Remark, page 284 in [89]). *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 $D = \operatorname{diag}(\pi)$.

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 6.11 from stochastic to real matrices.

THEOREM 6.10 (page 194 in [78]). *Let $A \in \mathbb{R}^{m\times n}$, $w \in \mathbb{R}^m$, and P be 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 6.7 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

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

Now Theorem 6.8 and $\|x_k\|_1 + \|x_{m-k}\|_1 = \|x\|_1 = 1$ imply

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

The reverse inequality follows, as in the proof of Theorem 6.8, 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 $1 \le i, j \le k$, and whose trailing $n-k$ elements are zero. □

Tan extends the explicit expressions for $\tau_1(\pi, S)$ for irreducible stochastic matrices in Theorem 6.9 to the larger class of stochastic matrices S that have only a single eigenvalue of modulus 1 [89, pp. 278-279]. For such matrices S there exists a permutation matrix P so that

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

where the $k \times k$ matrix A represents the recurrent states, the (possibly empty) square matrix B represents the transient states, and the stationary distribution π satisfies $P\pi = \begin{pmatrix} \pi_{1:k}^T & 0 \end{pmatrix}^T$ with $\pi_{1:k} > 0$. The ergodicity coefficients of such matrices S can be expressed in terms of the ergodicity coefficient of the leading principal submatrix A.

COROLLARY 6.11 (Theorem 3 in [89]). *If S is a stochastic matrix where 1 is the only eigenvalue of modulus 1, and if S is permuted into the form* (6.1)*, then*

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

6.3. Explicit expressions for infinity-norm coefficients. We extend Theorem 4.2 from stochastic matrices to real matrices.

THEOREM 6.12 (section 3 in [78]). *If $A \in \mathbb{R}^{m\times n}$ and $w \in \mathbb{R}^m$, and P is a permutation matrix so that $Pw = \begin{pmatrix} w_{1:k}^T & 0 \end{pmatrix}^T$ with $|w_{1:k}| > 0$, then*

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

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

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

For any vector x with $x^T w = 0$, permutation matrix P, and scalar θ, one shows as in Lemma 3.2 that

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

Choose the 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 $a_1/w_1 \ge \cdots \ge a_k/w_k$. 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 $\tau_\infty(w, A) = |x^T a| \le \phi(a) \le \max_{1\le i\le n} \phi(Ae_i)$.

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

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

Then $\phi(s) = |y^T s|$ 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 $\max_{1\le i\le n} \phi(Ae_i) = \phi(a) = |y^T s| \le \tau_\infty(w, A)$. ∎

In the special case when S is a stochastic matrix and $w = \mathbb{1}$, the expression for $\tau_\infty(\mathbb{1}, S)$ from Theorem 6.12 reduces to that for $\tau_\infty(S)$ from Theorem 4.2. Below is the expression when $w = \pi$ is the stationary distribution of S.

Corollary 6.13 (Theorem 1 in [89]). *Let S be a stochastic matrix where 1 is the only eigenvalue of modulus 1, and let S be permuted into the form* (6.1). *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 $s_1/\pi_1 \ge \cdots \ge s_k/\pi_k$, l is the smallest integer such that $\sum_{j=1}^{l} \pi_j \ge \frac{1}{2}$, and

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

We can view $\tau_\infty(w, A)$ as the norm of an (oblique) projection of A, with the projection being onto $\mathrm{range}(w)^\perp$. This is an extension of Corollary 4.4 from stochastic matrices to real matrices.

Corollary 6.14. *Let the assumptions of Theorem* 6.12 *hold. Then for some* $1 \le k \le n$

$$\tau_\infty(w, A) = \left\|A^T\left(I - D^{-1} e_k w^T\right)\right\|_\infty, \qquad \textit{where} \quad D \equiv P^T \begin{pmatrix} \mathrm{diag}(w_{1:k}) & 0 \\ 0 & I_{n-k}\end{pmatrix} P.$$

Proof. The proof is analogous to that of Corollary 4.4. ∎

6.4. Explicit expressions for two-norm coefficients. We derive four different expressions for $\tau_2(w, A)$ and extend results in [76, 86, 89] for stochastic matrices to real rectangular matrices. We start by representing $\tau_2(w, A)$ as the norm of a matrix with one row less.

