# Abstract

SELEE, TERESA MARGARET. Stochastic Matrices: Ergodicity Coefficients, and Applications to Ranking. (Under the direction of Ilse C. F. Ipsen.)

We present two different views of (row) stochastic matrices, which are nonnegative matrices with row sums equal to one. For applications to ranking, we examine the computation of a dominant left eigenvector of a stochastic matrix. The stochastic matrix of interest is called the Google matrix and contains information about how pages of the Internet are linked to one another. The dominant left eigenvector of the Google matrix yields a ranking for each Web page, which helps to determine the order in which search results are returned. These results are presented in Chapter 1.

Chapter 2 presents results for coefficients of ergodicity, which measure the rate at which products of stochastic matrices, especially products whose number of factors is unbounded, converge to a matrix of rank one. Ergodicity arises in the context of Markov chains and signals the tendency of the rows of such products to equalize. We present unified notation and definitions for coefficients of ergodicity applied to stochastic matrices, extend the definitions to general complex matrices, and illustrate a connection between ergodicity coefficients and inclusion regions for eigenvalues.

# Stochastic Matrices: Ergodicity Coefficients, and Applications to Ranking

by
Teresa Margaret Selee

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Applied Mathematics

Raleigh, North Carolina

2009

APPROVED BY:

| | |
|---|---|
| I. C. F. Ipsen<br>Chair of Advisory Committee | S. L. Campbell |
| C. T. Kelley | C. D. Meyer |

# Dedication

This work is dedicated to my family and friends. Without your love and support, this would not have been possible.

# Biography

Teresa Selee was born with bright eyes and an adventurous spirit in the Cleveland suburb of Elyria, Ohio. She spent the first eighteen years of her life growing up in North Ridgeville, Ohio. She was always eager to learn, looked forward to school, and did well in most subjects. It wasn't until her senior year of high school that she realized her true interest in math. Upon graduation, she attended Youngstown State University on a full academic scholarship. After a year of studying math, she added a second major in economics. Academically she was growing, but she yearned for more. During her third year of college she traveled across the pond to spend a semester at Oxford Brookes University, in Oxford, England. Before returning to the U.S., she spent six weeks traveling in Europe, visiting more than ten countries. During her travels, she realized the world was a classroom, as she learned about art, history, culture, food, and so much more.

Reluctantly, Teresa left Europe and returned to YSU for her senior year, earning high honors for her Bachelor's degrees in Mathematics and Economics. Unsure what to do with her degrees, Teresa worked for a few months before realizing how much she missed the challenging environment of the classroom. After being accepted to a PhD

program in mathematics at North Carolina State University, Teresa took off traveling again. She spent two months in Australia, driving around the country and visiting friends. She then spent another two months in Northern/Eastern Europe and Ireland, visiting more than ten countries.

Since she started graduate school six years ago, Teresa has had the opportunities to travel for work and pleasure to Paris, Düsseldorf, Boston, New Orleans, San Diego, San Francisco, Livermore and Chicago, as well as three cross-country drives. Now that she has completed her degree, Teresa is moving to Atlanta to work as a Research Scientist at Georgia Tech Research Institute. She is looking forward to the adventures awaiting her in a new city, with a new job, and especially with her soon-to-be husband Steve Stanislav.

# Acknowledgements

**Thanks to all those involved in my academic life.**

To my high school mathematics teacher, Mr. Adam Baillie, who helped me to recognize my potential and develop my math skills.

To Doug Faires, Nate Ritchey, and the rest of the Mathematics department at YSU who were an endless source or enrichment, encouragement and advice. You were supportive, and provided me so many opportunities that have helped me to develop into the person I am today.

To Teresa Riley and Rochelle Ruffer, two amazing professors, without whom I would never have earned my Economics degree.

To Tammy Kolda, Philip Kegelmeyer, and Josh Griffin whose time and energy helped me to have a great experience as a summer intern at Sandia National Laboratory in Livermore, CA. Thank you for growing my interest in tensors, clustering and data mining.

To the faculty and students, past and present, of the student Numerical Analysis seminar. Thank you for being a willing audience for practice talks, and for the useful feedback.

To the administrative staff in the Math department at NCSU. You are the backbone of this department and the key to keeping everything moving smoothly. Thank you for everything you do.

To my committee, Steve Campbell, Tim Kelley, and Carl Meyer for your support, advice, and assistance, not only during my thesis writing, but throughout my career at NCSU.

To my advisor, Ilse Ipsen, without whom this work would not be possible. Thank you for your time, ideas, feedback, and advice. Thank you also for the opportunities you afforded me to work, travel and expand my mathematical horizons.

**Thanks to my family, past, present, and future.**

To my parents who always supported me, and knew I could do this even when I had my doubts. Thank you for all the time and effort you've invested in me.

To my sister, Jacquelyn, who in the last 25 years has gone from an annoyance to one of my best friends. Thank you for always being there. I am so thankful to have you as a friend.

To my countless other family, grandparents, aunts, uncles, cousins, and more for always being the fun, crazy, loving people I call family.

To my future in-laws and new family. Thank you for your support and I am excited to become a part of your family.

**Thanks to my friends.**

To my officemates: Kelly Dickson, Laura Ellwein, Morgan Root, and Julie Beier. I'm so happy to have met you and spent so many days with you. Without your daily support and encouragement, I never would have survived this experience.

To my fellow students and friends who became my family in North Carolina, including Brian Adams, Steve and Lindsay May, Kelly Sweetingham, Ryan and Anna Hart Siskind, Iti and Tony Klein, Kristen Devault, Sarah Muccio, Laura Ellwein, Kelly Dickson, Julie Beier, Keri Kehoe, John Telliho, Corey Winton, Morgan and Lis Root, Ross and Anna Braymer, and many more. You have always been there for work and fun, school and play. Whether it be working through a tricky math problem, studying for exams, preparing for presentations, going out for food and drinks, watching and playing sports, or leaving town for a weekend of fun. I am so thankful to have all of you in my life.

To Amory Starkey for always believing in me. You are such a wonderful person and I wish you every happiness for your future.

To Mary Spann who has been through many ups and downs with me for almost twenty years.

Finally, thanks to my soon-to-be husband Steve Stanislav. You are everything in my world. Thank you for your love, support, friendship, advice, and the countless meals you've cooked for me. I cannot wait to spend the rest of my life with you.

# Table of Contents

# List of Figures

# Chapter 1

# Introduction

Stochastic matrices have been studied in many contexts, in part because of their relation to Markov chain theory and the study of products of stochastic matrices. Additionally, stochastic matrices have many useful properties, such as a dominant eigenvalue of 1 and a dominant right eigenvector as a vector of all 1's. In this work, we consider two different applications for stochastic matrices. Chapter 2 describes an application to ranking with respect to Google and the PageRank vector. The material in Chapter 2 was published in December 2007 in the SIAM Journal on Matrix Analysis and Applications (54). Chapter 3 contains results on the coefficient of ergodicity, which can be thought of as a bound on the subdominant eigenvalues of a matrix.

For applications to ranking, we present a simple algorithm for computing the PageRank (stationary distribution) of the stochastic Google matrix $G$. The algorithm lumps all dangling nodes into a single node. We express lumping as a similarity transformation of $G$ and show that the PageRank of the nondangling nodes can be computed

separately from that of the dangling nodes. The algorithm applies the power method only to the smaller lumped matrix, but the convergence rate is the same as that of the power method applied to the full matrix $G$. The efficiency of the algorithm increases as the number of dangling nodes increases. We also extend the expression for PageRank and the algorithm to more general Google matrices that have several different dangling node vectors, when it is required to distinguish among different classes of dangling nodes. We analyze the effect of the dangling node vector on the PageRank and show that the PageRank of the dangling nodes depends strongly on that of the nondangling nodes but not vice versa.

We also consider stochastic matrices and their relationship to ergodicity coefficients. These coefficients were originally introduced in the context of rates of convergence of finite, inhomogeneous Markov chains. In general, ergodicity deals with the long term behavior of dynamical systems. Thus for systems of finite, inhomogeneous Markov chains, ergodicity refers to the long-term behavior of products of stochastic matrices. In this paper we focus only on finite dimensional matrices. For some information on infinite dimensional stochastic matrices, see Isaacson and Madsen (57), Pax (80), Paz and Reichaw (82), and Rhodius (89).

An ergodicity coefficient $\tau_1(S)$ is defined for an $n \times n$ stochastic matrix $S$, as

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

where the maximum is taken over all $z \in \mathbb{R}^n$ and $e$ is a column vector of all 1's Notice that this form of an ergodicity coefficient is simply the norm of a matrix restricted to

a subspace. In addition to being continuous and bounded, $\tau_1(S)$ is submultiplicative, and bounds the magnitude of the second largest eigenvalue of $S$. Thus $\tau_1(S_1S_2) \leq \tau_1(S_1)\tau_1(S_2)$ and $|\lambda| \leq \tau_1(S)$ for all eigenvalues $\lambda \neq 1$ of $S$ for stochastic matrices $S_1$ and $S_2$. These two properties combine to yield an upper bound on the magnitude of subdominant eigenvalues of the product of stochastic matrices. The upper bound is computed as the product of the ergodicity coefficients of the individual matrices. That is, for any eigenvalue $\lambda \neq 1$ of the stochastic matrix $S_1S_2$, $|\lambda| \leq \tau_1(S_1S_2) \leq \tau_1(S_1)\tau_1(S_2)$. This inequality gives an eigenvalue bound for products of stochastic matrices, and is useful since, in general the eigenvalues of a product of matrices are not equal to the product of the eigenvalues of the matrices.

Chapter 3 discusses coefficient of ergodicity from a linear algebra point of view. We choose to express these coefficients as matrix norms restricted to a subspace, although we mention other expressions, as well. We present properties, and explicit forms for various ergodicity coefficients and simplify results and notation from the last 100 years, writing everything in language more familiar to the numerical linear algebra community. We conclude with new and existing applications, including condition numbers, eigenvalue and singular value bounds.

# Chapter 2

# PageRank Computation, with Special Attention to Dangling Nodes

## 2.1 Introduction

The order in which the search engine Google displays the Web pages is determined, to a large extent, by the *PageRank* vector (19; 77). The PageRank vector contains, for every Web page, a ranking that reflects the importance of the Web page. Mathematically, the PageRank vector $\pi$ is the stationary distribution of the so-called *Google matrix*, a sparse stochastic matrix whose dimension exceeds 11.5 billion (39). The Google matrix

$G$ is a convex combination of two stochastic matrices

$$G = \alpha S + (1 - \alpha)E, \qquad 0 \leq \alpha < 1,$$

where the matrix $S$ represents the link structure of the Web, and the primary purpose of the rank-one matrix $E$ is to force uniqueness for $\pi$. In particular, element $(i, j)$ of $S$ is nonzero if Web page $i$ contains a link pointing to Web page $j$.

However, not all Web pages contain links to other pages. Image files or pdf files, and uncrawled or protected pages have no links to other pages. These pages are called *dangling nodes*, and their number may exceed the number of nondangling pages (31, section 2). The rows in the matrix $S$ corresponding to dangling nodes would be zero if left untreated. Several ideas have been proposed to deal with the zero rows and force $S$ to be stochastic (31). The most popular approach adds artificial links to the dangling nodes, by replacing zero rows in the matrix with the same vector, $w$, so that the matrix $S$ is stochastic.

It is natural as well as efficient to exclude the dangling nodes with their artificial links from the PageRank computation. This can be done, for instance, by "lumping" all the dangling nodes into a single node (70). In section 2.3, we provide a rigorous justification for lumping the dangling nodes in the Google matrix $G$, by expressing lumping as a similarity transformation of $G$ (Theorem 2.3.1). We show that the PageRank of the nondangling nodes can be computed separately from that of the dangling nodes (Theorem 2.3.2), and we present an efficient algorithm for computing PageRank by applying the power method only to the much smaller, lumped matrix (section 2.3.3). Because the dangling nodes are excluded from most of the computations,

the operation count depends, to a large extent, on only the number of nondangling nodes, as opposed to the total number of Web pages. The algorithm has the same convergence rate as the power method applied to $G$, but is much faster because it operates on a much smaller matrix. The efficiency of the algorithm increases as the number of dangling nodes increases.

Many other algorithms have been proposed for computing PageRank, including classical iterative methods (2; 16; 68), Krylov subspace methods (34; 35), extrapolation methods (17; 18; 48; 61; 60), and aggregation/disaggregation methods (20; 55; 69); see also the survey papers (8; 66) and the book (67). Our algorithm is faster than the power method applied to the full Google matrix $G$, but retains all the advantages of the power method: It is simple to implement and requires minimal storage. Unlike Krylov subspace methods, our algorithm exhibits predictable convergence behavior and is insensitive to changes in the matrix (34). Moreover, our algorithm should become more competitive as the Web frontier expands and the number of dangling nodes increases. The algorithms in (68; 70) are special cases of our algorithm because our algorithm allows the dangling node and personalization vectors to be different, and thereby facilitates the implementation of TrustRank (41). TrustRank is designed to diminish the harm done by link spamming and was patented by Google in March 2005 (113). Moreover, our algorithm can be extended to a more general Google matrix that contains several different dangling node vectors (section 2.3.4).

In section 2.4 we examine how the PageRanks of the dangling and nondangling nodes influence each other, as well as the effect of the dangling node vector $w$ on the PageRanks of dangling and nondangling nodes. In particular we show (Theorem 2.4.1)

that the PageRanks of the dangling nodes depend strongly on the PageRanks of the nondangling nodes but not vice versa. Finally, in section 2.5, we consider a (theoretical) extreme case, where the Web consists solely of dangling nodes. We present a Jordan decomposition for general rank-one matrices (Theorems 2.5.1 and 2.5.2) and deduce from it a Jordan decomposition for a Google matrix of rank one (Corollary 2.5.3).

## 2.2 The ingredients

Let $n$ be the number of Web pages and $k$ the number of nondangling nodes among the Web pages, $1 \leq k < n$. We model the link structure of the Web by the $n \times n$ matrix

$$H \equiv \begin{bmatrix} H_{11} & H_{12} \\ 0 & 0 \end{bmatrix},$$

where the $k \times k$ matrix $H_{11}$ represents the links among the nondangling nodes, and $H_{12}$ represents the links from nondangling to dangling nodes; see Figure 2.1. The $n-k$ zero rows in $H$ are associated with the *dangling nodes*.

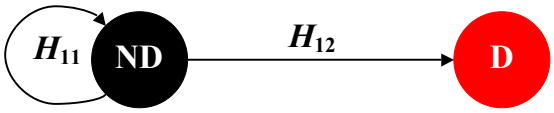


Figure 2.1: A simple model of the link structure of the Web. The sphere $ND$ represents the set of nondangling nodes, and $D$ represents the set of dangling nodes. The submatrix $H_{11}$ represents all the links from nondangling nodes to nondangling nodes, while the submatrix $H_{12}$ represents links from nondangling to dangling nodes.

The elements in the nonzero rows of $H$ are nonnegative and sum to one,

$$H_{11} \geq 0, \qquad H_{12} \geq 0, \qquad H_{11}e + H_{12}e = e, \qquad \text{where} \quad e \equiv \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},$$

and the inequalities are to be interpreted elementwise. To obtain a stochastic matrix, we add artificial links to the dangling nodes. That is, we replace each zero row in $H$ by the same *dangling node vector*

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \qquad w \geq 0, \qquad \|w\| \equiv w^T e = 1.$$

Here $w_1$ is $k \times 1$, $w_2$ is $(n-k) \times 1$, $\|\cdot\|$ denotes the one norm (maximal column sum), and the superscript $T$ denotes the transpose. The resulting matrix

$$S \equiv H + dw^T = \begin{bmatrix} H_{11} & H_{12} \\ ew_1^T & ew_2^T \end{bmatrix}, \qquad \text{where} \quad d \equiv \begin{bmatrix} 0 \\ e \end{bmatrix},$$

is stochastic, that is, $S \geq 0$ and $Se = e$.

Finally, so as to work with a stochastic matrix that has a unique stationary distribution, one selects a *personalization vector*

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \qquad v \geq 0, \qquad \|v\| = 1,$$

where $v_1$ is $k \times 1$ and $v_2$ is $(n-k) \times 1$, and defines the Google matrix as the convex combination

$$G \equiv \alpha S + (1-\alpha)ev^T, \qquad 0 \leq \alpha < 1.$$

Although the stochastic matrix $G$ may not be primitive or irreducible, its eigenvalue 1 is distinct and the magnitude of all other eigenvalues is bounded by $\alpha$ (32; 47; 60; 61; 107). Therefore $G$ has a unique stationary distribution,

$$\pi^T G = \pi^T, \qquad \pi \geq 0, \qquad \|\pi\| = 1.$$

The stationary distribution $\pi$ is called *PageRank*. Element $i$ of $\pi$ represents the PageRank for Web page $i$.

If we partition the PageRank conformally with $G$,

$$\pi = \begin{bmatrix} \pi_1 \\ \\ \pi_2 \end{bmatrix},$$

then $\pi_1$ represents the PageRank associated with the nondangling nodes and $\pi_2$ represents the PageRank of the dangling nodes.

The identity matrix of order $n$ will be denoted by $I_n \equiv [e_1 \cdots e_n]$, or simply by $I$.

## 2.3 Lumping

We show that lumping can be viewed as a similarity transformation of the Google matrix; we derive an expression for PageRank in terms of the stationary distribution

of the lumped matrix; we present an algorithm for computing PageRank that is based on lumping; and we extend everything to a Google matrix that has several different dangling node vectors, when it is required to distinguish among different classes of dangling nodes.

It was observed in (70) that the Google matrix represents a lumpable Markov chain. The concept of *lumping* was originally introduced for general Markov matrices, to speed up the computation of the stationary distribution or to obtain bounds (25; 40; 59; 62). Below we paraphrase lumpability (62, Theorem 6.3.2) in matrix terms: Let $P$ be a permutation matrix and

$$PMP^T = \begin{bmatrix} M_{11} & \cdots & M_{1,k+1} \\ \vdots & & \vdots \\ M_{k+1,1} & \cdots & M_{k+1,k+1} \end{bmatrix}$$

be a partition of a stochastic matrix $M$. Then $M$ is *lumpable* with respect to this partition if each vector $M_{ij}e$ is a multiple of the all-ones vector $e$, $i \neq j$, $1 \leq i, j \leq k+1$.

The Google matrix $G$ is lumpable if all dangling nodes are lumped into a single node (70, Proposition 1). We condense the notation in section 2.2 and write the Google matrix as

$$G = \begin{bmatrix} G_{11} & G_{12} \\ eu_1^T & eu_2^T \end{bmatrix}, \qquad \text{where} \quad u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \equiv \alpha w + (1-\alpha)v, \tag{2.1}$$

$G_{11}$ is $k \times k$, and $G_{12}$ is $(n-k) \times k$. Here element $(i,j)$ of $G_{11}$ corresponds to block $M_{ij}$, $1 \leq i, j \leq k$; row $i$ of $G_{12}$ corresponds to block $M_{i,k+1}$, $1 \leq i \leq k$; column $i$ of $eu_1^T$

corresponds to $M_{k+1,i}$, $1 \le i \le k$; and $eu_2^T$ corresponds to $M_{k+1,k+1}$.

### 2.3.1 Similarity transformation

We show that lumping the dangling nodes in the Google matrix can be accomplished by a similarity transformation that reduces $G$ to block upper triangular form.

**Theorem 2.3.1.** *With the notation in section* 2.2 *and the matrix* $G$ *as partitioned in* (2.1)*, let*

$$X \equiv \begin{bmatrix} I_k & 0 \\ 0 & L \end{bmatrix}, \qquad where \quad L \equiv I_{n-k} - \frac{1}{n-k}\hat{e}e^T \quad and \quad \hat{e} \equiv e - e_1 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

*Then*

$$XGX^{-1} = \begin{bmatrix} G^{(1)} & * \\ 0 & 0 \end{bmatrix}, \qquad where \quad G^{(1)} \equiv \begin{bmatrix} G_{11} & G_{12}e \\ u_1^T & u_2^T e \end{bmatrix}.$$

*The matrix* $G^{(1)}$ *is stochastic of order* $k+1$ *with the same nonzero eigenvalues as* $G$*.*

*Proof.* From

$$X^{-1} = \begin{bmatrix} I_k & 0 \\ 0 & L^{-1} \end{bmatrix}, \qquad L^{-1} = I_{n-k} + \hat{e}e^T,$$

it follows that

$$XGX^{-1} = \begin{bmatrix} G_{11} & G_{12}(I + \hat{e}e^T) \\ e_1 u_1^T & e_1 u_2^T (I + \hat{e}e^T) \end{bmatrix}$$

has the same eigenvalues as $G$. In order to reveal the eigenvalues, we choose a different partitioning and separate the leading $k+1$ rows and columns, and observe that

$$G_{12}(I + \hat{e}e^T)e_1 = G_{12}e, \qquad u_2^T(I + \hat{e}e^T)e_1 = u_2^T e$$

to obtain the block triangular matrix

$$XGX^{-1} = \begin{bmatrix} G^{(1)} & * \\ 0 & 0 \end{bmatrix}$$

with at least $n-k-1$ zero eigenvalues. □

### 2.3.2 Expression for PageRank

We give an expression for the PageRank $\pi$ in terms of the stationary distribution $\sigma$ of the small matrix $G^{(1)}$.