Theorem 6.15 (first expression). *Let $A \in \mathbb{R}^{m\times n}$, and $w \in \mathbb{R}^m$ with $w \ne 0$. 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,

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

With the help of Theorem 6.15 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 6.16 (second expression). *In addition to the conditions of Theorem 6.15, let $\tau \equiv \tau_2(w, A) = \|A^T y\|_2$, where $\|y\|_2 = 1$ and $y^T w = 0$. Then*

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

Proof. We represent the singular value problem $\|A_{m-1}\|_2$ from the proof of Theorem 6.15 as an eigenvalue problem,

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

Since $Q^T A A^T Q$ is real symmetric positive semidefinite, 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 $A^T Q = \begin{pmatrix} a & A_{m-1}^T \end{pmatrix}$, $Q^T y = \begin{pmatrix} 0 \\ \hat{y} \end{pmatrix}$, and $Q e_1 = w / \|w\|_2$ follows

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

Hence

$$a^T A_{m-1}^T \hat{y}\, Q e_1 = \frac{w^T A A^T y}{\|w\|_2^2} w. \qquad \square$$

Theorem 6.16 in turn leads to a third characterization which was introduced by Seneta and Tan [86, Proposition 1]. We replace the Lagrange multiplier based proof by one that is matrix based and involves the adjugate.

DEFINITION 6.17 (section 0.8.2 in [37]). *The* adjugate $\mathrm{adj}[A]$ *of an* $n \times n$ *matrix* A *is the* $n \times n$ *transposed matrix of cofactors. Its elements are* $(\mathrm{adj}[A])_{ij} = (-1)^{i+j}\det(A^{(ji)})$, *where* $A^{(ji)}$ *is obtained from* A *by deleting row* j *and column* i, $1 \le i, j \le n$. *For nonsingular matrices* A *this implies* $A\,\mathrm{adj}[A] = \det(A)\,I$.

THEOREM 6.18 (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 6.16, 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 6.16, 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

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

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

Since each product $(\omega_1 - \lambda)\prod_{j=2,j\neq i}^{m}(\omega_j - \lambda)$ in $w^T\mathrm{adj}[AA^T - \lambda I]w$ consists of $m-1$ factors, the quantity $w^T\mathrm{adj}[AA^T - \lambda I]w$ 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 6.15 to the adjugate gives

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

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

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

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

The fourth and last expression below is an extension to real matrices of the expression [89, Theorem 2] where S is a stochastic matrix and w a stationary distribution of S. The expression below suggests that $\tau_2(w, A)$ is equal to the norm of A when orthogonally projected onto the subspace $\mathrm{range}(w)^\perp$, and in this sense, it resembles Theorem 6.15.

THEOREM 6.19 (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. Because the two-norm corresponds to a quadratic form, we can use the constraint $z^T w = 0$ to eliminate an element from z and reduce the dimension of the maximization problem $\tau_2(w, A) = \max_{z^T w = 0, \|z\|_2 = 1} \|A^T z\|_2$.

Let P be a permutation that moves all nonzero elements in w to the top,

$$Pw = \begin{pmatrix} w_{1:m} \\ 0 \end{pmatrix}, \qquad \text{where} \quad |w_{1:m}| > 0.$$

Let $z \in \mathbb{R}^n$ be a vector with $z^T w = 0$ and $\|z\|_2 = 1$. Applying the permutation to z and A, and distinguishing the leading row gives

$$Pz = \begin{pmatrix} z_1 \\ z_{2:n} \end{pmatrix}, \qquad PA = \begin{pmatrix} a_1^T \\ A_{2:n}^T \end{pmatrix}.$$

From $0 = w^T z = (Pw)^T (Pz) = w_1 z_1 + w_{2:n}^T z_{2:n}$ and $w_1 \neq 0$ follows $z_1 = -w_{2:n}^T z_{2:n}/w_1$. Substituting this expression into $A^T z$ gives

$$A^T z = (PA)^T\,(Pz) = a_1 z_1 + A_{2:n} z_{2:n} = R z_{2:n}, \qquad \text{where} \quad R \equiv A_{2:n} - a_1 w_{2:n}^T / w_1.$$

Hence $\|A^T z\|_2 = \|R z_{2:n}\|_2$. Furthermore, substituting the expression for z_1 into the other constraint $1 = \|z\|_2^2 = (Pz)^T (Pz)$ gives

$$z_{2:n}^T Q z_{2:n} = 1, \qquad \text{where} \quad Q \equiv I_{n-1} + w_{2:n} w_{2:n}^T / w_1^2.$$

The matrix Q is real symmetric positive definite, and has a Cholesky factorization $Q = L^T L$, so that $\|L z_{2:n}\|_2 = 1$.

Thus the problem of maximizing $\|A^T z\|_2$ for $z \in \mathbb{R}^n$ subject to $z^T w = 0$ 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

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

Now we have reduced a constrained maximization problem of order n to an unconstrained maximization problem of order $n-1$. Looking inside the expression on the right gives

$$\|RL^{-1}\|_2^2 = \|RQ^{-1}R^T\|_2 = \left\|R\left(I + w_{2:n}w_{2:n}^T/w_1^2\right)^{-1} R^T\right\|_2.$$

From $Q^{-1} = I - w_{2:n}w_{2:n}^T/\|w\|_2^2$ and $RQ^{-1} = A_{2:n} - A^T ww_{2:n}^T/\|w\|_2^2$ follows

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

where the last equality is due to the fact that $I - ww^T/\|w\|_2^2$ is Hermitian and idempotent. ∎

Note that the two-norm expression for $\tau_2(w, A)$ in Theorem 6.19 represents an equality, while the Frobenius norm expression in Theorem 6.3 is only a bound.

In the special case when w is a dominant singular vector of A the expression in Theorem 6.19 reduces to the second largest singular value of A.

COROLLARY 6.20. *Let $A \in \mathbb{R}^{m\times n}$ have singular values $\sigma_1(A) \geq \sigma_2(A) \geq \cdots$ 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 6.19 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 the leading diagonal element of Σ is $\Sigma_{11} = \sigma_1(A)$, and the matrices 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 6.20 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 [86, p. 3].

6.5. 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

$$Au = \lambda_1 u, \qquad \text{where} \quad \lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_n|. \tag{6.2}$$

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 p-norm bound for nonnegative matrices below is an extension of the bounds for stochastic matrices in the one-norm in Theorem 3.6, and in the infinity-norm in Theorem 4.3.

THEOREM 6.21 (Theorem 3.1 in [71]). *If* $A \in \mathbb{R}^{n\times n}$ *is a nonnegative irreducible matrix as in* (6.2)*, 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 $\tau_p(u,A) = \max_{\substack{\|z\|_c=1 \\ z^Tu=0,\, z\in\mathbb{C}^n}} \left\|A^T z\right\|_c$. ☐

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

LEMMA 6.22 (page 293 in [6]). *If* $A \in \mathbb{R}^{n\times n}$ *is a nonnegative irreducible matrix as in* (6.2) *and* $D_u = \mathrm{diag}(u)$*, then* $(D_u^{-1} A D_u)\, \mathbb{1} = \lambda_1 \mathbb{1}$.

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 6.23 (page 63 in [71]). *If* $A \in \mathbb{R}^{n\times n}$ *is a nonnegative irreducible matrix as in* (6.2) *and* $D_u = \mathrm{diag}(u)$*, then*

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

Remark 6.24 (page 346 in [77]). The bound in Theorem 6.23, when applied to stochastic matrices, can be tighter than the bound in Theorem 6.21.

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 $\lambda_1 = 1$, $\lambda_2 = 3/16$, $\lambda_3 = \lambda_4 = 0$, so that the modulus of the largest subdominant eigenvalue is $\max_{i\geq 2} |\lambda_i| = 3/16$.

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 $\tau_1\left(D_u^{-1}SD_u\right)$ and $\tau_\infty\left(D_u^{-1}SD_u\right)$ represent tighter bounds for $\max_{i\geq 2}|\lambda_i| = 3/16$ than the other four coefficients.

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

COROLLARY 6.25. *If $A \in \mathbb{R}^{n\times n}$ is a nonnegative irreducible matrix as in* (6.2) *and $v^T A = \lambda_1 v^T$, then*

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

Proof. This follows from Theorem 6.21, and from Theorem 6.3 with $x = \lambda_1 v$. ☐

In particular, Corollary 6.25 implies an inclusion interval for $\tau_p(u, A)$ in terms of A deflated by its dominant spectral projector [78, (2)]

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

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

7. Complex matrices and general subspaces. The most general form of ergodicity coefficient [71, section 7], [32, section 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 general p-norm coefficients in section 7.1, and consider the special case when W is an invariant subspace of A in section 7.2. Then we focus on the two-norm coefficient, for we which derive explicit expressions in section 7.3, and establish its relation to singular values in section 7.4. In section 7.5 we derive inclusion regions for eigenvalues, and illustrate that for normal matrices, the two-norm coefficient is a Lehmann bound.

7.1. Properties common to all *p*-norm coefficients. A 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 7.1. *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. $\tau_p(W, BA) \le \|A^*\|_p\, \tau_p(W, B)$.

Proof. The proof is analogous to that for real matrices in Theorem 6.1. ☐

As in the case of real matrices in Theorem 6.3, we can represent $\tau_p(W, A)$ as the norm of a downdated matrix.

THEOREM 7.2. *If $A \in \mathbb{C}^{m\times n}$ and $W \in \mathbb{C}^{m\times k}$, then for all $X \in \mathbb{C}^{k\times n}$*

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

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

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

where the last inequality follows from Theorem 7.1. ☐

Theorem 7.2 represents an improvement over the bound in item 1 of Theorem 7.1 whenever $\|A^* - XW^*\|_p < \|A^*\|_p$.