**Theorem 2.3.2.** *With the notation in section* 2.2 *and the matrix* $G$ *as partitioned in* (2.1)*, let*

$$\sigma^T \begin{bmatrix} G_{11} & G_{12}e \\ u_1^T & u_2^T e \end{bmatrix} = \sigma^T, \qquad \sigma \geq 0, \qquad \|\sigma\| = 1$$

*and partition* $\sigma^T = \begin{bmatrix} \sigma_{1:k}^T & \sigma_{k+1} \end{bmatrix}$*, where* $\sigma_{k+1}$ *is a scalar. Then the PageRank equals*

$$\pi^T = \begin{bmatrix} \sigma_{1:k}^T & \sigma^T \begin{pmatrix} G_{12} \\ u_2^T \end{pmatrix} \end{bmatrix}.$$

*Proof.* As in the proof of Theorem 2.3.1, we write

$$XGX^{-1} = \begin{bmatrix} G^{(1)} & G^{(2)} \\ 0 & 0 \end{bmatrix},$$

where

$$G^{(2)} \equiv \begin{bmatrix} G_{12} \\ u_2^T \end{bmatrix} (I + \hat{e}e^T)[e_2 \cdots e_{n-k}].$$

The vector $\begin{bmatrix} \sigma^T & \sigma^T G^{(2)} \end{bmatrix}$ is an eigenvector for $XGX^{-1}$ associated with the eigenvalue $\lambda = 1$. Hence

$$\hat{\pi} \equiv \begin{bmatrix} \sigma^T & \sigma^T G^{(2)} \end{bmatrix} X$$

is an eigenvector of $G$ associated with $\lambda = 1$ and a multiple of the stationary distribution $\pi$ of $G$. Since $G^{(1)}$ has the same nonzero eigenvalues as $G$, and the dominant eigenvalue 1 of $G$ is distinct (32; 47; 60; 61; 107), the stationary distribution $\sigma$ of $G^{(1)}$ is unique.

Next we express $\hat{\pi}$ in terms of quantities in the matrix $G$. We return to the original partitioning which separates the leading $k$ elements,

$$\hat{\pi}^T = \begin{bmatrix} \sigma_{1:k}^T & \begin{pmatrix} \sigma_{k+1} & \sigma^T G^{(2)} \end{pmatrix} \end{bmatrix} \begin{bmatrix} I_k & 0 \\ 0 & L \end{bmatrix}.$$

Multiplying out

$$\hat{\pi}^T = \begin{bmatrix} \sigma_{1:k}^T & \begin{pmatrix} \sigma_{k+1} & \sigma^T G^{(2)} \end{pmatrix} L \end{bmatrix}$$

shows that $\hat{\pi}$ has the same leading $k$ elements as $\sigma$.

We now examine the trailing $n-k$ components of $\hat{\pi}^T$. To this end we partition the matrix $L = I_{n-k} - \frac{1}{n-k}\hat{e}e$ and distinguish the first row and column,

$$L = \begin{bmatrix} 1 & 0 \\ -\frac{1}{n-k}e & I - \frac{1}{n-k}ee^T \end{bmatrix}.$$

Then the eigenvector part associated with the dangling nodes is

$$z^T \equiv \begin{bmatrix} \sigma_{k+1} & \sigma^T G^{(2)} \end{bmatrix} L = \begin{bmatrix} \sigma_{k+1} - \frac{1}{n-k}\sigma^T G^{(2)} e & \sigma^T G^{(2)} \left(I - \frac{1}{n-k}ee^T\right) \end{bmatrix}.$$

To remove the terms containing $G^{(2)}$ in $z$, we simplify

$$(I + \hat{e}e^T)[e_2 \cdots e_{n-k}]e = (I + \hat{e}e^T)\hat{e} = (n-k)\hat{e}.$$

Hence

$$G^{(2)}e = (n-k)\begin{bmatrix} G_{12} \\ u_2^T \end{bmatrix}\hat{e} \tag{2.2}$$

and

$$\begin{aligned}\frac{1}{n-k}\sigma^T G^{(2)} e &= \sigma^T \begin{bmatrix} G_{12} \\ u_2^T \end{bmatrix} \hat{e} = \sigma^T \begin{bmatrix} G_{12} \\ u_2^T \end{bmatrix} e - \sigma^T \begin{bmatrix} G_{12} \\ u_2^T \end{bmatrix} e_1 \\ &= \sigma_{k+1} - \sigma^T \begin{bmatrix} G_{12} \\ u_2^T \end{bmatrix} e_1,\end{aligned}$$

where we used $\hat{e} = e - e_1$, and the fact that $\sigma$ is the stationary distribution of $G^{(1)}$, so

$$\sigma_{k+1} = \sigma^T \begin{bmatrix} G_{12} \\ u_2^T \end{bmatrix} e.$$

Therefore the leading element of $z$ equals

$$z_1 = \sigma_{k+1} - \frac{1}{n-k}\sigma^T G^{(2)} e = \sigma^T \begin{bmatrix} G_{12} \\ u_2^T \end{bmatrix} e_1.$$

For the remaining elements of $z$, we use (2.2) to simplify

$$G^{(2)}\left(I - \frac{1}{n-k} e e^T\right) = G^{(2)} - \frac{1}{n-k} G^{(2)} e e^T = G^{(2)} - \begin{bmatrix} G_{12} \\ u_2^T \end{bmatrix} \hat{e} e^T.$$

Replacing

$$(I + \hat{e} e^T)[e_2 \cdots e_{n-k}] = [e_2 \cdots e_{n-k}] + \hat{e} e^T$$

in $G^{(2)}$ yields

$$z_{2:n-k}^T = \sigma^T G^{(2)} \left( I - \frac{1}{n-k} e e^T \right) = \sigma^T \begin{bmatrix} G_{12} \\ u_2^T \end{bmatrix} [e_2 \cdots e_{n-k}].$$

Therefore the eigenvector part associated with the dangling nodes is

$$z = \begin{bmatrix} z_1 & z_{2:n-k}^T \end{bmatrix} = \sigma^T \begin{bmatrix} G_{12} \\ u_2^T \end{bmatrix}$$

and

$$\hat{\pi} = \begin{bmatrix} \sigma_{1:k}^T & \sigma^T \begin{pmatrix} G_{12} \\ u_2^T \end{pmatrix} \end{bmatrix}.$$

Since $\pi$ is unique, as discussed in section 2.2, we conclude that $\hat{\pi} = \pi$ if $\hat{\pi}^T e = 1$. This follows, again, from the fact that $\sigma$ is the stationary distribution of $G^{(1)}$ and $\sigma^T \left[ \begin{smallmatrix} G_{12} \\ u_2^T \end{smallmatrix} \right] e = \sigma_{k+1}$. □

### 2.3.3 Algorithm

We present an algorithm, based on Theorem 2.3.2, for computing the PageRank $\pi$ from the stationary distribution $\sigma$ of the lumped matrix

$$G^{(1)} \equiv \begin{bmatrix} G_{11} & G_{12} e \\ u_1^T & u_2^T e \end{bmatrix}.$$

The input to the algorithm consists of the nonzero elements of the hyperlink matrix $H$, the personalization vector $v$, the dangling node vector $w$, and the amplification factor $\alpha$. The output of the algorithm is an approximation $\hat{\pi}$ to the PageRank $\pi$, which is computed from an approximation $\hat{\sigma}$ of $\sigma$.

**Algorithm 2.3.3.** % Inputs: $H$, $v$, $w$, $\alpha$ Output: $\hat{\pi}$

% Power method applied to $G^{(1)}$:

Choose a starting vector $\hat{\sigma}^T = \begin{bmatrix} \hat{\sigma}_{1:k}^T & \hat{\sigma}_{k+1} \end{bmatrix}$ with $\hat{\sigma} \geq 0$, $\|\hat{\sigma}\| = 1$.

**While** not converged

$\hat{\sigma}_{1:k}^T = \alpha \hat{\sigma}_{1:k}^T H_{11} + (1-\alpha) v_1^T + \alpha \hat{\sigma}_{k+1} w_1^T$

$\hat{\sigma}_{k+1} = 1 - \hat{\sigma}_{1:k}^T e$

**end while**

% Recover PageRank:

$\hat{\pi}^T = \begin{bmatrix} \hat{\sigma}_{1:k}^T & \alpha \hat{\sigma}_{1:k}^T H_{12} + (1-\alpha) v_2^T + \alpha \hat{\sigma}_{k+1} w_2^T \end{bmatrix}$.

Each iteration of the power method applied to $G^{(1)}$ involves a sparse matrix vector multiply with the $k \times k$ matrix $H_{11}$ as well as several vector operations. Thus the dangling nodes are excluded from the power method computation. The convergence rate of the power method applied to $G$ is $\alpha$ (56). Algorithm 2.3.3 has the same convergence rate, because $G^{(1)}$ has the same nonzero eigenvalues as $G$ (see Theorem 2.3.1), but is much faster because it operates on a smaller matrix whose dimension does not depend on the number of dangling nodes. The final step in Algorithm 2.3.3 recovers $\pi$ via a single sparse matrix vector multiply with the $k \times (n-k)$ matrix $H_{12}$, as well as several vector operations.

Algorithm 2.3.3 is significantly faster than the power method applied to the full Google matrix $G$, but it retains all advantages of the power method: It is simple to implement and requires minimal storage. Unlike Krylov subspace methods, Algorithm 2.3.3 exhibits predictable convergence behavior and is insensitive to changes in the matrix (34). The methods in (68; 70) are special cases of Algorithm 2.3.3 because they allow the dangling node vector to be different from the personalization vector, thereby facilitating the implementation of TrustRank (41). TrustRank allows zero elements in the personalization vector $v$ in order to diminish the harm done by link spamming. Algorithm 2.3.3 can also be extended to the situation in which the Google matrix has several different dangling node vectors; see section 2.3.4.

The power method in Algorithm 2.3.3 corresponds to Stage 1 of the algorithm in (70). However, Stage 2 of that algorithm involves the power method on a rank-two matrix of order $n-k+1$. In contrast, Algorithm 2.3.3 simply performs a single matrix vector multiply with the $k \times (n-k)$ matrix $H_{12}$. There is no proof that the two-stage algorithm in (70) does compute the PageRank.

### 2.3.4 Several dangling node vectors

So far we have treated all dangling nodes in the same way, by assigning them the same dangling node vector $w$. However, one dangling node vector may be inadequate for an advanced Web search. For instance, one may want to distinguish different types of dangling node pages based on their functions (e.g., text files, image files, videos, etc.); or one may want to personalize a Web search and assign different vectors to dangling node pages pertaining to different topics, different languages, or different domains; see

the discussion in (70, section 8.2).

To facilitate such a model for an advanced Web search, we extend the single class of dangling nodes to $m \geq 1$ different classes, by assigning a different dangling node vector $w_i$ to each class, $1 \leq i \leq m$. As a consequence we need to extend lumping to a more general Google matrix that is obtained by replacing the $n - k$ zero rows in the hyperlink matrix $H$ by $m \geq 1$ possibly different dangling node vectors $w_1, \ldots, w_m$. The more general Google matrix is

$$F \equiv \begin{array}{c} \\ k \\ k_1 \\ \vdots \\ k_m \end{array} \begin{array}{c} \begin{array}{cccc} k & k_1 & \ldots & k_m \end{array} \\ \begin{pmatrix} F_{11} & F_{12} & \cdots & F_{1,m+1} \\ eu_{11}^T & eu_{12}^T & \cdots & eu_{1,m+1}^T \\ \vdots & \vdots & & \vdots \\ eu_{m1}^T & eu_{m2}^T & \cdots & eu_{m,m+1}^T \end{pmatrix} \end{array},$$

where

$$u_i \equiv \begin{bmatrix} u_{i1} \\ \vdots \\ u_{i,m+1} \end{bmatrix} \equiv \alpha w_i + (1-\alpha)v.$$

Let $\tilde{\pi}$ be the PageRank associated with $F$,

$$\tilde{\pi}^T F = \tilde{\pi}^T, \qquad \tilde{\pi} \geq 0, \qquad \|\tilde{\pi}\| = 1.$$

We explain our approach for the case where $F$ has two types of dangling nodes,

$$F = \begin{array}{c} \\ k \\ k_1 \\ k_2 \end{array} \begin{array}{c} \begin{array}{ccc} k & k_1 & k_2 \end{array} \\ \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ eu_{11}^T & eu_{12}^T & eu_{13}^T \\ eu_{21}^T & eu_{22}^T & eu_{23}^T \end{pmatrix} \end{array}.$$

We perform the lumping by a sequence of similarity transformations that starts at the bottom of the matrix. The first similarity transformation lumps the dangling nodes represented by $u_2$ and leaves the leading block of order $k + k_1$ unchanged,

$$X_1 \equiv \begin{array}{c} \\ k + k_1 \\ k_2 \end{array} \begin{array}{c} \begin{array}{cc} k + k_1 & k_2 \end{array} \\ \begin{pmatrix} I & 0 \\ 0 & L_1 \end{pmatrix} \end{array},$$

where $L_1$ lumps the $k_2$ trailing rows and columns of $F$,

$$L_1 \equiv I - \frac{1}{k_2}\hat{e}e^T, \quad L_1^{-1} \equiv I + \hat{e}e^T, \qquad \hat{e} = e - e_1 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Applying the similarity transformation to $F$ gives

$$X_1 F X_1^{-1} = \begin{array}{c} \\ k \\ k_1 \\ 1 \\ k_2 - 1 \end{array} \begin{array}{c} \begin{array}{cccc} k & k_1 & 1 & k_2 - 1 \end{array} \\ \begin{pmatrix} F_{11} & F_{12} & F_{13}e & \tilde{F}_{13} \\ eu_{11}^T & eu_{12}^T & (u_{13}^T e)e & e\tilde{u}_{13}^T \\ u_{21}^T & u_{22}^T & u_{23}^T e & \tilde{u}_{23}^T \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

with

$$\tilde{F}_{13} \equiv F_{13} L_1^{-1} \begin{bmatrix} e_2 & \cdots & e_{k_2} \end{bmatrix}, \qquad \tilde{u}_{j3}^T \equiv u_{j3}^T L_1^{-1} \begin{bmatrix} e_2 & \cdots & e_{k_2} \end{bmatrix}, \quad j = 1, 2.$$

The leading diagonal block of order $k + k_1 + 1$ is a stochastic matrix with the same nonzero eigenvalues as $F$. Before applying the second similarity transformation that lumps the dangling nodes represented by $u_1$, we move the rows with $u_1$ (and corresponding columns) to the bottom of the nonzero matrix, merely to keep the notation simple. The move is accomplished by the permutation matrix

$$P_1 \equiv [e_1 \cdots e_k \quad e_{k+k_1+1} \quad e_{k+1} \cdots e_{k+k_1} \quad e_{k+k_1+2} \cdots e_n].$$

The symmetrically permuted matrix

$$P_1 X_1 F X_1^{-1} P_1^T = \left[\begin{array}{ccc|c} F_{11} & F_{13}e & F_{12} & \tilde{F}_{13} \\ u_{21}^T & u_{23}^T e & u_{22}^T & \tilde{u}_{23}^T \\ eu_{11}^T & (u_{13}^T e)e & eu_{12}^T & e\tilde{u}_{13}^T \\ \hline 0 & 0 & 0 & 0 \end{array}\right]$$

retains a leading diagonal block that is stochastic. Now we repeat the lumping on dangling nodes represented by $u_1$, by means of the similarity transformation

$$X_2 \equiv \begin{array}{c} \\ k+1 \\ k_1 \\ k_2-1 \end{array} \begin{array}{c} \begin{array}{ccc} k+1 & k_1 & k_2-1 \end{array} \\ \begin{pmatrix} I & 0 & 0 \\ 0 & L_2 & 0 \\ 0 & 0 & I \end{pmatrix} \end{array},$$

where $L_2$ lumps the trailing $k_1$ nonzero rows,

$$L_2 \equiv I - \frac{1}{k_1}\hat{e}e^T, \qquad L_2^{-1} \equiv I + \hat{e}e^T.$$

The similarity transformation produces the lumped matrix

$$X_2P_1X_1FX_1^{-1}P_1^TX_2^{-1} = \begin{array}{c} \\ k \\ 1 \\ 1 \\ k_1-1 \\ k_2-1 \end{array} \begin{array}{c} \begin{array}{ccccc} k & 1 & 1 & k_1-1 & k_2-1 \end{array} \\ \begin{pmatrix} F_{11} & F_{13}e & F_{12}e & \tilde{F}_{12} & \tilde{F}_{13} \\ u_{21}^T & u_{23}^Te & u_{22}^Te & \tilde{u}_{22}^T & \tilde{u}_{23}^T \\ u_{11}^T & u_{13}^Te & u_{12}^Te & \tilde{u}_{12}^T & \tilde{u}_{13}^T \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}.$$

Finally, for notational purposes, we restore the original ordering of dangling nodes by permuting rows and columns $k+1$ and $k+2$,

$$P_2 \equiv [e_1 \cdots e_k \quad e_{k+2} \quad e_{k+1} \quad e_{k+3} \cdots e_n].$$

The final lumped matrix is

$$P_2X_2P_1X_1FX_1^{-1}P_1^TX_2^{-1}P_2^T = \begin{bmatrix} F_{11} & F_{12}e & F_{13}e & * \\ u_{11}^T & u_{12}^Te & u_{13}^Te & * \\ u_{21}^Te & u_{22}^Te & u_{23}^Te & * \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} F^{(1)} & * \\ 0 & 0 \end{bmatrix}.$$

The above discussion for $m = 2$ illustrates how to extend Theorems 2.3.1 and 2.3.2 to any number $m$ of dangling node vectors.

**Theorem 2.3.4.** *Define $X_i$ as*

$$\begin{array}{c} \\ k+(i-1)+\sum_{j=1}^{m-i} k_j \\ k_{m-i+1} \\ 1-i+\sum_{j=m-i+2}^{m} k_j \end{array} \begin{array}{c} \begin{array}{ccc} k+(i-1)+\sum_{j=1}^{m-i} k_j & k_{m-i+1} & 1-i+\sum_{j=m-i+2}^{m} k_j \end{array} \\ \begin{pmatrix} I & 0 & 0 \\ 0 & L_i & 0 \\ 0 & 0 & I \end{pmatrix} \end{array}$$

*and*

$$P_i \equiv [e_1 \cdots e_k \quad e_{r+i} \quad e_{k+1} \cdots e_{r+i-1} \quad e_{r+i+1} \cdots e_n], \qquad r = k + \sum_{j=1}^{m-i} k_j.$$

*Then*

$$P_m X_m P_{m-1} X_{m-1} \cdots P_1 X_1 F X_1^{-1} P_1^T \cdots X_m^{-1} P_m^T = \begin{bmatrix} F^{(1)} & * \\ 0 & 0 \end{bmatrix},$$

*where the lumped matrix*

$$F^{(1)} \equiv \begin{bmatrix} F_{11} & F_{12}e & \cdots & F_{1,m+1}e \\ u_{11}^T & u_{12}^T e & \cdots & u_{1,m+1}^T e \\ \vdots & \vdots & & \vdots \\ u_{m1}^T & u_{m2}^T e & \cdots & u_{m,m+1}^T e \end{bmatrix}$$

*is stochastic of order $k+m$ with the same nonzero eigenvalues as $F$.*

**Theorem 2.3.5.** *Let $\rho$ be the stationary distribution of the lumped matrix*

$$F^{(1)} \equiv \begin{bmatrix} F_{11} & F_{12}e & \cdots & F_{1,m+1}e \\ u_{11}^T & u_{12}^T e & \cdots & u_{1,m+1}^T e \\ \vdots & \vdots & & \vdots \\ u_{m1}^T & u_{m2}^T e & \cdots & u_{m,m+1}^T e \end{bmatrix}; \tag{2.3}$$

*that is,*

$$\rho^T F^{(1)} = \rho^T, \qquad \rho \geq 0, \qquad \|\rho\| = 1.$$

*With the partition $\rho^T = \begin{bmatrix} \rho_{1:k}^T & \rho_{k+1:k+m}^T \end{bmatrix}$, where $\rho_{k+1:k+m}$ is $m \times 1$, the PageRank of $F$ equals*

$$\tilde{\pi}^T = \begin{bmatrix} \rho_{1:k}^T & \rho^T \begin{pmatrix} F_{12} & \cdots & F_{1,m+1} \\ u_{12}^T & \cdots & u_{1,m+1}^T \\ \vdots & & \vdots \\ u_{m2}^T & \cdots & u_{m,m+1}^T \end{pmatrix} \end{bmatrix}.$$

## 2.4 PageRanks of dangling versus nondangling nodes

We examine how the PageRanks of dangling and nondangling nodes influence each other, as well as the effect of the dangling node vector on the PageRanks.