7.2. 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 7.3. *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 6.2. ☐

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 3.6 and for irreducible nonnegative matrices in Theorem 6.21.

THEOREM 7.4. *Let $A \in \mathbb{C}^{n\times n}$ have eigenvalues λ_j, labeled $|\lambda_1| \geq \cdots \geq |\lambda_n|$. 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*

$$|\lambda_{k+1}| \leq \tau_p(W, A).$$

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 $\|v\|_p = 1$. It is clear that $v^*W = 0$ if $\lambda_k \neq \lambda_{k+1}$, or if $\lambda_k = \lambda_{k+1}$ and λ_k is not defective. To see that it also holds if λ_k is defective, simply consider the Jordan block

$$A = \begin{pmatrix} \lambda_1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_1 \end{pmatrix}$$

of order $k+1$, where $\lambda_1 = \cdots = \lambda_k = \lambda_{k+1}$. For W to be an invariant subspace, it must be spanned by $e_1, \ldots, e_l$ for some $1 \leq l \leq k$. The only left eigenvector is $v = e_{k+1}$, so that $v^*W = 0$. Therefore we can conclude

$$|\lambda_{k+1}| = \|A^*v\|_p \leq \max_{\substack{\|z\|_p=1 \\ z^*W=0}} \|A^*z\|_p = \tau_p(W, A). \qquad ☐$$

If the columns of W are actually vectors from a Jordan decomposition, then we can say something about the tightness of $\tau_2(W, A)$ as an eigenvalue bound. Denote a Jordan decomposition of A by $A = X\Lambda X^{-1}$, 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 7.5. *Let $A \in \mathbb{C}^{n\times n}$ have eigenvalues λ_j, labeled $|\lambda_1| \geq \cdots \geq |\lambda_n|$. 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 7.4.

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 7.5 suggests that the tightness of $\tau_2(W, A)$ as an eigenvalue bound for the subdominant eigenvalues depends on the vectors in their Jordan decomposition. An analogous bound based on the Schur decomposition will be presented in Theorem 7.11.

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

7.3. Explicit expressions for two-norm coefficients. The four explicit expressions for two-norm coefficients of real matrices in section 6.4 also hold for complex matrices. In particular, we extend Theorems 6.15 and 6.19 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 $\text{range}(W)^\perp$.

THEOREM 7.6. *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 6.15 for real matrices. ∎

In contrast to the matrix A_{m-k} in Theorem 7.6, which has fewer rows than A, the projected matrix below has the same dimension as A.

THEOREM 7.7. *Let $A \in \mathbb{C}^{m \times n}$ and $W \in \mathbb{C}^{m \times k}$ with* $\text{rank}(W) = k$. *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 a 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, and 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

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

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 $BB^* = I - W(W^*W)^{-1}W^*$. ◻

7.4. Two-norm coefficients and singular values. We show that two-norm ergodicity coefficients are closely related to singular values. In particular, 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 7.5. The result below extends Corollary 6.20 from real matrices to complex matrices and arbitrary subspaces.

Corollary 7.8. *Let $A \in \mathbb{C}^{m\times n}$ have singular values $\sigma_1 \geq \sigma_2 \geq \cdots$, and let the columns of U_j and V_j consist of the respective left and right singular vectors associated with the j largest singular values $\sigma_1, \ldots, \sigma_j$. 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 7.7, and the proof is analogous to that of Corollary 6.20. ◻

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 7.9. *Let $A \in \mathbb{C}^{m\times n}$ have singular values $\sigma_1 \geq \sigma_2 \geq \cdots$. 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 [37, 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 7.8 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 maximum is attained if the columns of W are k left singular vectors associated with singular values σ_j for $j > 1$. ◻

7.5. Two-norm coefficients and eigenvalues. For a normal matrix, that is, a matrix A with $A^*A = AA^*$, Theorems 7.5 and 7.9 readily imply that the two-norm ergodicity coefficients based on dominant eigenvectors can reproduce the magnitude of any eigenvalue.

THEOREM 7.10. *Let $A \in \mathbb{C}^{n\times n}$ be a normal matrix with eigenvalues λ_j, labeled so that $|\lambda_1| \geq \cdots \geq |\lambda_n|$, 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 7.9 implies $|\lambda_{k+1}| \leq \tau_2(W, A)$. The matrices X in the Jordan decomposition $A = X\Lambda X^{-1}$ from section 7.2 are unitary, so that the submatrices X_2 and Y_2 in Theorem 7.5 have orthonormal columns and $\|X_2\|_2 = \|Y_2\|_2 = 1$. From Theorem 7.5 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 7.5 this was done by choosing for W those vectors from a Jordan decomposition that are associated with dominant eigenvalues. Below we derive an analogous bound when W consists of vectors from a Schur decomposition associated with dominant eigenvalues.

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

THEOREM 7.11. *Let $A \in \mathbb{C}^{n\times n}$ have eigenvalues λ_j, ordered so that $|\lambda_1| \geq \cdots \geq |\lambda_k|$ 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 [26, Lemma 7.1.2]. The remaining proof is similar to the proofs of Theorems 7.5 and 7.10. ∎

Theorem 7.11 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 region. They are expressed in terms of singular values of the matrix restricted to a subspace; see [60, section 10.5] for Hermitian matrices and [7] for general matrices. Theorem 7.12 below presents Lehmann bounds for normal matrices. We use $\sigma_i(B)$ to denote the ith largest singular value of the matrix B.

THEOREM 7.12 (Corollary 2.3 in [7]). *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*

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

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 7.13. *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*

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

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 $\operatorname{range}(X) = \operatorname{range}(W)^{\perp}$. Then

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

Now apply Theorem 7.12 with $m = n-k$ and $i = 1$. ∎

8. Summary and future work. Ergodicity coefficients were designed to estimate the rate at which a product of stochastic matrices converges to a rank-one matrix. An ergodicity coefficient for a stochastic matrix S can be defined as

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

where the maximum ranges over real vectors z. One can interpret this coefficient as the p-norm of the matrix S restricted to the subspace orthogonal to its dominant right eigenvector $\mathbb{1}$. Our goal was to introduce ergodicity coefficients to the numerical linear algebra community, and to present a coherent discussion of existing work.

Rhodius, Seneta, and Tan thoroughly analyzed ergodicity coefficients for real, and in particular stochastic, matrices, mostly for $p = 1$ and $p = \infty$. They proved properties like continuity and submultiplicativity, derived explicit expressions, and showed that $\tau_p(S)$ defines an inclusion region for the eigenvalues $\lambda \neq 1$ of S. Further work is necessary, though, to determine how tight of a bound $\tau_p(S)$ is for the subdominant eigenvalues closest to $\lambda = 1$.

In terms of applications, Kirkland, Meyer, and Seneta have shown how to use $\tau_1(S)$ as a condition number for the stationary distribution of S. In this context, it would be useful to develop fast estimators for $\tau_1(S)$.

Following Hartfiel, Rothblum, Seneta, and Tan, we extended ergodicity coefficients to complex rectangular matrices $A \in \mathbb{C}^{m\times n}$ and subspaces $W \in \mathbb{C}^{m\times k}$,

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

where the maximum ranges over complex vectors z. Our discussion focused on $p = 2$. In particular, we presented several explicit forms for $\tau_2(W, A)$, illustrated relations to Jordan and Schur forms, and demonstrated how to define inclusion regions for singular values and eigenvalues. We showed that $\tau_2(W, A)$ is bounded by singular values of A. For normal matrices A we pointed out that the inclusion regions defined by $\tau_2(W, A-\gamma I)$ are special cases of Lehmann bounds, and that for nonnormal matrices A the tightness of the inclusion regions depends on the departure of A from normality. Further work is necessary on explicit forms for $\tau_p(W, A)$ for general p and the tightness of the inclusion regions defined by these coefficients.

Acknowledgments. We thank Carl Meyer, Steve Campbell, and Steve Kirkland for their encouragement and many helpful discussions. We are also grateful to the two anonymous reviewers for the care with which they read our paper, and for their insightful comments.