From Theorem 2.3.2 and Algorithm 2.3.3, we see that the PageRank $\pi_1$ of the nondangling nodes can be computed separately from the PageRank $\pi_2$ of the dangling nodes, and that $\pi_2$ depends directly on $\pi_1$. The expressions below make this even clearer.

**Theorem 2.4.1.** *With the notation in section* 2.2*,*

$$\begin{aligned}\pi_1^T &= \left((1-\alpha)v_1^T + \rho w_1^T\right)(I - \alpha H_{11})^{-1},\\ \pi_2^T &= \alpha\pi_1^T H_{12} + (1-\alpha)v_2^T + \alpha(1 - \|\pi_1\|)w_2^T,\end{aligned}$$

*where*

$$\rho \equiv \alpha\frac{1-(1-\alpha)v_1^T(I-\alpha H_{11})^{-1}e}{1+\alpha w_1^T(I-\alpha H_{11})^{-1}e} \geq 0.$$

*Proof.* Rather than using Theorem 2.3.2 we found it easier just to start from scratch. From $G = \alpha(H + dw^T) + (1-\alpha)v^T$ and the fact that $\pi^T e = 1$, it follows that $\pi$ is the solution to the linear system

$$\pi^T = (1-\alpha)v^T\left(I - \alpha H - \alpha dw^T\right)^{-1},$$

whose coefficient matrix is a strictly row diagonally dominant M-matrix (2, equation (5)), (16, equation (2), Proposition 2.4). Since $R \equiv I - \alpha H$ is also an M-matrix, it is nonsingular, and the elements of $R^{-1}$ are nonnegative (9, section 6). The Sherman–Morrison formula (36, section 2.1.3) implies that

$$\left(R - \alpha dw^T\right)^{-1} = R^{-1} + \frac{\alpha R^{-1}dw^T R^{-1}}{1-\alpha w^T R^{-1}d}.$$

Substituting this into the expression for $\pi$ gives

$$\pi^T = (1-\alpha)v^T R^{-1} + \frac{\alpha(1-\alpha)v^T R^{-1}d}{1-\alpha w^T R^{-1}d}w^T R^{-1}. \tag{2.4}$$

We now show that the denominator $1 - \alpha w^T R^{-1} d > 0$. Using the partition

$$R^{-1} = (I - \alpha H)^{-1} = \begin{bmatrix} (I - \alpha H_{11})^{-1} & \alpha (I - \alpha H_{11})^{-1} H_{12} \\ 0 & I \end{bmatrix}$$

yields

$$1 - \alpha w^T R^{-1} d = 1 - \alpha \left( \alpha w_1^T (I - \alpha H_{11})^{-1} H_{12} e + w_2^T e \right). \tag{2.5}$$

Rewrite the term involving $H_{12}$ by observing that $H_{11}e + H_{12}e = e$ and that $I - \alpha H_{11}$ is an M-matrix, so

$$0 \leq \alpha (I - \alpha H_{11})^{-1} H_{12} e = e - (1 - \alpha) (I - \alpha H_{11})^{-1} e. \tag{2.6}$$

Substituting this into (2.5) and using $1 = w^T e = w_1^T e + w_2^T e$ shows that the denominator in the Sherman–Morrison formula is positive,

$$1 - \alpha w^T R^{-1} d = (1 - \alpha) \left( 1 + \alpha w_1^T (I - \alpha H_{11})^{-1} e \right) > 0.$$

Furthermore, $0 \leq \alpha < 1$ implies $1 - \alpha w^T R^{-1} d > 1 - \alpha$.

Substituting the simplified denominator into the expression (2.4) for $\pi$ yields

$$\pi^T = (1 - \alpha) v^T R^{-1} + \alpha \frac{v^T R^{-1} d}{1 + \alpha w_1^T (I - \alpha H_{11})^{-1} e} w^T R^{-1}. \tag{2.7}$$

We obtain for $\pi_1$

$$\pi_1^T = \left( (1-\alpha)v_1^T + \alpha \frac{v^T R^{-1} d}{1 + \alpha w_1^T (I - \alpha H_{11})^{-1} e} w_1^T \right) (I - \alpha H_{11})^{-1}.$$

Combining the partitioning of $R^{-1}$, (2.6), and $v_1^T e + v_2^T e = 1$ gives

$$\begin{aligned} 0 \leq v^T R^{-1} d &= \begin{bmatrix} v_1^T & v_2^T \end{bmatrix} \begin{bmatrix} (I - \alpha H_{11})^{-1} & \alpha (I - \alpha H_{11})^{-1} H_{12} \\ 0 & I \end{bmatrix} \begin{bmatrix} 0 \\ e \end{bmatrix} \\ &= \alpha v_1^T (I - \alpha H_{11})^{-1} H_{12} e + v_2^T e \\ &= 1 - (1-\alpha) v_1^T (I - \alpha H_{11})^{-1} e. \end{aligned}$$

Hence $\pi_1^T = \left( (1-\alpha)v_1^T + \rho w_1^T \right) (I - \alpha H_{11})^{-1}$ with $\rho > 0$.

To obtain the expression for $\pi_2$, observe that the second block element in

$$\pi^T (I - \alpha H - \alpha d w^T) = (1-\alpha) v^T$$

equals

$$-\alpha \pi_1^T H_{12} + \pi_2^T - \alpha \pi_2^T e w_2^T = (1-\alpha) v_2^T.$$

The result follows from $\pi_1^T e + \pi_2^T e = 1$. □

**Remark 2.4.2.** We draw the following conclusions from Theorem 2.4.1 with regard to how dangling and nondangling nodes accumulate PageRank; see Figure 2.2.

- The PageRank $\pi_1$ of the nondangling nodes does *not* depend on the connectivity among the dangling nodes (elements of $w_2$), the personalization vector for the

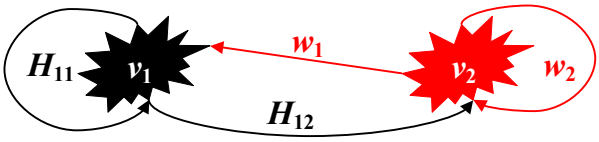


Figure 2.2: Sources of PageRank. Nondangling nodes receive their PageRank from $v_1$ and $w_1$, distributed through the links $H_{11}$. In contrast, the PageRank of the dangling nodes comes from $v_2$, $w_2$, and the PageRank of the nondangling nodes through the links $H_{12}$.

dangling nodes (elements of $v_2$), or the links from nondangling to dangling nodes (elements of $H_{12}$).

To be specific, $\pi_1$ does not depend on individual elements of $w_2$, $v_2$, and $H_{12}$. Rather, the dependence is on the norms, through $\|v_2\| = 1 - \|v_1\|$, $\|w_2\| = 1 - \|w_1\|$, and $H_{12}e = e - H_{11}e$.

- The PageRank $\pi_1$ of the nondangling nodes does not depend on the PageRank $\pi_2$ of the dangling nodes or their number, because $\pi_1$ can be computed without knowledge of $\pi_2$.

- The nondangling nodes receive their PageRank $\pi_1$ from their personalization vector $v_1$ and the dangling node vector $w_1$, both of which are distributed through the links $H_{11}$.

- The dangling nodes receive their PageRank $\pi_2$ from three sources: the associated part $v_2$ of the personalization vector; the associated part $w_2$ of the dangling node vector; and the PageRank $\pi_1$ of the nondangling nodes filtered through the connecting links $H_{12}$.

The links $H_{12}$ determine how much PageRank flows from nondangling to dangling nodes.

- The influence of the associated dangling node vector $w_2$ on the PageRank $\pi_2$ of the dangling nodes diminishes as the combined PageRank $\|\pi_1\|$ of the nondangling nodes increases.

Taking norms in Theorem 2.4.1 gives a bound on the combined PageRank of the nondangling nodes. As in section 2.2, the norm is $\|z\| \equiv z^T e$ for $z \geq 0$.

**Corollary 2.4.3.** *With the assumptions of Theorem* 2.4.1*,*

$$\|\pi_1\| = \frac{(1-\alpha)\|v_1\|_H + \alpha\|w_1\|_H}{1+\alpha\|w_1\|_H},$$

*where* $\|z\|_H \equiv z^T(I-\alpha H_{11})^{-1}e$ *for any* $z \geq 0$ *and*

$$(1-\alpha)\|z\| \leq \|z\|_H \leq \frac{1}{1-\alpha}\|z\|.$$

*Proof.* Since $(I-\alpha H_{11})^{-1}$ is nonsingular with nonnegative elements, $\|\cdot\|_H$ is a norm. Let $\|\cdot\|_\infty$ be the infinity norm (maximal row sum). Then the Hölder inequality (36, section 2.2.2) implies for any $z \geq 0$,

$$\|z\|_H \leq \|z\| \, \|(I-\alpha H_{11})^{-1}\|_\infty \leq \frac{1}{1-\alpha}\|z\|.$$

As for the lower bound,

$$\|z\|_H \geq \|z\| - \alpha z^T H_{11} e \geq (1-\alpha)\|z\|.$$

□

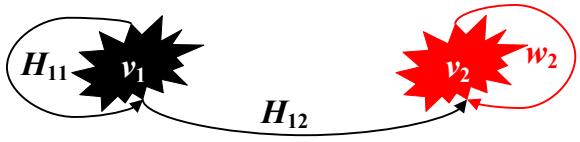


Figure 2.3: Sources of PageRank when $w_1 = 0$. The nondangling nodes receive their PageRank only from $v_1$. The dangling nodes, in contrast, receive their PageRank from $v_2$ and $w_2$, as well as from the PageRank of the nondangling nodes filtered through the links $H_{12}$.

Corollary 2.4.3 implies that the combined PageRank $\|\pi_1\|$ of the nondangling nodes is an increasing function of $\|w_1\|$. In particular, when $w_1 = 0$, the combined PageRank $\|\pi_1\|$ is minimal among all $w$ and the dangling vector $w_2$ has a stronger influence on the PageRank $\pi_2$ of the dangling nodes. The dangling nodes act like a sink and absorb more PageRank because there are no links back to the nondangling nodes; see Figure 2.3. When $w_1 = 0$ we get

$$\begin{aligned} \pi_1^T &= (1-\alpha)v_1^T(I-\alpha H_{11})^{-1}, \\ \pi_2^T &= \alpha\pi_1^T H_{12} + (1-\alpha)v_2^T + \alpha(1-\|\pi_1\|)w_2^T. \end{aligned} \tag{2.8}$$

In the other extreme case when $w_2 = 0$, the dangling nodes are not connected to each other; see Figure 2.4:

$$\begin{aligned} \pi_1^T &= \left((1-\alpha)v_1^T + \rho w_1^T\right)(I-\alpha H_{11})^{-1}, \\ \pi_2^T &= \alpha\pi_1^T H_{12} + (1-\alpha)v_2^T. \end{aligned} \tag{2.9}$$

In this case the PageRank $\pi_1$ of the nondangling nodes has only a positive influence on the PageRank of the dangling nodes.

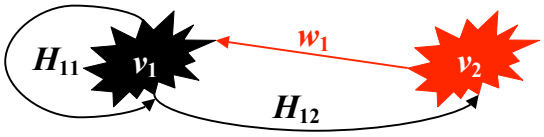


Figure 2.4: Sources of PageRank when $w_2 = 0$. The dangling nodes receive their PageRank only from $v_2$, and from the PageRank of the nondangling nodes filtered through the links $H_{12}$.

An expression for $\pi$ when dangling node and personalization vectors are the same, i.e., $w = v$, was given in (26),

$$\pi^T = (1-\alpha)\left(1 + \frac{\alpha v^T R^{-1} d}{1 - \alpha v^T R^{-1} d}\right) v^T R^{-1}, \qquad \text{where} \quad R \equiv I - \alpha H.$$

In this case the PageRank vector $\pi$ is a multiple of the vector $v^T(I - \alpha H)^{-1}$.

## 2.5 Only dangling nodes

We examine the (theoretical) extreme case when all Web pages are dangling nodes. In this case the matrices $S$ and $G$ have rank one. We first derive a Jordan decomposition for general matrices of rank one, before we present a Jordan form for a Google matrix of rank one.

We start with rank-one matrices that are diagonalizable. The vector $e_j$ denotes the $j$th column of the identity matrix $I$.

**Theorem 2.5.1** (eigenvalue decomposition)**.** *Let $A = yz^T \neq 0$ be a real square matrix with $\lambda \equiv z^T y \neq 0$. If $z$ has an element $z_j \neq 0$, then $X^{-1}AX = \lambda\, e_j e_j^T$, where*

$$X \equiv I + ye_j^T - \frac{1}{z_j} e_j z^T, \qquad X^{-1} = I - e_j e_j^T - \frac{1}{\lambda} yz^T + \frac{1+y_j}{\lambda} e_j z^T.$$

*Proof.* The matrix $A$ has a repeated eigenvalue zero and a distinct nonzero eigenvalue $\lambda$ with right eigenvector $y$ and left eigenvector $z$. From $\lambda y e_j^T = AX = \lambda X e_j e_j^T$ and $X^{-1}A = e_j z^T$ it follows that $X^{-1}X = I$ and $X^{-1}AX = \lambda e_j e_j^T$. □

Now we consider rank-one matrices that are not diagonalizable. In this case all eigenvalues are zero, and the matrix has a Jordan block of order two.

**Theorem 2.5.2** (Jordan decomposition)**.** *Let $A = yz^T \neq 0$ be a real square matrix with $z^T y = 0$. Then $y$ and $z$ have elements $y_j z_j \neq 0 \neq y_k z_k$, $j < k$. Define a symmetric permutation matrix $P$ so that $Pe_k = e_{j+1}$ and $Pe_j = e_j$. Set $\hat{y} \equiv Py$ and $\hat{u} \equiv Pz - e_{j+1}$. Then $X^{-1}AX = e_j e_{j+1}^T$ with*

$$X \equiv P\left(I + \hat{y}e_j^T - \frac{1}{\hat{u}_j} e_j \hat{u}^T\right), \qquad X^{-1} = \left(I - e_j e_j^T + \frac{1}{\hat{y}_k}\hat{y}\hat{u}^T - \frac{1+\hat{y}_j}{\hat{y}_k} e_j \hat{u}^T\right) P.$$

*Proof.* To satisfy $z^T y = 0$ for $y \neq 0$ and $z \neq 0$, we must have $y_j z_j \neq 0$ and $y_k z_k \neq 0$ for some $j < k$.

Since $A$ is a rank-one matrix with all eigenvalues equal to zero, it must have a

Jordan block of the form $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$. To reveal this Jordan block, set $\hat{z} \equiv Pz$,

$$\hat{X} \equiv \left( I + \hat{y}e_j^T - \frac{1}{\hat{u}_j} e_j \hat{u}^T \right), \qquad \hat{X}^{-1} = \left( I - e_j e_j^T + \frac{1}{\hat{y}_k} \hat{y}\hat{u}^T - \frac{1+\hat{y}_j}{\hat{y}_k} e_j \hat{u}^T \right).$$

Then the matrix $\hat{A} \equiv \hat{y}\hat{z}^T$ has a Jordan decomposition $\hat{X}^{-1}\hat{A}\hat{X} = e_j e_{j+1}^T$. This follows from $u_j = z_j$, $\hat{y}e_{j+1}^T = \hat{A}\hat{X} = \hat{X}e_j e_{j+1}^T$, and $\hat{X}^{-1}\hat{A} = e_j \hat{z}^T$.

Finally, we undo the permutation by means of $X \equiv P\hat{X}$, $X^{-1} = \hat{X}^{-1}P$, so that $X^{-1}X = I$ and $X^{-1}AX = e_j e_{j+1}^T$. □

Theorems 2.5.1 and 2.5.2 can also be derived from (51, Theorem 1.4).

In the (theoretical) extreme case when all Web pages are dangling nodes, the Google matrix is diagonalizable of rank one.

**Corollary 2.5.3** (rank-one Google matrix)**.** *With the notation in section* 2.2 *and* (2.1)*, let* $G = eu^T$*. Let* $u_j \neq 0$ *be a nonzero element of* $u$*. Then* $X^{-1}GX = e_j e_j^T$ *with*

$$X = I + ee_j^T - \frac{1}{v_j} e_j u^T$$

*and*

$$X^{-1} = I - e_j e_j^T - eu^T + 2e_j u^T.$$

*In particular,* $\pi^T = e_j^T X^{-1} = u^T$*.*

*Proof.* Since $1 = u^T e \neq 0$, the Google matrix is diagonalizable, and the expression in Theorem 2.5.1 applies. □

Corollary 2.5.3 can also be derived from (107, Theorems 2.1, 2.3).

# Chapter 3

# Ergodicity Coefficients

## 3.1 Introduction to ergodicity coefficients

Ergodicity coefficients were originally introduced in the context of rates of convergence of finite, inhomogeneous Markov chains. In general, ergodicity deals with the long term behavior of dynamical systems. Thus for systems of finite, inhomogeneous Markov chains, ergodicity refers to the long-term behavior of products of stochastic matrices.

An ergodicity coefficient $\tau_1(S)$ is defined for an $n \times n$ stochastic matrix $S$, as

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

where the maximum is taken over all $z \in \mathbb{R}^n$ and $e$ is a column vector of all 1's Notice that this form of an ergodicity coefficient is simply the norm of a matrix restricted to a subspace. In addition to being continuous and bounded, $\tau_1(S)$ is submultiplicative, and bounds the magnitude of the second largest eigenvalue of $S$. Thus $\tau_1(S_1 S_2) \leq$

$\tau_1(S_1)\tau_1(S_2)$ and $|\lambda| \leq \tau_1(S)$ for all eigenvalues $\lambda \neq 1$ of $S$ for stochastic matrices $S_1$ and $S_2$. These two properties combine to yield an upper bound on the magnitude of subdominant eigenvalues of the product of stochastic matrices. The upper bound is computed as the product of the ergodicity coefficients of the individual matrices. That is, for any eigenvalue $\lambda \neq 1$ of the stochastic matrix $S_1S_2$, $|\lambda| \leq \tau_1(S_1S_2) \leq \tau_1(S_1)\tau_1(S_2)$. This inequality gives an eigenvalue bound for products of stochastic matrices, and is useful since, in general the eigenvalues of a product of matrices are not equal to the product of the eigenvalues of the matrices.

This chapter discusses coefficient of ergodicity from a linear algebra point of view. We choose to express these coefficients as matrix norms restricted to a subspace, although we mention other expressions, as well. We present properties, and explicit forms for various ergodicity coefficients and simplify results and notation from more than 75 references spanning more than 100 years, writing everything in language more familiar to the numerical linear algebra community. We conclude with new and existing applications, including condition numbers, eigenvalue and singular value bounds.

## 3.2 Definitions and Notation

We adopt the following notation for this chapter. We denote matrices as capital letters, e.g., $A$. All vectors are assumed to be column vectors, and are denoted by lowercase letters, e.g., $a$. Unless otherwise noted in a theorem or proof, the $(i, j)$ entry of $A$ is $a_{ij}$, and the $i^{th}$ element of vector $a$ is $a_i$.

We use the common notation of $I$ as the appropriate-sized identity matrix. We use

0 as a matrix, vector, and scalar, depending on the situation. The column vector $e_i$ is the canonical vector with an 1 in the $i^{th}$ position and 0's elsewhere. The column vector of all ones is denoted $e$. We denote the transpose of $x \in \mathbb{R}^{n\times 1}$ as $x^T \in \mathbb{R}^{1\times n}$, and the conjugate transpose of $x \in \mathbb{C}^{n\times 1}$ as $x^* \in \mathbb{C}^{1\times n}$.

We define a (row) stochastic matrix, $S \in \mathbb{R}^{n\times n}$ so that $s_{ij} \geq 0$ for all $i, j$ and the elements in each row of $S$ sum to 1 $\left(\sum_{j=1}^{n} s_{ij} = 1 \text{ for } i = 1, \ldots, n\right)$, or $Se = e$. Further, we define a doubly stochastic matrix $P \in \mathbb{R}^{n\times n}$ as a matrix that is both row and column stochastic, so that $p_{ij} \geq 0$ for all $i, j$ and the elements of each row and each column sum to one, so that $Pe = e$ and $P^T e = e$. A probability vector $v$ is defined so that $v_i \geq 0$ for all $i$ and $\sum_i v_i = 1$.

We also employ several norms throughout this chapter. The one-norm of a column vector $x$ is $\|x\|_1 = \sum_i |x_i|$, and the infinity-norm of a column vector $x$ is $\|x\|_\infty = \max_i |x_i|$. The $p$-norm of a column vector $x$ is $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$. The one-norm of a matrix $A$ is the maximal column sum, $\|A\|_1 = \max_j \sum_i |a_{ij}|$, and the infinity-norm is the maximal row sum, $\|A\|_\infty = \max_i \sum_j |a_{ij}|$. In general, the $p$-norm of a matrix $A \in \mathbb{C}^{m\times n}$ is $\|A\|_p = \max_{\|x\|_p=1} \|Ax\|_p$ for $x \in \mathbb{C}^n$. The Frobenius matrix norm is $\|A\|_F$, with $\|A\|_F^2 = \sum_{i,j} |a_{ij}|^2 = \text{trace}(A^*A)$ (76, p 279).