REFERENCES

[1] M. Akelbek and S. Kirkland, *Coefficients of ergodicity and the scrambling index*, Linear Algebra Appl., 430 (2009), pp. 1111–1130. pages
[2] 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. pages 181
[3] M. Artzrouni, *The local coefficient of ergodicity of a nonnegative matrix*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 507–516. pages 159
[4] M. Artzrouni and O. Gavart, *Nonlinear matrix iterative processes and generalized coefficients of ergodicity*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1343–1353. pages 159
[5] 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. pages 159
[6] F. L. Bauer, E. Deutsch, and J. Stoer, *Abschätzungen für Eigenwerte positiver linearer Operatoren*, Linear Algebra Appl., 2 (1969), pp. 275–301. pages 163, 164, 190
[7] C. Beattie and I. C. Ipsen, *Inclusion regions for matrix eigenvalues*, Linear Algebra Appl., 358 (2003), pp. 281–291. pages 195
[8] S. N. Bernštein, *Sur le théorème limite du calcul des probabilités*, Math. Ann., 85 (1922), pp. 237–241. pages 156
[9] S. N. Bernštein, *Sur les sommes de quantités dépendantes*, Izv. Akad. Nauk SSSR, 15-17 (1926), pp. 1459–1478. pages 156
[10] S. N. Bernštein, *Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes*, Math. Ann., 97 (1926), pp. 1–59; Uspekhi Mat. Nauk, 10 (1944), pp. 65–144 (in Russian). pages 156
[11] S. N. Bernštein, *Sur les sommes de quantités dépendantes*, Dokl. Akad. Nauk SSSR., A, 4 (1928), pp. 55–60. pages 156
[12] S. N. Bernštein, *Teoriya Veroiatnostei*, 4th ed., Gostehizdat, Moscow, Leningrad, 1946. Relevant portions are republished in [13]. pages 156
[13] S. N. Bernštein, *Classification of Markov chains and their matrices*, in Sobranie Sochineniy: Tom IV, Teoriya Veroiatnostei i Matematicheskaia Statistika [1911-1946], Nauka, Moscow, 1964, pp. 455–483. pages 156, 197
[14] A. Brauer, *Limits for the characteristic roots of a matrix* IV*: Applications to stochastic matrices*, Duke Math. J., 19 (1952), pp. 75–91. pages 155
[15] S. Chatterjee and E. Seneta, *Towards consensus: Some convergence theorems on repeated averaging*, J. Appl. Probab., 14 (1977), pp. 89–97. pages 157
[16] G. E. Cho and C. D. Meyer, *Markov chain sensitivity measured by mean first passage time*, Linear Algebra Appl., 316 (2000), pp. 21–28. pages 168
[17] 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. pages 168
[18] H. Cohn, *Finite non-homogeneous Markov chains: Asymptotic behaviour*, Adv. in Appl. Probab., 8 (1976), pp. 502–516. pages 156
[19] H. Cohn, *On a paper by Doeblin on non-homogeneous Markov chains*, Adv. in Appl. Probab., 13 (1981), pp. 388–401. pages 156
[20] E. Deutsch and C. Zenger, *Inclusion domains for the eigenvalues of stochastic matrices*, Numer. Math., 18 (1971), pp. 182–192. pages 165
[21] R. L. Dobrushin, *Central limit theorem for nonstationary Markov chains*, Dokl. Akad. Nauk SSSR. (N.S.), 102 (1955), pp. 5–8. pages 156
[22] R. L. Dobrushin, *Central limit theorem for nonstationary Markov chains.* I, II, Theory Probab. Appl., 1 (1956), pp. 65–80. pages 156, 159, 164, 165, 166
[23] W. Doeblin, *Le cas discontinu des probabilités en chaîne*, Publ. Fac. Sci. Univ. Masaryk (Brno), 236 (1937), pp. 3–13. pages 156
[24] E. B. Dynkin, *On certain limit theorems for Markov chains*, Ukrain. Mat. Zh., 6 (1954), pp. 21–27. pages 156
[25] L. Elden, *The eigenvalues of the Google matrix*, Technical report LiTH-MAT-R-04-01, Department of Mathematics, Linköping University, Linköping, Sweden, 2003. pages 154
[26] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, 1996. pages 195
[27] 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. pages 153, 160, 179, 180
[28] J. Hajnal, *The ergodic properties of non-homogeneous finite Markov chains*, Proc. Cambridge Philos. Soc., 52 (1956), pp. 67–77. pages 156

[29] J. Hajnal, *Weak ergodicity in non-homogeneous Markov chains*, Proc. Cambridge Philos. Soc., 54 (1958), pp. 233–246. pages 156, 157, 166, 167
[30] J. Hajnal, *On products of non-negative matrices*, Math. Proc. Cambridge Philos. Soc., 79 (1976), pp. 521–530. pages 159
[31] D. Hartfiel, *Coefficients of ergodicity for imprimitive matrices*, Comm. Statist. Stochastic Models, 15 (1999), pp. 81–88. pages 160
[32] D. Hartfiel and U. G. Rothblum, *Convergence of inhomogeneous products of matrices and coefficients of ergodicity*, Linear Algebra Appl., 277 (1998), pp. 1–9. pages 160, 191
[33] D. J. Hartfiel, *Nonhomogeneous Matrix Products*, World Scientific, River Edge, NJ, 2002. pages 157, 159, 167
[34] T. H. Haveliwala and S. D. Kamvar, *The Second Eigenvalue of the Google Matrix*, Technical report 2003-20, Computer Science Department, Stanford University, Palo Alto, CA, 2003. pages 154
[35] M. Haviv and L. Van DER Heyden, *Perturbation bounds for the stationary probabilities of a finite Markov chain*, Adv. in Appl. Probab., 16 (1984), pp. 804–818. pages 161, 168
[36] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002. pages 174
[37] R. A. Horn and C. A. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985. pages 175, 187, 194
[38] R. A. Horn and S. Serra-Capizzano, *A general setting for the parametric Google matrix*, Internet Math., 3 (2006), pp. 385–411. pages 155
[39] J. J. Hunter, *Stationary distributions and mean first passage times of perturbed Markov chains*, Linear Algebra Appl., 410 (2005), pp. 217–243. pages 168
[40] J. J. Hunter, *Mixing times with applications to perturbed Markov chains*, Linear Algebra Appl., 417 (2006), pp. 108–123. pages 168
[41] M. Iosifescu, *On two recent papers on ergodicity in nonhomogeneous Markov chains*, Ann. Math. Statist., 43 (1972), pp. 1732–1736. pages 166
[42] M. Iosifescu, *Finite Markov Processes and Their Applications*, John Wiley & Sons, Chichester, UK, Editura Tehnică, Bucharest, 1980. pages 157, 158
[43] 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. pages 154
[44] 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. pages 154
[45] I. C. F. Ipsen and R. S. Wills, *Mathematical properties and analysis of Google's PageRank*, Bol. Soc. Esp. Mat. Apl. S$\vec{e}$MA, No. 34 (2006), pp. 191–196. pages 154
[46] D. Isaacson and R. Madsen, *Positive columns for stochastic matrices*, J. Appl. Probab., 11 (1974), pp. 829–835. pages 155, 166
[47] D. L. Isaacson and R. Madsen, *Markov Chains. Theory and Applications*, John Wiley & Sons, New York, London, Sydney, 1976. pages 160
[48] J. F. C. Kingman, *Geometrical aspects of the theory of non-homogeneous Markov chains*, Math. Proc. Cambridge Philos. Soc., 77 (1975), pp. 171–183. pages 156, 157, 164
[49] S. Kirkland, *On a question concerning condition numbers for Markov chains*, SIAM J. Matrix Anal. Appl., 23 (2002), pp. 1109–1119. pages 168
[50] S. J. Kirkland and M. Neumann, *The case of equality in the Dobrushin-Deutsch-Zenger bound*, Linear Algebra Appl., 431 (2009), pp. 2372–2394. pages 164, 165
[51] S. J. Kirkland, M. Neumann, and B. L. Shader, *Applications of Paz's inequality to perturbation bounds for Markov chains*, Linear Algebra Appl., 268 (1998), pp. 183–196. pages 168
[52] S. J. Kirkland, M. Neumann, and N.-S. Sze, *On optimal condition numbers for Markov chains*, Numer. Math., 110 (2008), pp. 521–537. pages 168
[53] A. N. Kolmogorov, *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung*, Math. Ann., 104 (1931), pp. 415–458. pages 156
[54] 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. pages 160
[55] A. Lešanovský, *Coefficient of ergodicity generated by non-symmetrical vector norms*, Czechoslovak Math. J., 40 (1990), pp. 284–294. pages 159, 160, 170
[56] A. A. Markov, *Extension of the law of large numbers to dependent quantities*, Izvestiia Fiz.-Matem. Obsch. Kazan Univ., 2 (1906), pp. 135–156 (in Russian). English translation [58]. pages 165, 199
[57] A. A. Markov, *Izbrannie Trudy (Selected Works)*, A.N.S.S.S.R., Leningrad, 1951. pages 165
[58] A. A. Markov, *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 [56] into English. pages 165, 198

[59] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, 2000. pages 167

[60] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice–Hall, Englewood Cliffs, NJ, 1980. pages 195

[61] A. Paz, *Ergodic theorems for infinite probabilistic tables*, Ann. Math. Statist., 41 (1970), pp. 539–550. pages 155, 157, 158, 163

[62] A. Paz, *Introduction to Probabilistic Automata*, Academic Press, New York, London, 1971. pages 160, 165

[63] A. Paz and M. Reichaw, *Ergodic theorems for sequences of infinite stochastic matrices*, Proc. Cambridge Philos. Soc., 63 (1967), pp. 777–784. pages 155, 156, 166

[64] H. Perfect and L. Mirsky, *Spectral properties of doubly-stochastic matrices*, Monatsh. Math., 69 (1965), pp. 35–57. pages 172

[65] 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. pages 166, 181

[66] A. Rhodius, *The maximal value for coefficients of ergodicity*, Stochastic Process. Appl., 29 (1988), pp. 141–145. pages 160, 165, 170, 173, 174, 175, 176, 177