Finally, we use the terms *ergodicity coefficient* and *coefficient of ergodicity* interchangeably.

## 3.3 Weak ergodicity

To understand the history and development of coefficients of ergodicity we begin with a short discussion of the history of weak ergodicity for products and sequences of stochastic matrices. Seneta (95, §1),(97, §1), (98, §§4.3-4.4) credits Kolmogorov (65) with the following definition of weak ergodicity for a sequence.

**Definition 3.3.1.** *Let* $\{S_k\}$, $k \geq 1$, *be a sequence of* $n \times n$ *stochastic matrices, and define* $T^{(p,r)} = S_{p+1}S_{p+2}\cdots S_{p+r}$, *with* $(i,j)$ *entry* $t_{ij}^{(p,r)}$. *The sequence* $\{S_k\}$ *is weakly ergodic if for all* $i, j, s = 1, \ldots, n$ *and* $p \geq 0$,

$$\left(t_{is}^{(p,r)} - t_{js}^{(p,r)}\right) \to 0 \text{ as } r \to \infty.$$

Essentially, a sequence of stochastic matrices is weakly ergodic if, as the number of factors in a product approaches infinity, the rows of the product tend to equalize.

Throughout the ergodicity literature, there appears to be some debate over the origins of necessary and sufficient conditions for weak ergodicity. Seneta gives excellent summaries of the discussion of the history and development of these conditions in (95, §1), (97, §1), and (98, Ch 3-4). For additional information, some ergodicity papers include those by Cohn (23; 24), Dobrushin (28), Hajnal (42; 43), Kingman (63), and Paz and Reichaw (82). These papers are available in English, and we've found them to be more accesible and available than some of the older or foreign-language papers, including some from the 1920s-1950s by Bernštein, Doeblin, Dynkin, Sapogov and Sarymsakov (10; 11; 12; 13; 14; 15; 27; 29; 30; 92; 93; 94).

## 3.4 General ergodicity coefficients for stochastic matrices

As Lešanovský (72) points out, there are two forms for ergodicity coefficient that came about due to different problems and applications. We first present the more basic definitions, which is somewhat restrictive since the ergodicity coefficient in this case only takes values between 0 and 1. Second, we introduce an expression that, although bounded, is not restricted to lie in $[0, 1]$. This second definition provides a foundation for the most common form of the ergodicity coefficient, and allows for extension beyond stochastic matrices (see Section 3.6).

### 3.4.1 First class of ergodicity coefficients

**Definition 3.4.1** (p 509 in (95), Def 4.6 in (98))**.** *A coefficient of ergodicity, $\mu(\cdot)$ is a continuous scalar function defined for stochastic matrices $S$, so that $0 \leq \mu(S) \leq 1$.*

Along with this definition, we define a special case of the ergodicity coefficient called a *proper ergodicity coefficient*. When an ergodicity coefficient of a stochastic matrix is proper, we know that it achieves its minimum value of 0 if and only if the matrix has rank equal to 1.

**Definition 3.4.2** (Definition 4.6 in (98))**.** *For a stochastic $n \times n$ matrix $S$, a coefficient of ergodicity is proper if*

$$\mu(S) = 0 \iff S = ev^T$$

*for some probability vector $v$.*

The only way for a stochastic matrix $S$ to be a rank-one matrix is for all rows of $S$ to be identical. Thus, the statement $\text{rank}(S) = 1$ is equivalent to $S = ev^T$ in Definition 3.4.2.

There is some contradictory notation associated with the term proper defined in Definition 3.4.2. Sometimes the coefficient is defined as $\hat{\mu}(S) = 1 - \mu(S)$ and is called proper if: $\hat{\mu}(S) = 1 \iff \text{rank}(S) = 1$ (21), (43, §2), (53, p 56), (63, §4), (95, §2).

We now define weak ergodicity of products of stochastic matrices. From the definition of weak ergodicity of a sequence of stochastic matrices given in Definition 3.3.1, we learn that a sequence is weakly ergodic if as the number of factors in a products grows, the rows of the product tend towards equality. When the rows are equal, we have obtained a rank-one matrix, since our original matrices are stochastic. Proper ergodicity coefficients are zero for stochastic, rank-one matrices. Combining the concepts presented in this paragraph, we get the following relation between weak ergodicity and ergodicity coefficients.

**Theorem 3.4.3** (p 136 in (98))**.** *The product $T^{(p,r)} = S_{p+1}S_{p+2}\cdots S_{p+r}$ formed from a sequence of $n \times n$ stochastic matrices $\{S_k\}$, $k \geq 1$, is weakly ergodic if for $p \geq 0$, as $r \to \infty$,*

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

*for a proper ergodicity coefficient $\mu(\cdot)$.*

Some examples of proper ergodicity coefficients for a stochastic matrix $S \in \mathbb{R}^{n\times n}$

include (98, p 137)

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

$$\alpha(S) = \max_{k} \max_{ij} |s_{ik} - s_{jk}|$$

$$\beta(S) = 1 - \sum_{k=1}^{n} \min_{i} s_{ik}$$

We present a basic relation between these proper ergodicity coefficients and an improper ergodicity coefficient, $\gamma(S) = 1 - \max_{k} \min_{i} s_{ik}$. It is easy to see that $\gamma(S)$ is not proper. Consider the rank-1 stochastic matrix $S = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix}$. Then $\gamma(S) = 1 - \max\{1/3, 2/3\} = 1/3 \neq 0$, so $\gamma(S)$ is not proper.

**Theorem 3.4.4** (p 56 in (53), pp. 137–138 in (98))**.** *Let* $\gamma(S) = 1 - \max_{k} \min_{i} s_{ik}$ *be an improper ergodicity coefficient. Then*

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

*Proof.* We will present this proof as three separate parts:

1. $\alpha(S) \leq \tau_1(S)$
2. $\tau_1(S) \leq \beta(S)$
3. $\beta(S) \leq \gamma(S)$

We begin with the first inequality, and the least intuitive of the three parts.

1. $\alpha(S) \leq \tau_1(S)$.

First, recall $\alpha(S) = \max_k \max_{ij} |s_{ik} - s_{jk}|$ and $\tau_1(S) = \frac{1}{2} \max_{ij} \sum_{k=1}^{n} |s_{ik} - s_{jk}|$. Choose indices $l, m$ and $r$ so that $\alpha(S) = s_{mr} - s_{lr} \geq 0$. We begin with

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

The absolute values are removed by breaking up the sums as

$$2\tau_1(S) \geq \sum_{k \in \mathcal{P}_m} (s_{mk} - s_{lk}) + \sum_{k \in \mathcal{P}_l} (s_{lk} - s_{mk})$$

where $\mathcal{P}_m$ is the set of all indices $k$ with $s_{mk} \geq s_{lk}$ and $\mathcal{P}_l$ is the set of all indices $k$ with $s_{mk} < s_{lk}$. Breaking up the sums further yields

$$2\tau_1(S) \geq \sum_{k \in \mathcal{P}_m} s_{mk} - \sum_{k \in \mathcal{P}_m} s_{lk} + \sum_{k \in \mathcal{P}_l} s_{lk} - \sum_{k \in \mathcal{P}_l} s_{mk}.$$

Since $S$ is a row stochastic matrix, $\sum_{k \in \mathcal{P}_m} s_{mk} = 1 - \sum_{k \in \mathcal{P}_l} s_{mk}$ and $\sum_{k \in \mathcal{P}_l} s_{lk} = 1 - \sum_{k \in \mathcal{P}_m} s_{lk}$. Thus,

$$\begin{aligned} 2\tau_1(S) &\geq \left(1 - \sum_{k \in \mathcal{P}_l} s_{mk}\right) - \sum_{k \in \mathcal{P}_m} s_{lk} + \left(1 - \sum_{k \in \mathcal{P}_m} s_{lk}\right) - \sum_{k \in \mathcal{P}_l} s_{mk} \\ &= 2\left(1 - \sum_{k \in \mathcal{P}_l} s_{mk} - \sum_{k \in \mathcal{P}_m} s_{lk}\right). \end{aligned}$$

Thus

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

Now, we relate this to our $\alpha(S) = s_{mr} - s_{lr}$. Since $\alpha(S) \geq 0$, the index $r$ must be in $\mathcal{P}_m$. Then again using the fact that $1 - \sum_{k\in\mathcal{P}_l} s_{mk} = \sum_{k\in\mathcal{P}_m} s_{mk}$

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

Finally, we pull out the $r^{th}$ term from each of these sums to obtain

$$\tau_1(S) \geq \sum_{\substack{k\in\mathcal{P}_m \\ k\neq r}} s_{mk} + s_{mr} - \sum_{\substack{k\in\mathcal{P}_m \\ k\neq r}} s_{lk} - s_{lr} = \sum_{\substack{k\in\mathcal{P}_m \\ k\neq r}} s_{mk} - \sum_{\substack{k\in\mathcal{P}_m \\ k\neq r}} s_{lk} + \alpha(S).$$

Since the indices $k \in \mathcal{P}_m$ are those with $s_{mk} \geq s_{lk}$ we see that $\sum_{\substack{k\in\mathcal{P}_m \\ k\neq r}} s_{mk} - \sum_{\substack{k\in\mathcal{P}_m \\ k\neq r}} s_{lk} \geq 0$ and conclude

$$\tau_1(S) \geq \alpha(S).$$

2. $\tau_1(S) \leq \beta(S)$

In Corollary 3.5.6 we will prove that there are two explicit forms for $\tau_1(S)$,

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

For now, we choose to simply focus on the choice of $\tau_1(S) = 1 - \min_{ij}\sum_{k=1}^{n}\min\{s_{ik}, s_{jk}\}$ and recall $\beta(S) = 1 - \sum_{k=1}^{n}\min_i s_{ik}$. We now prove $1 - \tau_1(S) \leq 1 - \beta(S)$.

Choose $l$ and $m$ as the indices $i$ and $j$ of the rows of $S$ that achieve the minimum in the expression for $1 - \tau_1(S)$. Then $1 - \tau_1(S) = \sum_{k=1}^{n}\min\{s_{lk}, s_{mk}\}$, and clearly, $\sum_{k=1}^{n}\min\{s_{lk}, s_{mk}\} \geq \sum_{k=1}^{n}\min_i s_{ik} = 1 - \beta(S)$.

3. $\beta(S) \leq \gamma(S)$

Recall $\beta(S) = 1 - \sum_{k=1}^{n} \min_i s_{ik}$ and $\gamma(S) = 1 - \max_k \ \min_i s_{ik}$. The value of $1 - \beta(S)$ is the sum of the smallest elements in each column. This value is greater than or equal to the largest value from the set of smallest elements taken from each column, or

$$\sum_{k=1}^{n} \min_i s_{ik} \geq \max_k \min_i s_{ik}.$$

□

For additional relations between these coefficients see (80), (95, §1).

### 3.4.2 Second class of ergodicity coefficients

This second form provides the framework for the most common form of an ergodicity coefficient.

**Definition 3.4.5** ((72), §2 in (97), §3.1 in (98))**.** *For every $n \times n$ stochastic matrix $S$, and metric $d$ defined on the set $D = \{x : x \in \mathbb{R}^n, x \geq 0, \sum_{i=1}^{n} x_i = 1\}$*

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

*is called an ergodicity coefficient.*

Unfortunately, there exist situations where the two forms of ergodicity coefficients, $\mu(S)$ and $\tau(S)$, can contradict one another (72). First consider choosing the metric $d(x^T, y^T) = \|x^T - y^T\|_1$ which satisfies both ergodicity coefficient definitions because

it is continuous, bounded between 0 and 1, and generated by a metric $d$ (see Section 3.5.1). Contrast this with $\mu(S) = s_{11}$. Clearly $\mu(S)$ is continuous and $0 \leq \mu(S) \leq 1$, however, $\mu(S) = s_{11}$ cannot be generated by a metric $d$ on $D$, which contradicts Definition 3.4.5. Finally, choose $d(x^T, y^T) = \|x^T - y^T\|_\infty$ (see Section 3.5.2), so that $\tau(S)$ can take values greater than 1, and thus contradicts the initial definition of an ergodicity coefficient in Definition 3.4.1.

There are two conventional choices for the metric $d$ in Definition 3.4.5. For more information on these, and other metric choices, see (72; 86; 87; 91), (97, §2) .

1. Let
$$d(x, y) = \ln\left(\frac{\max_i \frac{x_i}{y_i}}{\min_i \frac{x_i}{y_i}}\right) = \max_{ij} \ln\left(\frac{x_i y_j}{x_j y_i}\right)$$
for $x, y \in D$, $x > 0, y > 0$.

   This choice of coefficient, often called Birkhoff's Contraction Coefficient, will not be our main focus in this work. For more information, Hajnal (44, (7)) discusses some basic properties of this ergodicity coefficient. Seneta devoted a section of his book to this coefficient (98, §3.4). More recently, Artzrouni and colleagues, wrote several papers that employ this form of an ergodicity coefficient (3; 4; 5). Instead of stochastic matrices, this coefficient is often employed with row-allowable matrices (3; 4; 5), (98, Chapter 3), which are nonnegative matrices with at least one positive entry in each row (98, Definition 3.1).

2. Let $d(x, y) = \|x - y\|_p$ for any vector norm on $\mathbb{R}^n$. Then
$$\tau(S) = \sup_{\substack{x,y,\in D \\ x \neq y}} \frac{\left\|(x-y)^T S\right\|_p}{\|x - y\|_p}.$$

The continuity of $\|\cdot\|_p$, with $z = x - y$, allows us to re-write this expression as

$$\tau(S) = \sup_{\substack{\|z\|_p=1 \\ z^T e=0}} \left\| z^T S \right\|_p \tag{3.1}$$

taken over all such $z \in \mathbb{R}^n$.

We believe the expressions for normed ergodicity coefficients presented here first appeared in a 1956 paper by Dobrushin (28, §1.4-1.5), although Rhodius (87, §1) cites Seneta (97, §2) with the introduction of the normwise form of the ergodicity coefficient in equation (3.1).

The expression in equation (3.1) is the form of the ergodicity coefficient for stochastic matrices that seems to appear most often. Unfortunately, there is some confusion associated with this definition, because many authors, e.g. (38; 45; 46; 71; 72; 85; 87; 88; 89; 91; 97; 99; 101; 105; 109; 110), have assumed that $\|x\| = \left\|x^T\right\|$ for vector $x$. We choose not to make this assumption and instead present new notation in order to be consistent with later defined explicit forms for specific $p$-norms. We will use the following definition for $z \in \mathbb{R}^n$, with $S \in \mathbb{R}^{n \times n}$:

$$\tau_p(S) = \max_{\substack{\|z\|_p=1 \\ z^T e=0}} \left\| S^T z \right\|_p, \tag{3.2}$$

where the subscript $p$ has been added to $\tau(S)$ to indicate our focus on the $p$-norm. In addition to changing the order of the matrix-vector multiplication, we have changed the supremum to a maximum. We can make the switch from supremum to maximum because $\tau_p(S)$ is a continuous, real-valued function on a bounded subset of a finite-

dimensional space (37, p 206-207).

The ergodicity coefficient defined in equation (3.2) is our focus throughout the remainder of this work.

## 3.5 Specific ergodicity coefficients for stochastic matrices

Recall the coefficient of ergodicity expression from (3.2), $\tau_p(S) = \max_{\|z\|_p=1,\ z^T e=0} \left\|S^T z\right\|_p$ for $z \in \mathbb{R}^n$ and stochastic $S \in \mathbb{R}^{n\times n}$. This section focuses on bounds, relations and explicit expressions for two common choices for $p$-norm, $p = 1$ and $p = \infty$. Additionally, we mention results for the ergodicity coefficient for stochastic matrices with the general $p$-norm.

### 3.5.1 Ergodicity coefficients of stochastic matrices in the 1-norm

To place emphasis on the choice of the $l_1$-norm, write $\tau_p(\cdot)$ from equation (3.2) as $\tau_1(\cdot)$, and obtain

$$\tau_1(S) = \max_{\substack{\|z\|_1=1 \\ z^T e=0}} \left\|S^T z\right\|_1 \tag{3.3}$$

where the maximum ranges over all $z \in \mathbb{R}^n$ (97, §2) (98, §4.3). This form of the ergodicity coefficient is also commonly referred to as the Dobrushin or delta coefficient (58; 81; 109).

The ergodicity coefficient $\tau_1(S)$ is the norm of a matrix restricted to a subspace.

Specifically, $e$ is a right eigenvector of $S$ associated with dominant eigenvalue 1, ($Se = e$). The following argument shows that left eigenvectors associated with eigenvalues $\lambda \neq 1$ are orthogonal to $e$. Let $v$ be a left eigenvector of $\lambda \neq 1$ of $S$, ($v^T S = \lambda v^T$). Then from $\lambda v^T e = v^T S e = v^T e$ and $\lambda \neq 1$, we conclude that $v^T e = 0$. Since left eigenvectors associated with eigenvalues $\lambda \neq 1$ are orthogonal to $e$, the subspace associated with $\tau_1(S)$ is the space spanned by left eigenvectors associated with eigenvalues $\lambda \neq 1$.

We have already shown that $\tau_1(\cdot)$ is an ergodicity coefficient in the sense of Definition 3.4.5 in Section 3.4. We now illustrate that the original definition, Definition 3.4.1, also holds, by proving that $0 \leq \tau_1(S) \leq 1$ and $\tau_1(S)$ is continuous for any stochastic matrix $S$.

**Theorem 3.5.1** (§4.3 in (98))**.** *The function $\tau_1(\cdot)$ is a coefficient of ergodicity in the sense of Definition 3.4.1. That is, for stochastic matrices $S$, $S_1$ and $S_2$*

1. $0 \leq \tau_1(S) \leq 1$

2. $|\tau_1(S_1) - \tau_1(S_2)| \leq \|S_1 - S_2\|_\infty$.

*Proof.* 1. We prove this directly using the definition of a norm. Clearly $0 \leq \tau_1(S) = \max_{\|z\|_1=1, z^T e=0} \left\|S^T z\right\|_1$. Moreover,

$$\max_{\|z\|_1=1, z^T e=0} \left\|S^T z\right\|_1 \leq \max_{\|z\|_1=1} \left\|S^T z\right\|_1$$

because the left side of the inequality is more restrictive than the right side. We

conclude

$$0 \leq \tau_1(S) = \max_{\substack{\|z\|_1=1 \\ z^T e=0}} \left\|S^T z\right\|_1 \leq \max_{\|z\|_1=1} \left\|S^T z\right\|_1 = \left\|S^T\right\|_1 = \|S\|_\infty = 1.$$

2. Without loss of generality, assume $\tau_1(S_1) \geq \tau_1(S_2)$. Also, let the maximum of $\tau_1(S_1)$ be assumed by $y$, so that $\tau_1(S_1) = \left\|S_1^T y\right\|_1$, with $\|y\|_1 = 1$ and $y^T e = 0$. We write

$$\tau_2(S_2) = \max_{\substack{\|z\|_1=1 \\ z^T e=0}} \left\|S_2^T z\right\|_1 \geq \left\|S_2^T y\right\|_1 .$$

Then

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

Because $\|y\|_1 = 1$ and $\left\|S^T\right\|_1 = \|S\|_\infty$ we obtain

$$\left\|S_1^T y\right\|_1 - \left\|S_2^T y\right\|_1 \leq \left\|\left(S_1^T - S_2^T\right) y\right\|_1 \leq \left\|(S_1 - S_2)^T\right\|_1 = \|S_1 - S_2\|_\infty$$

We have now proven that $|\tau_1(S_1) - \tau_1(S_2)| \leq \|S_1 - S_2\|_\infty$, and thus the coefficient of ergodicity $\tau_1(S)$ is continuous.

Clearly, $\tau_1(S)$ satisfies the conditions for an ergodicity coefficient as stated in Definition 3.4.1. □

The ergodicity coefficient $\tau_1(S)$ is a proper ergodicity coefficient for any stochastic matrix $S$.

**Theorem 3.5.2** (Definition 4.6 in (98)). *The function $\tau_1(\cdot)$ is a proper coefficient of ergodicity in the sense of Definition 3.4.2. That is, for stochastic matrix $S \in \mathbb{R}^{n\times n}$,*

$$\tau_1(S) = 0 \iff \text{rank}(S) = 1.$$

*Proof.* If $\tau_1(S) = 0$ then $\left\|S^T z\right\|_1 = 0$ for any $z$ with $z^T e = 0$ and $\|z\|_1 = 1$. In particular, let $z = \frac{1}{2}(e_i - e_j)$, with $i, j = 1, \ldots, n$, $i \neq j$, so that any two rows of $S^T$ are identical. Thus, $S$ is a rank-one matrix.