[67] A. Rhodius, *Über explizite Formeln für Ergodizitätskoeffizienten*, Wiss. Z. Pädagog. Hochsch. "Karl Friedrich Wilhelm Wander" Dresden, 23 (1989), pp. 69–74. pages 159

[68] A. Rhodius, *On explicit forms for ergodicity coefficients*, Linear Algebra Appl., 194 (1993), pp. 71–83. pages 159, 160

[69] A. Rhodius, *On the maximum of ergodicity coefficients, the Dobrushin ergodicity coefficient, and products of stochastic matrices*, Linear Algebra Appl., 253 (1997), pp. 141–154. pages 160, 165

[70] A. Rhodius, *On ergodicity coefficients of infinite stochastic matrices*, Z. Anal. Anwendungen, 19 (2000), pp. 873–887. pages 155, 160, 165

[71] 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. pages 153, 159, 160, 179, 190, 191

[72] N. A. Sapogov, *The Laplace-Lyapunov limit theorem for singular Markov chains*, Dokl. Akad. Nauk SSSR., 58 (1947), pp. 1905–1908. pages 156

[73] N. A. Sapogov, *On singular Markov chains*, Dokl. Akad. Nauk SSSR., 58 (1947), pp. 193–196. pages 156

[74] T. A. Sarymsakov, *On inhomogeneous Markov chains*, Dokl. Akad. Nauk S.S.S.R., 120 (1958), pp. 465–467. pages 156

[75] E. Seneta, *On the historical development of the theory of finite inhomogeneous Markov chains*, Proc. Cambridge Philos. Soc., 74 (1973), pp. 507–513. pages 156, 157, 158, 166, 167

[76] E. Seneta, *Coefficients of ergodicity: Structure and applications*, Adv. in Appl. Probab., 11 (1979), pp. 576–590. pages 157, 159, 160, 163, 164, 167, 168, 171, 173, 181, 185

[77] E. Seneta, *Spectrum localization by ergodicity coefficients for stochastic matrices*, Linear Multilinear Algebra, 14 (1983), pp. 343–347. pages 160, 172, 173, 190

[78] E. Seneta, *Explicit forms for ergodicity coefficients and spectrum localization*, Linear Algebra Appl., 60 (1984), pp. 187–197. pages 154, 163, 164, 169, 177, 181, 182, 183, 184, 191

[79] E. Seneta, *Perturbations of the stationary distribution measured by ergodicity coefficients*, Adv. in Appl. Probab., 20 (1988), pp. 228–230. pages 160, 167, 168

[80] E. Seneta, *Sensitivity analysis, ergodicity coefficients, and rank-one updates for finite Markov chains*, in Numerical Solutions of Markov Chains, W. J. Stewart, ed., Marcel Dekker, New York, 1991, pp. 121–129. pages 168

[81] E. Seneta, *Applications of ergodicity coefficients to homogeneous Markov chains*, in Doeblin and Modern Probability, Contemp. Math. 149, H. Cohn, ed., AMS, Providence, RI, 1993, pp. 189–199. pages 164, 181

[82] E. Seneta, *Sensitivity of finite Markov chains under perturbation*, Statist. Probab. Lett., 17 (1993), pp. 163–168. pages 168

[83] E. Seneta, *Markov and the birth of chain dependence theory*, Int. Stat. Rev., 64 (1996), pp. 255–263. pages 165

[84] E. Seneta, *Markov and the creation of Markov chains*, in MAM 2006: Markov Anniversary Meeting, A. N. Langville and W. J. Stewart, eds., Boson Books, Raleigh, NC, 2006, pp. 1–20. pages 154, 157, 160, 161, 165, 168

[85] E. Seneta, *Non-Negative Matrices and Markov Chains*, Springer-Verlag, New York, 2006 (revised reprint of 1981 edition). pages 157, 158, 159, 160, 162, 165, 166

[86] E. Seneta and C. P. Tan, *The Euclidean and Frobenius ergodicity coefficients and spectrum localization*, Bull. Malaysian Math. Soc. (2), 7 (1984), pp. 1–7. pages 185, 187, 189
[87] 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. pages 154
[88] C. P. Tan, *A functional form for a particular coefficient of ergodicity*, J. Appl. Prob., 19 (1982), pp. 858–863. pages 160, 168, 169, 170
[89] C. P. Tan, *Coefficients of ergodicity with respect to vector norms*, J. Appl. Probab., 20 (1983), pp. 277–287. pages 160, 165, 173, 178, 183, 184, 185, 188
[90] R. S. Wills and I. C. F. Ipsen, *Ordinal ranking for Google's PageRank*, SIAM J. Matrix Anal. Appl., 30 (2009), pp. 1677–1696. pages 154