If $\text{rank}(S) = 1$, then $S$ can be written $S = ev^T$ for some probability vector $v$. Then $S^T z = ve^T z = 0$ if $z$ is chosen so that $z^T e = 0$. Hence $\tau_1(S) = 0$. □

The ergodicity coefficient can provide bounds on the subdominant eigenvalues of a matrix or product of matrices. We now show that $\tau_1(\cdot)$ is an upper bound on all non-unit eigenvalues of stochastic matrices and that $\tau_1(\cdot)$ is submultiplicative.

**Theorem 3.5.3** (§5.2 in (6)). *For stochastic matrices $S$, $S_1$ and $S_2$*

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

*Proof.* 1. If $\lambda \neq 1$ is a real eigenvalue of $S$, then there is a real left eigenvector $x$ with $x^T S = \lambda x^T$ and $\|x\|_1 = 1$. Since $x$ is a left eigenvector and $e$ is a right eigenvector for a different eigenvalue, $v$ and $e$ are orthogonal, so $v^T e = 0$. Hence

$$|\lambda| = |\lambda|\|x\|_1 = \|\lambda x\|_1 = \left\|S^T x\right\|_1 \leq \max_{\substack{\|x\|=1 \\ x^T e = 0}} \left\|S^T x\right\|_1 = \tau_1(S).$$

Hence, $\tau_1(S)$ is an upper bound for all subdominant eigenvalues of $S$.

If $\lambda \neq 1$ is a complex eigenvalue, then its eigenvectors $v$ are complex as well. Seneta (97, §3.2) proves the bound by making use of our Lemma 3.5.4 to show that the explicit expression for $\tau_1$ in Theorem 3.5.5 holds when the maximum ranges over complex vectors. A different approach is presented in (6, §5.2) and (91, Theorem 3.1), in which seminorms are constructed that give a representation of the complex eigenvector as a real vector.

2. Let $y$ achieve the maximum of $\tau_1(S_1 S_2)$, so that $\tau_1(S_1 S_2) = \left\| (S_1 S_2)^T y \right\|_1$ with $\|y\|_1 = 1$ and $y^T e = 0$. Define $x = S_1^T y / \left\| S_1^T y \right\|_1$. Clearly, $\|x\|_1 = \left\| S_1^T y \right\|_1 / \left\| S_1^T y \right\|_1 = 1$ and $x^T e = y^T S_1 e / \left\| S_1^T y \right\|_1 = y^T e / \left\| S_1^T y \right\|_1 = 0$. Then

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

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

Thus $\tau_1(S)$ is submultiplicative for stochastic matrices $S$.

□

According to papers published in 1972 by Zenger (114) and in 1993 by Seneta (103, p 191), the ergodicity coefficient as an eigenvalue bound originated with E. Deutsch (6, p 295) in a paper by Bauer, Deutsch, and Stoer. In the 1970's Seneta (96) and Pykh (84) stated the eigenvalue bound but did not prove it. In 1979 it was stated and proven by Seneta (97, §3) who is sometimes credited with the introduction of this eigenvalue bound, e.g. Rhodius (87; 88). In their 1985 paper, Rothblum and Tan (91, §1) give a survey of various subdominant eigenvalue bounds. The eigenvalue bound also appears in (45; 89).

The origin of the submultiplicative property of the ergodicity coefficient is unclear. In 1993 Seneta (103, p 191) credited this property to a 1956 paper by Dobrushin (28). This property has also been ascribed to Kingman, based on a paper from 1975 (63, (4.10)). Additionally, the submultiplicative property is given by Seneta (97, §2), (98, Lemma 4.2), Paun (79), Rhodius (89), and Tan (110).

**Explicit expressions for $\tau_1(S)$**

In a paper from 1906 (74, pp 358-359), (73; 75), Markov may have been the first to present an explicit form of the ergodicity coefficient $\tau_1(S)$ in a construction of the Weak Law of Large Numbers. Markov defines a quantity $0 < H < 1$ where

$$H = \frac{1}{2} \max_{ij} \sum_k |s_{ik} - s_{jk}| \tag{3.4}$$

for a stochastic matrix $S$. Seneta comments on this use of $H$, and Markov's work in (104; 105). It is not clear why Markov did not realize that $H$ can take the values 0 and 1, so that $0 \leq H \leq 1$.

In a 1956 paper on the Central Limit Theorem for Markov chains Dobrushin (28) defined quantities $N(S)$ and $\alpha(S)$ for any stochastic matrix $S$, with

$$N(S) = \sup_{x,y} \frac{\|Sx - Sy\|}{\|x - y\|} \qquad \text{and} \qquad \alpha(S) = \min_{ij} \sum_k \min\{s_{ik}, s_{jk}\}$$

to show that

$$N(S) = 1 - \alpha(S) = 1 - \min_{ij} \sum_k \min\{s_{ik}, s_{jk}\}.$$

These two definitions may explain why Seneta's refers to $\tau_1(S)$ as the "Markov-Dobrushin coefficient of ergodicity" (105). Perhaps Paz in 1971 (81) was the first to introduce the term $\delta$-coefficient for $\delta(S) = 1 - \min_{ij} \sum_k \min\{s_{ik}, s_{jk}\}$, which was later adopted by other authors, including Tan (110) and Rhodius (85; 88; 89). Notice, you have seen the expressions for $H$ and $1 - \alpha(S)$ in Subsection 3.4.1 as $\tau_1(S)$ (also note, the $\alpha$ in this section differs from the $\alpha$ in Subsection 3.4.1).

Before presenting the theorem for the explicit form, we state and show that a vector $x$ that satisfies $x^T e = 0$, can be represented as a linear combination with non-negative coefficients of vectors of the form $e_i - e_j$. This lemma forms the basis of the proof of Theorem 3.5.5.

**Lemma 3.5.4** (Lemma 2.4, p 62 in (98))**.** *Let $x \in \mathbb{R}^n$ satisfy $x \neq 0$ and $x^T e = 0$. Then $x$ can be written*

$$x = \sum_{i \neq j} \frac{y_{ij}}{2} (e_i - e_j)$$

*for $y_{ij} \geq 0$ and $\sum_{i \neq j} y_{ij} = \|x\|_1$ with $i, j = 1, \ldots, n$.*

*Proof.* We prove this lemma by induction on the length $n$ of the vector $x$. When $n = 2$, assume $x_1 > 0$ so that $x = x_1 \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, because $x^T e = 0$. If $y_{12}/2 = x_1$, then $x = y_{12}\frac{1}{2}(e_1 - e_2)$ and $y_{12} = 2x_1 = \|x\|_1 > 0$.

Now assume the lemma holds for some $n \geq 2$ and we will show it holds for $n+1$. Let $x$ be a vector of length $n+1$ with $x^T e = 0$ and $x \neq 0$. Without loss of generality, assume

$x$ has already been permuted so that $x_n > 0$, $x_{n+1} < 0$ and $x_n = \max_{i=1,\ldots,n+1} |x_i|$. Then

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

Set $\hat{x} = \begin{pmatrix} x_{1:n-1} \\ x_n + x_{n+1} \end{pmatrix}$, where $\hat{x} \neq 0$ and $\hat{x}^T e = x^T e = 0$. Applying the induction hypothesis to $\hat{x}$ yields

$$\hat{x} = \sum_{i \neq j} \frac{\hat{y}_{ij}}{2} (e_i - e_j) \text{ with } \hat{y}_{ij} \geq 0 \text{ and } \sum_{i \neq j} \hat{y}_{ij} = \|\hat{x}\|_1 .$$

We now need to relate $\hat{x}$ to our $n+1$ dimensional vector $x$. Let $y_{n,n+1} = -2x_{n+1} > 0$, then

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

Since $x_n > 0$ and $x_{n+1} < 0$, then $x_n + x_{n+1} > 0$, so

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

To complete the proof, notice that $x$ takes the appropriate form, since

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

□

Applying this form of vector $x$, we express an explicit form for $\tau_1(S)$.

**Theorem 3.5.5** (§3.1, §4.3 in (98)). *For every stochastic matrix $S \in \mathbb{R}^{n\times n}$*

$$\tau_1(S) = \frac{1}{2}\max_{ij}\left\|S^T\left(e_i - e_j\right)\right\|_1$$

*Proof.* Let $y = \frac{1}{2}(e_i - e_j)$ for some $i \neq j$, $i, j = 1, \ldots, n$. Clearly, $\|y\|_1 = 1$ and $y^T e = 0$. Then

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

Now for the other direction, we apply Lemma 3.5.4 to some vector $x$ chosen to achieve the maximum of $\tau_1(S)$. That is, $\tau_1(S) = \left\|S^T x\right\|_1$, with $\|x\|_1 = 1$ and $x^T e = 0$. Then from the lemma, $x = \sum_{i\neq j} y_{ij}\frac{1}{2}\left(e_i - e_j\right)$ and $\sum_{i\neq j} y_{ij} = \|x\|_1 = 1$. The triangle inequality applied to $x$ implies

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

□

Notice that since $\max_{ij}\left\|S^T\left(e_i - e_j\right)\right\|_1 = \max_{ij}\sum_k |s_{ik} - s_{jk}|$ then $\tau_1(S) = H$ where $H$ was defined in equation (3.4). Thus, this expression for $\tau_1(S)$ dates to 1906 and Markov (74, pp 358-359), (73; 75).

We now give a second explicit form for $\tau_1(S)$.

**Corollary 3.5.6** (§3.1, §4.3 in (98), pp 1733–1734 in (52))**.** *For every stochastic matrix* $S \in \mathbb{R}^{n\times n}$

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

*Proof.* We begin with the explicit expression, from Theorem 3.5.5, $\tau_1(S) = \frac{1}{2}\max_{ij} \left\|S^T(e_i - e_j)\right\|_1$, by defining indices $l, m$ so that

$$2\tau_1(S) = \left\|S^T(e_m - e_l)\right\|_1 = \left\|(e_m - e_l)^T S\right\|_\infty = \sum_{k=1}^{n} |s_{mk} - s_{lk}|.$$

Re-writing this expression by breaking up the sum to remove the absolute values yields

$$2\tau_1(S) = \sum_{k\in\mathcal{P}_m} (s_{mk} - s_{lk}) + \sum_{k\in\mathcal{P}_l} (s_{lk} - s_{mk})$$

where $\mathcal{P}_m$ is the set of all indices $k$ with $s_{mk} \geq s_{lk}$ and $\mathcal{P}_l$ is the set of all indices $k$ with $s_{lk} > s_{mk}$. Breaking up these sums even further yields

$$2\tau_1(S) = \sum_{k\in\mathcal{P}_m} s_{mk} - \sum_{k\in\mathcal{P}_m} s_{lk} + \sum_{k\in\mathcal{P}_l} s_{lk} - \sum_{k\in\mathcal{P}_l} s_{mk}.$$

Since $S$ is a row stochastic matrix, $\sum_{k\in\mathcal{P}_m} s_{mk} = 1 - \sum_{k\in\mathcal{P}_l} s_{mk}$ and $\sum_{k\in\mathcal{P}_l} s_{lk} = 1 - \sum_{k\in\mathcal{P}_m} s_{lk}$, we obtain

$$\begin{aligned} 2\tau_1(S) &= \left(1 - \sum_{k\in\mathcal{P}_l} s_{mk}\right) - \sum_{k\in\mathcal{P}_m} s_{lk} + \left(1 - \sum_{k\in\mathcal{P}_m} s_{lk}\right) - \sum_{k\in\mathcal{P}_l} s_{mk} \\ &= 2\left(1 - \sum_{k\in\mathcal{P}_l} s_{mk} - \sum_{k\in\mathcal{P}_m} s_{lk}\right). \end{aligned}$$

Thus

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

The elements of $S$ in the above sums are the smaller of the pair $(s_{mk}, s_{lk})$ for each $k$, i.e., for all $k \in \mathcal{P}_l$, $s_{mk} < s_{lk}$ and for all $k \in \mathcal{P}_m$, $s_{lk} \leq s_{mk}$. Thus, $\tau_1(S)$ can be written

$$\tau_1(S) = 1 - \sum_{k=1}^{n} \min\left\{s_{mk}, s_{lk}\right\}.$$

Because $\left\|S^T\left(e_m - e_l\right)\right\|_1 = \max_{ij} \left\|S^T\left(e_i - e_j\right)\right\|_1$, we obtain

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

□

In the ergodicity literature, the choice of expression for the coefficient often depends on the application. This may explain why some authors, (28; 43; 52; 57; 82; 84), chose $1 - \tau_1(S) = \min_{ij} \sum_{k=1}^{n} \min\{s_{ik}, s_{jk}\}$, which also takes values in the range $[0, 1]$.

**Scrambling matrices: the restriction that $\tau_1(S) < 1$**

When using $\tau_1(S)$ as an upper bound on eigenvalues, it is important to know when $\tau_1(S) < 1$, since it is this case that gives the most information. Knowing $\tau_1(S) = 1$ is not helpful for bounding eigenvalues of stochastic matrices since we know their largest eigenvalue in magnitude is 1. We present definitions and properties to illustrate when

$\tau_1(S) < 1$.

We begin with a review of several matrix definitions that are useful for this discussion. Recall, the definitions for irreducibility and primitivity of matrices, and regular (or aperiodic) matrices.

**Definition 3.5.7** (p 671 in (76)). *A square matrix $A$ is irreducible if there does not exist a permutation matrix $P$ such that*

$$P^T AP = \begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix}, \text{ where } X \text{ and } Z \text{ are square matrices.}$$

**Definition 3.5.8** (p 674 in (76)). *A square, nonnegative, irreducible matrix $A$ is primitive if $\lambda_1 > |\lambda|$, where $\lambda_1$ is the dominant eigenvalue of $A$ and $\lambda$ is any other eigenvalue of $A$.*

There is a special name for primitive matrices that are also stochastic.

**Definition 3.5.9** (p 694 in (76)). *A primitive stochastic matrix $S$ is called regular or aperiodic.*

One might think that investigation into the eigenvalues of a stochastic matrix $S$ could provide insight for when $\tau_1(S)$ might be less than 1. Unfortunately, this approach is not helpful. It follows from Theorem 3.5.3 that if $S$ has another eigenvalue with magnitude 1, then $\tau_1(S) = 1$. There are other cases when $\tau_1(S) = 1$ even if $S$ only has

one eigenvalue with magnitude 1. Consider

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

with eigenvalues $1, -0.6058, -0.7021 \pm 0.6383i$. The matrix $S$ is a primitive matrix, so we know that $S$ has a simple eigenvalue 1, and has only one eigenvalue with magnitude equal to one.

It is not clear how to discuss the value of the ergodicity coefficient using the concepts of irreducibility, primitivity, or regularity. Thus we present a definition for stochastic matrices that yields the property $\tau_1(S) < 1$.

**Definition 3.5.10** (§2 in (43), §3.2 in (95), p 82 in (98))**.** *A stochastic matrix $S$ is scrambling if $\tau_1(S) < 1$.*

There are several easy ways to identify scrambling matrices. First, notice that the set of scrambling matrices is a subset of the set of regular matrices (see Definition 3.5.9) (43; 98). Secondly, it follows from the explicit forms for $\tau_1(S)$ in Theorem 3.5.5 that a matrix cannot be scrambling if each column contains a single nonzero element. Thus permutation matrices are not scrambling matrices. Finally, if no two rows are orthogonal, then a stochastic matrix is scrambling (97, §4). This is another way of saying that given any two rows, there is at least one position where both have a positive element in the same column (43, §2).

One particular case of scrambling matrices is Markov matrices.

**Definition 3.5.11** (§3.2 in (95))**.** *A stochastic matrix $S$ is a Markov matrix if*

$$\max_j \left( \min_i s_{ij} \right) > 0.$$

Basically, a Markov matrix is a stochastic matrix with an entirely positive column. Clearly Markov matrices are also scrambling matrices, since all rows have a positive element in the same position.

We observe that both the Markov and scrambling properties are preserved by multiplication.

**Theorem 3.5.12** (Lemma 2 in (43), §4 in (97))**.** *If $S$ and $Q$ are stochastic matrices and either of them is scrambling, then $SQ$ and $QS$ are also scrambling.*

*Proof.* This follows from the submultiplicative property, $\tau_1(SQ) \leq \tau_1(S)\tau_1(Q)$, presented as item 2 of Theorem 3.5.3. □

**Corollary 3.5.13** (§3.2 in (95))**.** *If $S$ and $Q$ are stochastic matrices and either of them is a Markov matrix, then $SQ$ and $QS$ are also Markov matrices.*

### 3.5.2 Ergodicity coefficients of stochastic matrices in the $\infty$-norm

In this section we focus on the case $p = \infty$ from expression (3.2) and get

$$\tau_\infty(S) = \max_{\substack{\|z\|_\infty = 1 \\ z^T e = 0}} \left\| S^T z \right\|_\infty \tag{3.5}$$

where the maximum ranges over all $z \in \mathbb{R}^n$ (97, §2). This choice of norm is of particular interest because it is easy to compute.

Unlike $\tau_1(S)$ which takes values between 0 and 1, $\tau_\infty(S)$ has no such fixed upper bound. This is an example of a coefficient of ergodicity that does not satisfy the original definition (see Definition 3.4.1), however it is still an ergodicity coefficient based on Definition 3.4.5.

One of the easiest example of $\tau_\infty(S) > 1$ comes from the $n \times n$ matrix $S$, for $n$ even. Then the first $n/2$ rows of $S$ are the same, and the remaining $n/2$ rows are the same.

$$S = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \qquad \text{with } \tau_\infty(S) = \frac{n}{2}. \tag{3.6}$$

Tan presents an example for a $6 \times 6$ stochastic matrix with $\tau_\infty(S) > 1$ in (109, p 861), and Seneta has a specific example for a $62 \times 62$ stochastic matrix in (97, §3).

We now present results to give some insight into $\tau_\infty(S)$. Similar statements were proven for $\tau_1(S)$ in Section 3.5.1. We show that $\tau_\infty(S)$ is bounded, continuous, well-conditioned and submultiplicative. Additionally, $\tau_\infty(S)$ is proper and an upper bound on subdominant eigenvalues of $S$.

**Theorem 3.5.14** ((7) & (8) in (109))**.** *For stochastic matrices $S$, $S_1$ and $S_2$*

*1.* $0 \leq \tau_\infty(S) \leq \|S\|_1$

2. $|\tau_\infty(S_1) - \tau_\infty(S_2)| \leq \|S_1 - S_2\|_1$

3. $\tau_\infty(S) = 0$ *if and only if* $\text{rank}(S) = 1$

4. $|\lambda| \leq \tau_\infty(S)$ *for any real eigenvalue* $\lambda \neq 1$ *of* $S$

5. $\tau_\infty(S_1 S_2) \leq \tau_\infty(S_1)\ \tau_\infty(S_2)$

*Proof.* The proofs for each of these items are analogous to proofs presented in Section 3.5.1 for $\tau_1(S)$. Specifically, the proof of items 1 and 2 is analogous to the proof of Theorem 3.5.1, the proof of item 3 is analogous to the proof of Theorem 3.5.2, and the proof of items 4 and 5 is analogous to the proof of Theorem 3.5.3. □

We present the results of Tan (109) in which he describes an explicit form for $\tau_\infty(S)$.

**Theorem 3.5.15** (§2 in (109))**.** *For stochastic* $S \in \mathbb{R}^{n \times n}$

$$\tau_\infty(S) = \max_s \phi(s)$$

*where* $s$ *is some column of* $S$ *and*

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

*assuming the elements of* $s$ *have been permuted so that* $s_1 \geq \cdots \geq s_n \geq 0$*, where* $s_i$ *is the* $i^{th}$ *element of column* $s$ *of stochastic matrix* $S$*.*

*Proof.* First, write for arbitrary column $s$ of $S$

$$\tau_\infty(S) = \max_{\substack{\|z\|_\infty=1 \\ z^T e=0}} \left\|S^T z\right\|_\infty = \max_s \max_{\substack{\|z\|_\infty=1 \\ z^T e=0}} \left|z^T s\right|$$

Also, define $\phi(s)$ for any nonnegative $s$ as

$$\phi(s) = \min_k \sum_{i=1}^{n} \left|s_i - s_k\right|, \quad \text{for } k = 1, \ldots, n.$$

We now present the proof in three parts:

1. $\left|z^T s\right| \leq \|z\|_\infty \phi(s)$ where $z^T e = 0$,

2. When the elements of $s$ have been permuted so that $s_1 \geq \cdots \geq s_n \geq 0$ then

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

3. $\phi(s) = \max_{\substack{\|z\|_\infty=1 \\ z^T e=0}} \left|z^T s\right|$.

For all parts of the proof, notice that $s$ is a nonnegative vector because it is some column of the stochastic matrix $S$.

1. For any $k$, $z^T s = \sum_{i=1}^{n} z_i \left(s_i - s_k\right)$ because $z^T e = 0$.

Since $k$ is arbitrary,

$$\left|z^T s\right| = \left|\sum_{i=1}^{n} z_i(s_i - s_k)\right| = \min_k \left|\sum_{i=1}^{n} z_i(s_i - s_k)\right|.$$

Finally, we obtain

$$\begin{aligned}
\left|z^T s\right| &= \min_k \left|\sum_{i=1}^{n} z_i(s_i - s_k)\right| \\
&\leq \min_k \sum_{i=1}^{n} |z_i(s_i - s_k)| \\
&\leq \left(\max_i |z_i|\right) \min_k \sum_{i=1}^{n} |s_i - s_k| \\
&= \|z\|_\infty \phi(s)
\end{aligned}$$

2. Assume, without loss of generality that $S$ has been permuted so that the elements of a specific

column $s$ are ordered $s_1 \geq \cdots \geq s_n \geq 0$. We want to find $z \in \mathbb{R}^n$ so that $|z^T s| = \phi(s)$. This requires $\|z\|_\infty = 1$ and $z^T e = 0$. Then if $n$ is even, choose $z = \pm \begin{pmatrix} e & -e \end{pmatrix}$ so that the first $n/2$ elements are $\pm 1$ and the remaining $n/2$ elements are $\mp 1$. If $n$ is odd, choose $z = \pm \begin{pmatrix} e & 0 & -e \end{pmatrix}$ with the first $n/2$ elements as $\pm 1$, the middle element is 0, and the last $n/2$ elements are $\mp 1$.

In order for $|z^T s| = \sum_{i=1}^{n} |s_i - s_k|$, choose $k = \frac{n}{2}$ or $k = \frac{n}{2} + 1$ when $n$ is even and choose $k = \frac{1}{2}(n+1)$ if $n$ is odd.

3. From part 2, the vector $z$ that achieves the maximum of $\left|z^T s\right|$ under the restrictions $\|z\|_\infty = 1$ and

$z^T e = 0$ is of the form $z = \pm \begin{pmatrix} e & -e \end{pmatrix}$ or $z = \pm \begin{pmatrix} e & 0 & -e \end{pmatrix}$, depending on whether the matrix has an even or odd number of rows, respectively. Then $\phi(s) = \max_{\substack{\|z\|_\infty = 1 \\ z^T e = 0}} \left| z^T s \right|$

We conclude

$$\tau_\infty(S) = \max_{\substack{\|z\|_\infty = 1 \\ z^T e = 0}} \left\| S^T z \right\|_\infty = \max_s \max_{\substack{\|z\|_\infty = 1 \\ z^T e = 0}} \left| z^T s \right| = \max_s \phi(s)$$

for arbitrary column $s$ of matrix $S$. □

Based on the explicit form from Theorem 3.5.15, Tan also presents bounds on $\tau_\infty(S)$.

**Theorem 3.5.16** (Proposition 4 in (109))**.** *For $n \times n$ stochastic matrix $S$,*

$$\alpha(S) \leq \tau_\infty(S) \leq \tfrac{n}{2}\, \alpha(S), \qquad \text{if } n \text{ is even}$$

$$\text{and}$$

$$\alpha(S) \leq \tau_\infty(S) \leq \tfrac{n-1}{2}\, \alpha(S), \qquad \text{if } n \text{ is odd.}$$

*Recall from Section 3.4 that $\alpha(S) = \max_k \max_{ij} |s_{ik} - s_{jk}|$.*

*Proof.* For all $n$, the lower bound is straightforward when $\tau_\infty(S)$ is written appropriately. With the explicit form from Theorem 3.5.15 and the formula $\phi(s) = \min_k \sum_{i=1}^n |s_i - s_k|$, for $k = 1, \ldots, n$,

$$\tau_\infty(S) = \max_m \phi(s_m) = \max_m \min_l \sum_{i=1}^n |s_{im} - s_{lm}|$$

where $s_m$ denotes the $m^{th}$ column of $S$, and $s_{jm}$ denotes the $j^{th}$ element of $s_m$. Then

$$\alpha(S) = \max_k \min_{ij} |s_{ik} - s_{jk}| \leq \max_m \min_l \sum_{i=1}^{n} |s_{im} - s_{lm}| = \tau_\infty(S).$$

Without loss of generality, assume the rows of $S$ have been permuted, such that the elements have been ordered, and $s_{1m} \geq s_{2m} \geq \cdots \geq s_{nm} \geq 0$. Then from the explicit forms for $\tau_\infty(S)$ and $\phi(\cdot)$ in Theorem 3.5.15 we prove the upper bound for even values of $n$:

$$\begin{aligned}
\tau_\infty(S) &= \max_m \phi(s_m) \\
&= \max_m \left[ (s_{1m} - s_{nm}) + (s_{2m} - s_{n-1,m}) + \cdots + \left( s_{n/2,m} - s_{(n+2)/2,m} \right) \right] \\
&\leq \max_m \left[ (s_{1m} - s_{nm}) + (s_{1m} - s_{nm}) + \cdots + (s_{1m} - s_{nm}) \right] \\
&\leq \frac{n}{2} \alpha(S).
\end{aligned}$$

When $n$ is odd we obtain

$$\begin{aligned}
\tau_\infty(S) &= \max_m \phi(s_m) \\
&= \max_m \left[ (s_{1m} - s_{nm}) + (s_{2m} - s_{n-1,m}) + \cdots + \left( s_{(n-1)/2,m} - s_{(n+3)/2,m} \right) \right] \\
&\leq \max_m \left[ (s_{1m} - s_{nm}) + (s_{1m} - s_{nm}) + \cdots + (s_{1m} - s_{nm}) \right] \\
&\leq \frac{n-1}{2} \alpha(S).
\end{aligned}$$

□

Another common bound on $\tau_\infty(S)$ follows from the previous relationship between

$\tau_\infty(S)$ and $\alpha(S)$. Rhodius (85) and Lešanovský (72) credit Tan (109) with this result.

**Corollary 3.5.17.** *For $n \times n$ stochastic matrix $S$,*

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

*Proof.* From Theorems 3.4.4 and 3.5.1 we know that $\alpha(S) \leq \tau_1(S) \leq 1$. Then from Theorem 3.5.16 when $n$ is even, we have

$$\tau_\infty(S) \leq \frac{n}{2}\alpha(S) \leq \frac{n}{2}$$

and for odd values of $n$

$$\tau_\infty(S) \leq \frac{n-1}{2}\alpha(S) \leq \frac{n-1}{2}.$$

□

It is easy to see that this is a tight bound. Recall the $n \times n$ matrix for even $n$, presented earlier in this section:

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

### 3.5.3 Ergodicity coefficients of stochastic matrices in the $p$-norm

We briefly consider the general ergodicity coefficient for a stochastic matrix in the $p$-norm.

$$\tau_p(S) = \max_{\substack{\|z\|_p=1 \\ z^T e=0}} \left\|S^T z\right\|_p \tag{3.7}$$

where the maximum ranges over all $z \in \mathbb{R}^n$ (97, §2).

Like $\tau_1(S)$ and $\tau_\infty(S)$, we have several properties and relationships for $\tau_p(S)$. Although previously listed only for $p = 1$ and $p = \infty$, we now present several properties and relationships that hold for all positive integer values of $p$, including continuity, submultiplicativity, proper, upper bounds, and eigenvalue bounds.

**Theorem 3.5.18** (§1 in (110), (85))**.** *For stochastic matrices $S$, $S_1$ and $S_2$ and positive integer values of $p$*

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

The proofs for each of these items are analogous to proofs presented in Section 3.5.1 for $\tau_1(S)$.

In 1979, Seneta (97) investigated the values of $\tau_p(S)$ for various dimensions of $S$. Specifically, for $S \in \mathbb{R}^{2\times 2}$, for any $p$-norm, $\tau_p(S) = |\lambda_2|$ where $\lambda_2$ is a subdominant eigenvalue of $S$.

**Theorem 3.5.19** (p 585 in (97))**.** *For $S \in \mathbb{R}^{2\times 2}$ with eigenvalues $|\lambda_1| \geq |\lambda_2|$. Then $\tau_p(S) = |\lambda_2|$ for all $p$-norms.*

*Proof.* Let

$$S = \begin{pmatrix} s_1 & 1-s_1 \\ s_2 & 1-s_2 \end{pmatrix}, \qquad \text{for } 0 \leq s_1, s_2 \leq 1$$

have eigenvalues 1 and $s_1 - s_2$. Clearly $s_1 - s_2$ is less than one, unless $S = I$, in which case both eigenvalues of $S$ are 1. Since $\tau_p(S)$ requires $z^T e = 0$, $z$ must take the form

$$z = \alpha \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \text{for } \alpha \neq 0.$$

The value of $\alpha$ is chosen to meet the second restriction on $\tau_p(S)$, that $\|z\|_p = 1$ for $p = 1, 2, \ldots$. From

$$1 = \|z\|_p = |\alpha| \left\| \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\|_p = |\alpha|\, 2^{1/p}$$

we obtain $|\alpha| = 2^{-1/p}$. When $p = \infty$, $|\alpha| = 1$.

Then

$$\begin{aligned}
\tau_p(S) &= \max_{\substack{\|z\|_p=1 \\ z^T e=0}} \left\|S^T z\right\|_p \\
&= \left\| \begin{pmatrix} s_1 & s_2 \\ 1-s_1 & 1-s_2 \end{pmatrix} \begin{pmatrix} \alpha \\ -\alpha \end{pmatrix} \right\|_p, \text{ with } |\alpha| = 2^{-1/p} \\
&= \left\| \alpha \left(s_1 - s_2\right) \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\|_p \\
&= |\alpha||s_1 - s_2| \, 2^{1/p} \\
&= |s_1 - s_2|
\end{aligned}$$

□

**The maximum value of $\tau_p(S)$**

In addition to the basic properties in Theorem 3.5.18, some work has been done on the maximal value of $\tau_p(S)$. In 1988 Rhodius (85, p 145) determined maximal values of the ergodicity coefficient of stochastic matrices, for any $p$-norm. The first theorem in this section relates the ergodicity coefficient to the set of extreme points of $n \times n$ stochastic matrices. The remaining corollary and theorems in this section develop computable bounds for the ergodicity coefficient by exploiting the structure of the extreme points of the set of $n \times n$ stochastic matrices. Before we can present any of these results, we prove some properties of the set of $n \times n$ stochastic matrices, $\mathcal{S}_n$, and the set $\mathcal{H} = \left\{x \mid x \in \mathbb{R}^n, x^T e = 0 \text{ and } \|x\|_p = 1\right\}$.

**Lemma 3.5.20.** *The set $\mathcal{S}_n$ of $n \times n$ stochastic matrices is convex and compact. Its extreme points are stochastic matrices that have a single one in each row.*

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

$$(\alpha S_1 + (1-\alpha)S_2)e = \alpha S_1 e + (1-\alpha)S_2 e = \alpha e + (1-\alpha)e = e.$$

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

Because of its $n-1$ zero elements, a canonical vector $e_j$ cannot be expressed as convex combination of other stochastic vectors. Therefore stochastic matrices whose rows are canonical vectors are extreme points of $\mathcal{S}_n$.

The set $\mathcal{S}_n$ is bounded because for any stochastic matrix $S$ we have

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

where the first inequality follows from (49, (6.19)), and the second from the fact that the elements of a stochastic matrix are bounded above by 1.

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 $k \to \infty$. Because this implies $(S_k)_{ij} \to Z_{ij}$ and the elements $(S_k)_{ij}$ are in the closed interval $[0,1]$, $Z_{ij}$ must also be in $[0,1]$. We still need to show that the elements in each row of $Z$ sum to 1. For any $\varepsilon > 0$ there exists a $k$ so that

$$\varepsilon > \|S_k - Z\|_p \geq \frac{\|(S_k - Z)e\|_p}{\|e\|_p} = n^{-p}\|e - Ze\|_p.$$

Therefore $Ze = e$. We have shown that the limit of any converging sequence of stochastic matrices is also a stochastic matrix. Therefore the set $\mathcal{S}_n$ is closed. Since a closed and bounded subset of a finite dimensional vector space is compact, $\mathcal{S}_n$ must be compact. □

We also prove the set of real vectors $x$, with $x^T e = 0$ and $\|x\|_p = 1$ is compact. This result will be necessary in the proof of Theorem 3.5.22.

**Lemma 3.5.21.** *The set* $\mathcal{H} = \left\{ x \mid x \in \mathbb{R}^n, x^T e = 0 \text{ and } \|x\|_p = 1 \right\}$ *is compact.*

*Proof.* We will prove $\mathcal{H}$ is closed and bounded, and since a closed and bounded subset of a finite dimensional vector space is compact, we will have proven $\mathcal{H}$ is compact.

Clearly $\mathcal{H}$ is bounded since $\|x\|_p = 1$.

To prove that $\mathcal{H}$ is closed, let $x_k$ be a sequence of vectors in $\mathcal{H}$ that converge to $z$, so $\|x_k - z\|_p \to 0$ as $k \to \infty$. For any $\varepsilon > 0$ there exists a $k$ such that

$$\varepsilon > \|x_k - z\|_p = \left\| (x_k - z)^T \right\|_q \geq \frac{\left\| (x_k - z)^T e \right\|_q}{\|e\|_q} = \frac{\left\| -z^T e \right\|_q}{\|e\|_q} = n^{-q} \left\| z^T e \right\|_q$$

if $\frac{1}{p} + \frac{1}{q} = 1$. Therefore $z^T e = 0$, which means $z \in \mathcal{H}$. Thus the set $\mathcal{H}$ is closed because the limit of any converging sequence of vectors from $\mathcal{H}$ is also a vector in $\mathcal{H}$. □

**Theorem 3.5.22** (Theorem 1(a) in (85))**.** *Let* $\mathcal{S}_n$ *be the set of all* $n \times n$ *stochastic matrices, and* $S \in \mathcal{S}_n$*. For any p-norm,*

$$\max_{S \in \mathcal{S}_n} \tau_p(S) = \max_{\substack{\|z\|_p = 1 \\ z^T e = 0}} \max_{Q \in \operatorname{Extr} \mathcal{S}_n} \left\| Q^T z \right\|_p . \tag{3.8}$$

*where* $\operatorname{Extr} \mathcal{S}_n$ *is the set of all extreme points of* $\mathcal{S}_n$.

*Proof.* For a stochastic matrix $S \in \mathbb{R}^{n \times n}$ and $z \in \mathbb{R}^n$, the $p$-norm ergodicity coefficient is defined as

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

We are interested in finding the maximum $\tau_p(S)$ for all $n \times n$ stochastic matrices, so

$$\max_{S \in \mathcal{S}_n} \tau_p(S) = \max_{S \in \mathcal{S}_n} \max_{\substack{\|z\|_p = 1 \\ z^T e = 0}} \left\| S^T z \right\|_p = \max_{\substack{\|z\|_p = 1 \\ z^T e = 0}} \max_{S \in \mathcal{S}_n} \left\| S^T z \right\|_p .$$

The order of the maxima can be switched since $\mathcal{S}_n$ and $\mathcal{H}$ are both compact sets, for $\mathcal{H} = \left\{ z \mid z \in \mathbb{R}^n, z^T e = 0 \text{ and } \|z\|_p = 1 \right\}$, as proven in Lemmas 3.5.20 and 3.5.21, and because $\| \cdot \|$ is continuous. The extreme value theorem states that a continuous real-valued function on a nonempty compact space is bounded and attains its maximum (108, p 807).

Now we examine a function on the set of all $z \in \mathbb{R}^n$ with $\|z\|_p = 1$ and $z^T e = 0$. Specifically, the function $\|S^T z\|_p$ is convex. On a compact, convex set the maximum of a convex function is attained at an extreme point (50, p 535). Thus, it follows for $z \in \mathbb{R}^n$ with $\|z\|_p = 1$ and $z^T e = 0$ that

$$\max_{S \in \mathcal{S}_n} \left\| S^T z \right\|_p = \max_{Q \in \operatorname{Extr} \mathcal{S}_n} \left\| Q^T z \right\|_p .$$

and we obtain (3.8)

$$\max_{S \in \mathcal{S}_n} \tau_p(S) = \max_{\substack{\|z\|_p = 1 \\ z^T e = 0}} \max_{Q \in \operatorname{Extr} \mathcal{S}_n} \left\| Q^T z \right\|_p .$$

□

As a corollary to the above theorem, we exploit the form of matrices $Q \in \operatorname{Extr} \mathcal{S}_n$ to write $Q^T z$, as well as

$$\max_{S \in \mathcal{S}_n} \tau_p(S) = \max_{\|z\|_p = 1, z^T e = 0} \max_{Q \in \operatorname{Extr} \mathcal{S}_n} \left\| Q^T z \right\|_p$$

in an alternate form. We proceed with an example.

A stochastic matrix $Q \in \mathcal{S}_n$ is in $\operatorname{Extr} \mathcal{S}_n$ if $Q$ has a single 1 in each of its rows. It follows that for any vector $z \in \mathbb{R}^n$,

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

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

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

Then $D_1 = \{1, 2\}$ and $D_3 = \{3\}$ because $q_{11} = q_{21} = q_{33} = 1$, and

$$Q^T z = \begin{pmatrix} z_1 + z_2 \\ 0 \\ z_3 \end{pmatrix}.$$

**Corollary 3.5.23** (Theorem 1(b) in (85))**.** *For $Q \in \operatorname{Extr} \mathcal{S}_n$ and for $z \in \mathbb{R}^n$ with $\|z\|_p = 1$ and $z^T e = 0$,*

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

*where $q_{ij} = 1 \iff i \in D_j$ for $i, j = 1, \ldots, n$, and*

$$\max_{S \in \mathcal{S}_n} \tau_p(S) = \max_{\substack{\|z\|_p = 1 \\ z^T e = 0}} \max_{Q \in \operatorname{Extr} \mathcal{S}_n} \left\| \begin{pmatrix} \sum_{i \in D_1} z_i \\ \vdots \\ \sum_{i \in D_n} z_i \end{pmatrix} \right\|_p .$$

Substituting the expression for $Q^T z$ from Corollary 3.5.23 into equation (3.8) from Theorem 3.5.22 yields

$$\max_{S \in \mathcal{S}_n} \tau_p(S) = \max_{\substack{\|z\|_p = 1 \\ z^T e = 0}} \max_{Q \in \operatorname{Extr} \mathcal{S}_n} \left\| Q^T z \right\|_p = \max_{\substack{\|z\|_p = 1 \\ z^T e = 0}} \max_{Q \in \operatorname{Extr} \mathcal{S}_n} \left\| \begin{pmatrix} \sum_{i \in D_1} z_i \\ \vdots \\ \sum_{i \in D_n} z_i \end{pmatrix} \right\|_p .$$

We now present a second theorem in which a specific function achieves its maximum. We assume for $z \in \mathbb{R}^n$ with $\|z\|_p = 1$ and $z^T e = 0$ that the elements of $z$ are ordered, so that $z_1, \ldots, z_l > 0$ and $z_{l+1}, \ldots, z_n \leq 0$ for some $l = 1, \ldots, n-1$. Then from Corollary 3.5.23, for $Q \in \operatorname{Extr} \mathcal{S}_n$

$$\left\| Q^T z \right\|_p^p = \sum_{j=1}^{n} \left| \sum_{i \in D_j} z_i \right|^p \leq \sum_{j=1}^{n} \sum_{i \in D_j} |z_i|^p = \|z\|_1^p = \left( \|z_{1:l}\|_1 + \|z_{l+1:n}\|_1 \right)^p .$$

From $0 = z^T e = z_{1:l}^T e + z_{l+1:n}^T e$ and the assumption that the first $l$ components are $z$ are nonnegative, it follows that

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

Hence

$$\left\|Q^T z\right\|_p^p \leq 2 \left(\|z_{1:l}\|_1\right)^p$$

and taking the $p^{th}$ root of both sides, we have

$$\left\|Q^T z\right\|_p \leq 2^{1/p} \|z_{1:l}\|_1 = 2^{1/p} \sum_{i=1}^{l} z_i.$$

This results leads us to the next theorem.

**Theorem 3.5.24** (Theorem 2 in (85))**.** *For $1 \leq l \leq n$, define the function $f(\cdot, l)$ as*

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

*where $\|z\|_p = 1$, $z^T e = 0$, $z_i > 0$ for $i = 1, \ldots, l$ and $z_i \leq 0$ for $i = l+1, \ldots, n$. Then $f(z, l)$ achieves a maximum when $z_1 = \cdots = z_l$ and $z_{l+1} = \cdots = z_n$.*

*Proof.* Choose $x \in \mathbb{R}^n$ to achieve a maximum of $f(z, l)$, where $\|x\|_p = 1$, $x^T e = 0$, $x_1, \ldots, x_l > 0$, and $x_{l+1}, \ldots, x_n \leq 0$,

$$\max_{\substack{\|z\|_p=1 \\ z^T e=0 \\ z_1,\ldots,z_l \geq 0 \\ z_{l+1},\ldots,z_n \leq 0}} f(z, l) = f(x, l).$$

Define $y \in \mathbb{R}^n$ such that

$$y_1 = \cdots = y_l = \tfrac{1}{l}\left(x_1 + \cdots + x_l\right)$$

and

$$y_{l+1} = \cdots = y_n = \tfrac{1}{n-l}\left(x_{l+1} + \cdots + x_n\right).$$

Then $y^T e = 0$ and $f(y,l) = f(x,l)$.

We now turn our attention to $\|y\|_p$ using the above formulas for $y_i$ in terms of $x_i$. Then

$$\|y\|_p^p = \sum_{i=1}^{l} |y_i|^p + \sum_{i=l+1}^{n} |y_i|^p = l \left| \frac{1}{l} \sum_{i=1}^{l} x_i \right|^p + (n-l) \left| \frac{1}{n-l} \sum_{i=l+1}^{n} x_i \right|^p .$$

From $x_i > 0$ for $i = 1, \ldots, l$, we have $\left|\frac{1}{l}\sum_{i=1}^{l} x_i\right| = \frac{1}{l}\sum_{i=1}^{l} x_i$. Then the Hölder inequality can be applied to $\sum_{i=1}^{l} x_i/l$, so that

$$\sum_{i=1}^{l} \frac{x_i}{l} \leq \left( \sum_{i=1}^{l} x_i^p \right)^{1/p} \left( \sum_{i=1}^{l} \frac{1}{l^q} \right)^{1/q},$$

with the requirement that $1/q + 1/p = 1$ for $1 \leq p, q < \infty$. Solving for $q = \frac{p}{p-1}$ and raising both sides of the inequality to the $p^{th}$ power, yields

$$\left( \sum_{i=1}^{l} \frac{x_i}{l} \right)^p \leq \sum_{i=1}^{l} x_i^p \left( \sum_{i=1}^{l} \frac{1}{l^q} \right)^{p-1}.$$

Now, from the simplification $\sum_{i=1}^{l} \frac{1}{l^q} = l\left(\frac{1}{l^q}\right) = l^{1-q}$, we write

$$\begin{aligned}\left(\sum_{i=1}^{l} \frac{x_i}{l}\right)^p &\leq \sum_{i=1}^{l} x_i^p \left(l^{1-q}\right)^{p-1} \\ &= \sum_{i=1}^{l} x_i^p \left(l^{-1/(p-1)}\right)^{p-1} \\ &= \frac{1}{l}\sum_{i=1}^{l} x_i^p\end{aligned}$$

Through a similar argument we obtain the second relation needed for $\|y\|_p^p$

$$\left|\frac{1}{n-l}\sum_{i=l+1}^{n} x_i\right|^p \leq \frac{1}{n-l}\left(|x_{l+1}|^p + \cdots + |x_n|^p\right).$$

Returning to $\|y\|_p^p$,

$$\begin{aligned}\|y\|_p^p &= l\left|\frac{1}{l}\sum_{i=1}^{l} x_i\right|^p + (n-l)\left|\frac{1}{n-l}\sum_{i=l+1}^{n} x_i\right|^p \\ &\leq l\left(\frac{1}{l}\sum_{i=1}^{l} x_i^p\right) + (n-l)\left(\frac{1}{n-l}\sum_{i=l+1}^{n} |x_i|^p\right) \\ &= \sum_{i=1}^{l} x_i^p + \sum_{i=l+1}^{n} |x_i|^p \\ &= \|x\|_p^p = 1.\end{aligned}$$

Since $f(x,l) = f(y,l)$ and $\|y\|_p \leq 1$, we obtain

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

However, since by assumption $x$ achieves the maximum of $f(z,l)$, then $f\left(y/\|y\|_p, l\right) \leq f(x,l)$, so

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

□

We now combine the previous two theorems and corollary to obtain a value for $\max_{S\in\mathcal{S}_n} \tau_p(S)$.

**Theorem 3.5.25** (Theorem 3 in (85))**.** *For stochastic matrices $S \in \mathbb{R}^{n\times n}$ and positive integer $p$*

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

*where the maximum ranges over all $n \times n$ stochastic matrices $S$.*

*Proof.* Combining the results from Theorem 3.5.22, Corollary 3.5.23, and Theorem

3.5.24

$$\begin{aligned}
\max_{S\in\mathcal{S}_n}\tau_p(S) &= \max_{\substack{\|z\|_p=1\\ z^Te=0}}\ \max_{Q\in\operatorname{Extr}\mathcal{S}_n}\left\|Q^Tz\right\|_p, \\
&= \max_{\substack{\|z\|_p=1\\ z^Te=0}}\ \max_{Q\in\operatorname{Extr}\mathcal{S}_n}\left\|\begin{pmatrix}\sum_{i\in D_1}z_i\\ \vdots\\ \sum_{i\in D_n}z_i\end{pmatrix}\right\|_p, \\
&= \max_{l=1,\ldots,n-1}\ \max_{\substack{\|z\|_p=1\\ z^Te=0\\ z_1,\ldots,z_l\geq 0\\ z_{l+1},\ldots,z_n\leq 0}} f(z,l).
\end{aligned}$$

Recall from Theorem 3.5.24, that $f(z,l)$ is defined as $f(z,l) = 2^{1/p}\left(z_1+\cdots+z_l\right)$, where $\|z\|_p = 1$, $z^Te = 0$, $z_i > 0$ for $i = 1,\ldots,l$ and $z_i \leq 0$ for $i = l+1,\ldots,n$. Theorem 3.5.24 says the function $f(z,l)$ achieves a maximum when $z_1 = \cdots = z_l$ and $z_{l+1} = \cdots = z_n$, so

$$\max_{\substack{\|z\|_p=1\\ z^Te=0\\ z_1,\ldots,z_l\geq 0\\ z_{l+1},\ldots,z_n\leq 0}} f(z,l) = \max_{\substack{\|z\|_p=1\\ z^Te=0\\ z_1,\ldots,z_l\geq 0\\ z_{l+1},\ldots,z_n\leq 0}} 2^{1/p}\left(z_1+\cdots+z_l\right) = 2^{1/p}\cdot l\cdot z_1.$$

Then

$$\max_{S\in\mathcal{S}_n}\tau_p(S) = 2^{1/p}\max_{l=1,\ldots,n-1} lz_1.$$

It remains to solve for $lz_1$ so that $\max_{S\in\mathcal{S}_n}\tau_p(S)$ does not depend on $l$ or $z$. From $z^Te = 0$, we obtain $lz_1 + (n-l)z_n = 0$, so

$$z_n = \frac{-lz_1}{n-l}.$$

Substituting this value for $z_n$ into the equation $\|z\|_p^p = 1$ gives

$$lz_1^p + (n-l)\left(\frac{lz_1}{n-l}\right)^p = 1.$$

Solving for $lz_1$ gives

$$lz_1 = \left(\frac{1}{l^{1-p} + (n-l)^{1-p}}\right)^{1/p},$$

and substituting the expression for $lz_1$ into the above expression for $\max f(z,l)$ yields

$$\max_{\substack{\|z\|_p=1 \\ z^T e=0 \\ z_1,\ldots,z_l \geq 0 \\ z_{l+1},\ldots,z_n \leq 0}} f(z,l) = \left(\frac{2}{l^{1-p} + (n-l)^{1-p}}\right)^{1/p}$$

for $l = 1, \ldots, n-1$. The values of $l$ that maximize $2/\left(l^{1-p} + (n-l)^{1-p}\right)$ are $l = n/2$ when $n$ is even, and $l = (n \pm 1)/2$ if $n$ is odd. □

### 3.5.4 Connections between $\tau_1(S)$ and $\tau_\infty(S)$

In this section we present relations between $\tau_1(S)$ and $\tau_\infty(S)$. The situation when $\tau_\infty(S) \leq \tau_1(S)$ has been investigated, specifically because under this condition, $\tau_\infty(S)$ is a tighter bound for subdominant eigenvalues.

In section 3.5.2, we presented a relation between $\tau_\infty(S)$ and $\alpha(S)$ from Tan in 1982 (109, Proposition 4). Recall the result from Theorem 3.5.16 for $n \times n$ stochastic matrix $S$ with $n \geq 2$

$$\alpha(S) \leq \tau_\infty(S) \leq \tfrac{n}{2}\,\alpha(S), \qquad \text{if } n \text{ is even,}$$

and

$$\alpha(S) \leq \tau_\infty(S) \leq \tfrac{n-1}{2}\,\alpha(S), \qquad \text{if } n \text{ is odd,}$$

where $\alpha(S) = \max_k \max_{ij} |s_{ik} - s_{jk}|$ (see Section 3.4). In Theorem 3.4.4 we proved $\alpha(S) \leq \tau_1(S)$ for all stochastic matrices $S$. We combine these two results to prove the following relation between $\tau_1(S)$ and $\tau_\infty(S)$.

**Theorem 3.5.26.** *For stochastic matrix $S \in \mathbb{R}^{n\times n}$ with $n \geq 2$*

$$\frac{2}{n}\tau_1(S) \leq \tau_\infty(S) \leq \frac{n}{2}\tau_1(S), \qquad \textit{for } n \textit{ even,}$$

*and*

$$\frac{2}{n}\tau_1(S) \leq \tau_\infty(S) \leq \frac{n-1}{2}\tau_1(S), \qquad \textit{for } n \textit{ odd.}$$

*Proof.* We start by proving the first inequality $\frac{2}{n}\tau_1(S) \leq \tau_\infty(S)$, which holds for both even and odd values of $n$. According to Theorem 3.5.5, we can choose $i \neq j$ so $\tau_1(S) = \frac{1}{2}\left\|S^T(e_i - e_j)\right\|_1$. This together with $\|x\|_1 \leq n\|x\|_\infty$ implies

$$2\tau_1(S) = \left\|S^T(e_i - e_j)\right\|_1 \leq n\left\|S^T(e_i - e_j)\right\|_\infty.$$

Since $y = e_i - e_j$ satisfies $y^T e = 0$ and $\|y\|_\infty = 1$ we obtain

$$2\tau_1(S) \leq n\left\|S^T y\right\|_\infty \leq n\,\tau_\infty(S).$$

We now show that, the second inequality follows directly from Theorems 3.4.4 and 3.5.16. Theorem 3.4.4 implies $\alpha(S) \leq \tau_1(S)$. Theorem 3.5.16 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 $n$. Combining these two relations yields:

$$\tau_\infty(S) \leq \tfrac{n}{2}\,\tau_1(S), \qquad \text{if } n \text{ is even,}$$

and

$$\tau_\infty(S) \leq \tfrac{n-1}{2}\,\tau_1(S), \qquad \text{if } n \text{ is odd.}$$

□

We next present a special case where $\tau_\infty(S) = \frac{n}{2}\tau_1(S)$

**Theorem 3.5.27.** *For every even $n \geq 2$ there is a stochastic matrix $S$ for which $\tau_\infty(S) = \frac{n}{2}\tau_1(S)$.*

*Proof.* The matrix $S$ from expression (3.6),

$$S = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

has the first $n/2$ rows equal to $e_1^T$ and the remaining $n/2$ rows equal to $e_n^T$, where $n$ is even, and $\tau_\infty(S) = \frac{n}{2}$.

The explicit form for $\tau_1(S)$ in Theorem 3.5.5 implies

$$\tau_1(S) = \frac{1}{2}\left\|S^T\left(e_{n/2} - e_{n/2+1}\right)\right\|_1 = 1.$$

So

$$\tau_\infty(S) = \frac{n}{2} = \frac{n}{2}\tau_1(S).$$

□

For the special cases of $2 \times 2$ and $3 \times 3$ stochastic matrices, we show $\tau_1(S) = \tau_\infty(S)$.

**Theorem 3.5.28** (§3.3 in (97))**.** *For stochastic $S \in \mathbb{R}^{2\times 2}$ and $S \in \mathbb{R}^{3\times 3}$,*

$$\tau_1(S) = \tau_\infty(S).$$

*Proof.* For $n = 2$ this result follows from Theorem 3.5.19, since $\tau_p(S) = |\lambda_2|$, when $S$ has two eigenvalues $|\lambda_1| \geq |\lambda_2|$.

We now proceed with the case where $n = 3$. Theorem 3.5.5 implies $\tau_1(S) = \frac{1}{2}\max_{1\leq i,j\leq 3} \left\|S^T(e_i - e_j)\right\|_1$. Theorem 3.5.15 shows that

$$\tau_\infty(S) = \max_{1\leq i,j\leq 3} \left\|S^T(e_i - e_j)\right\|_\infty.$$

In general, $S^T(e_i - e_j)$ takes the form $\begin{pmatrix} \alpha & \beta & -\alpha-\beta \end{pmatrix}^T$. We distinguish two cases:

1. $\alpha$ and $\beta$ have the same sign.

2. $\alpha$ and $\beta$ have different signs.

When $\alpha$ and $\beta$ have the same sign, then $|-\alpha-\beta| = |\alpha|+|\beta|$ and $\max\{|\alpha|, |\beta|, |-\alpha-\beta|\} = |\alpha| + |\beta|$. Hence,

$$\tau_1(S) = \frac{1}{2}\left\|S^T(e_i - e_j)\right\|_1 = \frac{1}{2}(|\alpha| + |\beta| + |-\alpha-\beta|) = |\alpha| + |\beta|,$$

and

$$\tau_\infty(S) = \max\{|\alpha|, |\beta|, |-\alpha-\beta|\} = |\alpha| + |\beta|.$$

So, $\tau_1(S) = \tau_\infty(S)$.

When $\alpha$ and $\beta$ have different signs, assume without loss of generality that $|\alpha| \geq |\beta|$. Then $|-\alpha-\beta| = |\alpha| - |\beta|$ and $\max\{|\alpha|, |\beta|, |-\alpha-\beta|\} = |\alpha|$. Hence,

$$\tau_1(S) = \frac{1}{2}\left\|S^T(e_i - e_j)\right\|_1 = \frac{1}{2}(|\alpha| + |\beta| + |-\alpha-\beta|) = |\alpha|,$$

and

$$\tau_\infty(S) = \max\{|\alpha|, |\beta|, |-\alpha-\beta|\} = |\alpha|.$$

So, $\tau_1(S) = \tau_\infty(S)$. □

## 3.6 Ergodicity coefficients for general matrices

In 1984 Seneta (100) extended the coefficient of ergodicity from stochastic to general matrices. Since different general matrices do not have a common dominant eigenvalue and eigenvector, the definition and notation has an added vector. We present a slightly altered version of Seneta's notation in order to be consistent with previous notation for stochastic matrices.

**Definition 3.6.1** ((1) in (100)). *Let $A \in \mathbb{C}^{m\times n}$ and $w \in \mathbb{C}^m$. Then for integer $1 \leq p \leq \infty$*

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

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

This is similar to notation for stochastic matrices. Recall (3.2) which gave us

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

where the maximum ranges over $z \in \mathbb{R}^n$. The new notation for $A \in \mathbb{C}^{m\times n}$ specifies the choice of norm as well as the vector used to define our subspace. For stochastic matrices, we would write

$$\tau_p(S) = \tau_p(e, S)$$

but only for certain values of $p$. The difference between $\tau_p(S)$ and $\tau_p(e, S)$ is that the former is maximized over the space of $z \in \mathbb{R}^n$, but the later is maximized over $z \in \mathbb{C}^n$. In 1984 Seneta (100, §2) acknowledged this difference, and argued that in the cases of $p = 1$ and $p = \infty$, there is no difference between maximizing over the real and the complex n-dimensional spaces because the vector $z$ that attains the maximization is a real vector.

Several years later, an even more general definition of an ergodicity coefficient developed from Rothblum and Tan (91, §7) and Hartfiel (46, §3). Specifically, this ergodicity coefficient for rectangular, complex matrices is maximized over the space orthogonal to a specified matrix.

**Definition 3.6.2.** *Let* $A \in \mathbb{C}^{m\times n}$ *and* $W \in \mathbb{C}^{m\times k}$. *Then for integer* $1 \le p \le \infty$

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

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

We will use this more general definition throughout this section. We begin with an upper bound on the ergodicity coefficient and two bounds on the difference between two coefficients. Additionally, we are proving that $\tau_p(W, A)$ is continuous, well-conditioned, and subadditive.

**Theorem 3.6.3.** *For* $A \in \mathbb{C}^{m\times n}$ *and* $W \in \mathbb{C}^{m\times k}$

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

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

2. Without loss of generality, assume $\tau_p(W, A_1) \geq \tau_p(W, A_2)$. Let $y \in \mathbb{C}^m$ achieve the maximum for $\tau_p(W, A_1)$, i.e., $\tau_p(W, A_1) = \|A_1^*y\|_p$, with $\|y\|_p = 1$ and $y^*W = 0$. Then

$$\begin{aligned}
\tau_p(W, A_1) - \tau_p(W, A_2) &= \|A_1^*y\|_p - \max_{\substack{\|z\|_p=1 \\ z^*W=0}} \|A_2^*z\|_p \\
&\leq \|A_1^*y\|_p - \|A_2^*y\|_p \\
&\leq \|(A_1 - A_2)^* y\|_p \\
&\leq \max_{\substack{\|z\|_p=1 \\ z^*W=0}} \|(A_1 - A_2)^* z\|_p \\
&= \tau_p(W, A_1 - A_2)
\end{aligned}$$

3. This follows directly from items 1 and 2.

□

We now present a lemma that will be used to prove a weak submultiplicative statement as well as the subdominant eigenvalue bound.

**Lemma 3.6.4** (Lemma 7.1-1 in (36)). *Let $A \in \mathbb{C}^{n\times n}$ have eigenvalues $\lambda_i \neq \lambda_j$ for integers $i, j$ with $1 \leq i \leq k$, $k+1 \leq j \leq n$. Let $X \in \mathbb{C}^{n\times k}$ have columns that span an invariant subspace associated with $\lambda_i$, $1 \leq i \leq k$. Then for $B \in \mathbb{C}^{k\times k}$,*

$$AX = XB.$$

*The eigenvalues $\lambda_1, \cdots, \lambda_k$ are eigenvalues of both matrices $A$ and $B$, and the eigenvalues $\lambda_{k+1}, \cdots, \lambda_n$ are eigenvalues of $A$ only.*

*Proof.* Let $X$ have the QR factorization $X = Q \begin{pmatrix} R \\ 0 \end{pmatrix}$, with $Q \in \mathbb{C}^{n\times n}$ and $R \in \mathbb{C}^{k\times k}$.

From $AX = XB$ and the QR decomposition of $X$, we obtain $Q^T AQ \begin{pmatrix} R \\ 0 \end{pmatrix} = \begin{pmatrix} R \\ 0 \end{pmatrix} B$.

Let $Q^T AQ = T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}$.

From the equations $T_{11}R = RB$, $T_{21}R = 0$ and the fact that $R$ is nonsingular, we conclude that $T_{21} = 0$. Then, the eigenvalues of $T_{11}$ are the eigenvalues of $B$. The eigenvalues of $A$ are the eigenvalues of $T$, which are the eigenvalues of $T_{11}$ and $T_{22}$. □

We now obtain a weak submultiplicative statement for general square matrices.

**Theorem 3.6.5.** *Let $A \in \mathbb{C}^{n\times n}$ and $W \in \mathbb{C}^{n\times k}$ satisfy Lemma 3.6.4, and choose $l, m \geq 0$. Then*

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

*Proof.* Let $y \in \mathbb{C}^n$ achieve the maximum for $\tau_p\left(W, A^{l+m}\right)$. That is,

$$\tau_p\left(W, A^{l+m}\right) = \left\|(A^*)^{l+m} y\right\|_p, \text{ with } \|y\|_p = 1 \text{ and } y^*W = 0.$$

Let $x = (A^*)^m y / \left\|(A^*)^m y\right\|_p$, so that $\|x\|_p = 1$. Then from $AW = WC$ we obtain $A^m W = WC^m$. Along with $y^*W = 0$ we get

$$W^*x = \frac{W^*(A^*)^m y}{\left\|(A^*)^m y\right\|_p} = \frac{(C^*)^m W^* y}{\left\|(A^*)^m y\right\|_p} = 0.$$

Then

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

□

We can also say something about the submultiplicativity of the ergodicity coefficient if it operates on the product of two different matrices.

**Theorem 3.6.6.** *For $A \in \mathbb{C}^{n\times r}$, $B \in \mathbb{C}^{m\times n}$, and $W \in \mathbb{C}^{m\times k}$*

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

*Proof.* Let $y \in \mathbb{C}^m$ achieve the maximum for $\tau_p(W, BA)$. That is, $\tau_p(W, BA) =$

$\|(BA)^* y\|_p = \|A^* B^* y\|_p$, with $\|y\|_p = 1$ and $y^* W = 0$. Then

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

□

Like its stochastic counterparts, $\tau_p(W, A)$ can bound subdominant eigenvalues of $A$. See Theorems 3.5.3 and 3.5.14 for results with $\tau_1(S)$ and $\tau_\infty(S)$, respectively. Additionally, Theorem 3.7.9 from Rothblum and Tan (91, p 54) presents a version of this result for square, nonnegative matrices.

**Theorem 3.6.7.** *Let $A \in \mathbb{C}^{n \times n}$ and $W \in \mathbb{C}^{n \times k}$, where the columns of $W$ span an invariant subspace and $|\lambda_1| \geq \cdots \geq |\lambda_k| > |\lambda_{k+1}| \geq \cdots \geq |\lambda_n|$. Then*

$$|\lambda_i| \leq \tau_p(W, A), \qquad \text{for } i = k+1, \ldots, n.$$

*Proof.* Let $z$ be a unit left eigenvector of $A$ for some eigenvalue $\lambda_i$, so that $\|z\|_p = 1$ and $z^* A = \lambda_i z^*$. We see that $z^* W = 0$ because $W$ spans the space of the first $k$ eigenvalues. Then

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

for $i = k+1, \ldots, n$. □

We show that the ergodicity coefficient in Defintion 3.6.1 for $A \in \mathbb{C}^{n \times n}$ with eigenvalues $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$ can bound all the subdominant eigenvalues, i.e., all

eigenvalues $\neq \lambda_1$.

**Corollary 3.6.8.** *For $A \in \mathbb{C}^{n\times n}$, let $w$ be a right eigenvector of $\lambda_1$, i.e., $Aw = \lambda_1 w$. Then for every eigenvalue $\lambda$ such that $\lambda \neq \lambda_1$*

$$|\lambda| \leq \tau_p(w, A)$$

*for $1 \leq p \leq \infty$.*

*Proof.* Let $z$ be a unit left eigenvector of $A$ for some $\lambda \neq \lambda_1$ so that $z^*A = \lambda z^*$ or $A^*z = \overline{\lambda}z$ and $\|z\|_p = 1$. Then

$$\lambda z^* w = z^* A w = \lambda_1 z^* w.$$

Because $\lambda \neq \lambda_1$ it follows that $z^*w = 0$. Finally,

$$|\lambda| = |\overline{\lambda}| = \left|\overline{\lambda}\right| \|z\|_p = \left\|\overline{\lambda}z\right\|_p = \|A^*z\|_p \leq \max_{\substack{\|z\|_p=1 \\ w^*z=0}} \|A^*z\|_p = \tau_p(w, A).$$

□

### 3.6.1 Relations between different $p$-norm ergodicity coefficients

We now show that an ergodicity coefficient in one $p$-norm can provide upper and lower bounds on ergodicity coefficients in a different $p$-norm. This result follows from vector norm relations. We begin with a result for $p$-norms before extending the relations between different $p$-norms to ergodicity coefficients. Rhodius presents similar results for ergodicity coefficients in (88, Theorem 1).

**Theorem 3.6.9** (pp 27–28 in (33), (6.4) in (49)). *If $x \in \mathbb{C}^{n\times 1}$, and $p \leq q$ then*

$$\|x\|_q \leq \|x\|_p \leq n^{\left(\frac{1}{p}-\frac{1}{q}\right)}\|x\|_q.$$

*Proof.* We will first prove the upper bound on $\|x\|_p$. We begin by re-writing the $p$-norm of $x$ as

$$\|x\|_p^p = \sum_{i=1}^{n} |x_i|^p = \sum_{i=1}^{n} |x_i|^p \cdot 1 = \begin{pmatrix} |x_1|^p & \cdots & |x_n|^p \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Applying the Hölder inequality with $\frac{1}{k} + \frac{1}{l} = 1$,

$$\begin{aligned} \|x\|_p^p &\leq \left\| \begin{pmatrix} |x_1|^p \\ \vdots \\ |x_n|^p \end{pmatrix} \right\|_k \left\| \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right\|_l \\ &= \left( \sum_{i=1}^{n} |x_i|^{pk} \right)^{1/k} n^{1/l} \end{aligned}$$

Choosing $k = q/p$, the above expression is re-written as

$$\|x\|_p^p \leq \left( \sum_{i=1}^{n} |x_i|^q \right)^{p/q} n^{(1-\frac{p}{q})} = \|x\|_q^p \, n^{(1-\frac{p}{q})}.$$

Finally, taking the $p^{th}$ root of both sides of the inequality yields the upper bound as stated in the theorem

$$\|x\|_p \leq n^{\left(\frac{1}{p}-\frac{1}{q}\right)}\|x\|_q.$$

Now for the lower bound on $\|x\|_p$, let $y_i = |x_i|/\|x\|_p$. Then $y_i \geq 0$ and $\sum_{i=1}^n y_i^p = 1$. Clearly it follows that $0 \leq y_i \leq 1$ for each $i = 1, \ldots, n$, and if $q > p$, then $y_i^q \leq y_i^p$ for all $i$. Finally, from

$$\frac{\sum_{i=1}^n |x_i|^q}{\|x\|_p^q} = \sum_{i=1}^n y_i^q \leq \sum_{i=1}^n y_i^p = 1$$

we obtain the bound

$$\|x\|_q^q = \sum_{i=1}^n |x_i|^q \leq \|x\|_p^q$$

which is of the form we are trying to prove when both sides are raised to the $1/q$ power. □

We now extend these results for $p$-norms to the general ergodicity coefficient $\tau_p(W, A)$.

**Corollary 3.6.10.** *For $A \in \mathbb{C}^{m\times n}$ and $W \in \mathbb{C}^{n\times k}$*

$$n^{\left(\frac{1}{q}-\frac{1}{p}\right)}\tau_q(W, A) \leq \tau_p(W, A) \leq n^{\left(\frac{1}{p}-\frac{1}{q}\right)}\tau_q(W, A)$$

*with $p \leq q$.*

*Proof.* Recall

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

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

To prove the lower bound, choose $y \in \mathbb{C}^m$ to achieve the maximzation of $\tau_q(W, A)$, so that $\tau_q(W, A) = \|A^*y\|_q$ with $\|y\|_q = 1$ and $y^*W = 0$. From Theorem 3.6.9, if $p \leq q$,

then $\|x\|_p \leq n^{(1/p-1/q)}\|x\|_q$ for any complex vector $x$, which is rewritten as

$$\frac{1}{\|x\|_q} \leq n^{(1/p-1/q)}\frac{1}{\|x\|_p}.$$

Then since $\|y\|_q = 1$,

$$\tau_q(W,A) = \frac{\|A^*y\|_q}{\|y\|_q} \leq n^{\left(\frac{1}{p}-\frac{1}{q}\right)}\frac{\|A^*y\|_q}{\|y\|_p}.$$

Again from Theorem 3.6.9, since $\|x\|_q \leq \|x\|_p$, then $\|A^*y\|_q \leq \|A^*y\|_p$ since $A^*y$ is a complex vector. Finally,

$$\tau_q(W,A) \leq n^{\left(\frac{1}{p}-\frac{1}{q}\right)}\frac{\|A^*y\|_q}{\|y\|_p} \leq n^{\left(\frac{1}{p}-\frac{1}{q}\right)}\frac{\|A^*y\|_p}{\|y\|_p} \leq n^{\left(\frac{1}{p}-\frac{1}{q}\right)}\tau_p(W,A).$$

Dividing both sides by $n^{(1/p-1/q)}$ yields the form in the theorem:

$$n^{\left(\frac{1}{q}-\frac{1}{p}\right)}\tau_q(W,A) \leq \tau_p(W,A).$$

For the second inequality, choose $y \in \mathbb{C}^m$ to acheive the maximzation of $\tau_p(W,A)$, so that $\tau_p(W,A) = \|A^*y\|_p$ with $\|y\|_p = 1$ and $y^*W = 0$. From Theorem 3.6.9, for any complex vector $x$, since $\|x\|_q \leq \|x\|_p$, we obtain

$$\frac{1}{\|x\|_p} \leq \frac{1}{\|x\|_q}.$$

Then since $\|y\|_p = 1$,

$$\tau_p(W,A) = \frac{\|A^*y\|_p}{\|y\|_p} \leq \frac{\|A^*y\|_p}{\|y\|_q}.$$

Again from Theorem 3.6.9, since $\|x\|_p \leq n^{(1/p-1/q)}\|x\|_q$ for any complex vector $x$, and because $A^*y$ is a complex vector, we have

$$\tau_p(W,A) \leq \frac{\|A^*y\|_p}{\|y\|_q} \leq n^{\left(\frac{1}{p}-\frac{1}{q}\right)}\frac{\|A^*y\|_q}{\|y\|_q} \leq n^{\left(\frac{1}{p}-\frac{1}{q}\right)}\tau_q(W,A).$$

□

When $p = 1$ and $q = \infty$, we obtain the bound $\frac{1}{n}\tau_\infty(W,A) \leq \tau_1(W,A) \leq n\tau_\infty(W,A)$. When $A$ is a stochastic matrix, and $W = e$, this bound is not as tight as the one given for stochastic matrices in the 1-norm and $\infty$-norm in Theorem 3.5.26, which says $\tau_\infty(S) \leq \frac{n}{2}\tau_1(S)$ if $n$ is even and $\tau_\infty(S) \leq \frac{n-1}{2}\tau_1(S)$ if $n$ is odd.

### 3.6.2 Explicit forms for $\tau_p(w,A)$ for specific values of $p$

In this section we present work from 1984 by Seneta alone (100), as well as in collaboration with Tan (106). In these works, Seneta and Tan develop explicit forms for $\tau_p(w,A)$, with $p$ taking the values 1, $\infty$ and 2. The expressions presented hold for $w \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$. Also in 1984, Rothblum (90) presented another view to finding the explicit solutions in the cases of $p = 1$ and $p = \infty$ for square, nonnegative, irreducible matrices. His approach focuses on quantifying the extreme points of intersection of a unit ball with a hyperplane, in order to solve a maximization, and is not presented here. Rothblum's argument is similar to that of Rhodius in 1988 for $\tau_p(S)$ (85), which is given in section 3.5.3.

**Explicit forms for $\tau_1(w, A)$**

In 1984 Seneta (100) developed an explicit form for $\tau_1(w, A)$ for real, rectangular matrices $A$. We begin with the special case that $w_i \neq 0$ for all $i$. Lemma 3.5.4 shows that for any $x \neq 0$ that satisfies $x^T e = 0$, we can write $x = \sum_{i \neq j} y_{ij} \frac{1}{2}(e_i - e_j)$ for $y_{ij} > 0$ and also $\|x\|_1 = \sum_{i \neq j} y_{ij}$. We now present a new lemma in which we write $x$ with $x^T w = 0$ in a form similar to that presented in Lemma 3.5.4.

**Lemma 3.6.11** (p 192 in (100))**.** *If $x, w \in \mathbb{R}^n$ have $x \neq 0$, $w^T x = 0$, and $w_i \neq 0$ for $i = 1, \ldots, n$, write $D = \begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_n \end{pmatrix}$. Then $x = \sum_{i \neq j} y_{ij} \frac{D^{-1}(e_i - e_j)}{\|D^{-1}(e_i - e_j)\|_1}$ for $y_{ij} \geq 0$ and $\|x\|_1 = \sum_{i \neq j} y_{ij}$.*

*Proof.* As in the proof of Lemma 3.5.4, we will prove this lemma by induction on the length $n$ of the vector $x$. We begin with $n = 2$. In this case $D = \begin{pmatrix} w_1 & \\ & w_2 \end{pmatrix}$. From $0 = x^T w = x^T De = w_1 x_1 + w_2 x_2$, we know $w_1 x_1 = -w_2 x_2$. Without loss of generality, assume $w_1 x_1 > 0$, then $Dx = w_1 x_1 \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, so $x = w_1 x_1 D^{-1}(e_1 - e_2)$. However, since $D^{-1}(e_1 - e_2) = \begin{pmatrix} 1/w_1 \\ 1/w_2 \end{pmatrix}$ and $\|D^{-1}(e_1 - e_2)\|_1 = |1/w_1| + |1/w_2|$, then

$$x = w_1 x_1 D^{-1}(e_1 - e_2) = w_1 x_1 \left( \frac{1}{|w_1|} + \frac{1}{|w_2|} \right) \frac{D^{-1}(e_1 - e_2)}{\|D^{-1}(e_1 - e_2)\|_1},$$

with $y_{12} = w_1 x_1 \left( \frac{1}{|w_1|} + \frac{1}{|w_2|} \right) > 0$. Finally, because $|w_1 x_1| = |w_2 x_2|$,

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

Now assume the lemma holds for some $n \geq 2$ and we will show it holds for $n+1$. Let $x$ be an $n+1$ dimensional real vector with $x \neq 0$, $x^T w = 0$, $w_i \neq 0$ for $i = 1, \ldots, n+1$ and diagonal matrix $D$ defined so that $De = w$.

From $x^T w = 0$ we obtain $x^T De = 0$. Then $Dx$ is orthogonal to $e$. Without loss of generality, assume $Dx$ is permuted so that $w_n x_n > 0$, $w_{n+1} x_{n+1} < 0$ and $w_n x_n = \max_{i=1,\ldots,n+1} |w_i x_i|$. Then

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

and multiplying by $D^{-1}$ yields

$$x = \begin{pmatrix} x_{1:n-1} \\ x_n + \frac{w_{n+1} x_{n+1}}{w_n} \\ 0 \end{pmatrix} - w_{n+1} x_{n+1} D^{-1} (e_n - e_{n+1}). \tag{3.9}$$

Let $z = \begin{pmatrix} x_{1:n-1} \\ x_n + \frac{w_{n+1}x_{n+1}}{w_n} \\ 0 \end{pmatrix}$. Then

$$z_{1:n}^T w_{1:n} = \sum_{i=1}^{n-1} w_i x_i + w_n \left( x_n + \frac{w_{n+1}x_{n+1}}{w_n} \right) = x^T w.$$

Recall $x^T w = 0$, so $z_{1:n}^T w_{1:n} = 0$, and since $z_{1:n} \in \mathbb{R}^n$ we can apply the induction hypothesis to $z_{1:n}$ and write

$$z_{1:n} = \sum_{i \neq j}^{n} y_{ij} \frac{D_{1:n}^{-1}(e_i - e_j)}{\left\| D_{1:n}^{-1}(e_i - e_j) \right\|_1}$$

where $D_{1:n}$ contains the first $n$ rows and columns of $D$, and $y_{ij} \geq 0$, $\|z_{1:n}\|_1 = \sum_{i \neq j} y_{ij}$.

We now relate $z$ to $z_{1:n}$. Clearly $\|z\|_1 = \|z_{1:n}\|_1 = \sum_{i \neq j} y_{ij}$. Also, by extending all the canonical vectors to be in $\mathbb{R}^{n+1}$ instead of $\mathbb{R}^n$,

$$z = \sum_{i \neq j}^{n+1} y_{ij} \frac{D^{-1}(e_i - e_j)}{\left\| D^{-1}(e_i - e_j) \right\|_1}. \tag{3.10}$$

Going back to the vector $x$ in expression 3.9, and specifically, to $-w_{n+1}x_{n+1}D^{-1}(e_n - e_{n+1})$. From $D^{-1}(e_n - e_{n+1}) = \frac{1}{w_n} - \frac{1}{w_{n+1}}$, we obtain $\left\| D^{-1}(e_n - e_{n+1}) \right\|_1 = \frac{1}{|w_n|} + \frac{1}{|w_{n+1}|}$. Then writing

$$y_{n,n+1} = -w_{n+1}x_{n+1} \left( \frac{1}{|w_n|} + \frac{1}{|w_{n+1}|} \right)$$

the vector $-w_{n+1}x_{n+1}D^{-1}(e_n - e_{n+1})$ becomes

$$-w_{n+1}x_{n+1}D^{-1}(e_n - e_{n+1}) = y_{n,n+1}\frac{D^{-1}(e_n - e_{n+1})}{\|D^{-1}(e_n - e_{n+1})\|_1}. \tag{3.11}$$

Clearly $y_{n,n+1} > 0$ from our initial assumption that $w_{n+1}x_{n+1} < 0$.

Combining the expressions in (3.10) and (3.11), we write $x$ from (3.9) as

$$x = \sum_{i\neq j} y_{ij}\frac{D^{-1}(e_i - e_j)}{\|D^{-1}(e_i - e_j)\|_1} + y_{n,n+1}\frac{D^{-1}(e_n - e_{n+1})}{\|D^{-1}(e_n - e_{n+1})\|_1}.$$

Since $y_{ij} \geq 0$ for all $i, j$, we have satisfied the first part of the theorem. It remains to prove that $\|x\|_1 = \sum_{i\neq j}^{n+1} y_{ij}$.

Begin by examining $\sum_{i\neq j} y_{ij} = \|z\|_1$,

$$\|z\|_1 = \|x_{1:n-1}\|_1 + \left|x_n + \frac{w_{n+1}}{w_n}x_{n+1}\right| = \|x_{1:n-1}\|_1 + \left|\frac{1}{w_n}\right| |w_n x_n + w_{n+1}x_{n+1}|$$

and because $w_n x_n + w_{n+1}x_{n+1} > 0$ we can remove the set of absolute value signs around this sum in the expression for $\|z\|_1$ to obtain

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

Additionally, from $y_{n,n+1}$ we obtain

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

Finally,

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

We have proven the theorem because $\|x\|_1 = \sum_{i\neq j} y_{ij}$. □

When $w = e$, then $D = I$ and we recover Lemma 3.5.4.

Like in section 3.5.1, we apply the expression in Lemma 3.6.11 to a vector $z$ with $z^T w = 0$ to find an explicit form for $\tau_1(w, A)$.

**Theorem 3.6.12** (p 193 in (100))**.** *Let $A \in \mathbb{R}^{m\times n}$, $w \in \mathbb{R}^m$ and $w_i \neq 0$ for all $i$. Then*

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

*Proof.* Recall the definition of an ergodicity coefficients from Definition 3.6.1:

$$\tau_1(w, A) = \max_{\substack{\|z\|_1=1 \\ z^T w=0}} \left\|A^T z\right\|_1$$

for $z \in \mathbb{R}^m$. Choose $z$ with $z^T w = 0$ and $\|z\|_1 = 1$, so that we can apply Lemma 3.6.11

to write

$$z = \sum_{i \neq j} y_{ij} D^{-1} (e_i - e_j) / \left\| D^{-1} (e_i - e_j) \right\|_1,$$

with $d_{ii} = w_i$ for $i = 1, \ldots, m$. Then

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

From Lemma 3.6.11, $\|z\|_1 = \sum_{i \neq j} y_{ij}$, and by assumption, $\|z\|_1 = 1$. Finally, choose $l$ and $m$, so

$$z = \frac{D^{-1}(e_l - e_m)}{\|D^{-1}(e_l - e_m)\|_1}$$

achieves the maximum of $\left\|A^T D^{-1}(e_i - e_j)\right\|_1 / \|D^{-1}(e_i - e_j)\|_1$. Then

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

□

It is easy to see that this result is analogous to that for $\tau_1(S)$ in section 3.5.1. As mentioned previously, $\tau_1(S)$ can be written as $\tau_1(e, S)$. In this situation, $w = e$, so $D = I$. Then from $\|e_i - e_j\|_1 = 2$,

$$\tau_1(e, S) = \max_{ij} \frac{\left\|S^T D^{-1}(e_i - e_j)\right\|_1}{\|D^{-1}(e_i - e_j)\|_1} = \frac{1}{2} \max_{ij} \left\|S^T (e_i - e_j)\right\|_1,$$

which is the result presented in Theorem 3.5.5.

In order to avoid the restriction that $w_i \neq 0$ for $i = 1, \ldots, m$ we present the following corollary to Theorem 3.6.12.

**Corollary 3.6.13** (p 194 in (100)). *Let $A \in \mathbb{R}^{m\times n}$ and $w \in \mathbb{R}^m$, $w \neq 0$. Choose permutation matrix $P$ so that $Pw = \begin{pmatrix} w_l \\ 0 \end{pmatrix}$, where $w_l \in \mathbb{R}^l$ contains the nonzero elements of $w$ and $1 < l \leq m$. Then*

$$\tau_1(w, A) = \max\left\{\tau_1(w_l, A_l),\ \left\|A_{m-l}^T\right\|_1\right\}$$

*with $PA = \begin{pmatrix} A_l \\ A_{m-l} \end{pmatrix}$, $A_l \in \mathbb{R}^{l\times n}$ and $A_{m-l} \in \mathbb{R}^{(m-l)\times n}$.*

*Proof.* Again, recall the general definition for $\tau_1(w, A)$, for $z \in \mathbb{R}^m$

$$\tau_1(w, A) = \max_{\substack{\|z\|_1=1 \\ z^T w=0}} \left\|A^T z\right\|_1 .$$

Permute $z \in \mathbb{R}^m$ accordingly, so $Pz = \begin{pmatrix} z_l \\ z_{m-l} \end{pmatrix}$, for $z_l \in \mathbb{R}^l$ and $z_{m-l} \in \mathbb{R}^{m-l}$. Then

$$A^T z = A^T P^T P z = A_l^T z_l + A_{m-l}^T z_{m-l}. \tag{3.12}$$

If $z_l = 0$, then $\left\|A^T z\right\|_1 \leq \|A_{m-l}\|_1 \|z_{m-l}\|_1$ and $\|z\|_1 = \|z_{m-l}\|_1$ imply

$$\frac{\left\|A^T z\right\|_1}{\|z\|_1} \leq \left\|A_{m-l}^T\right\|_1 .$$

Equality holds if $Pz = e_i$ for some $l + 1 \leq i \leq m$.

Now consider the possibility that $z_l \neq 0$. Then $0 = z^T w = z^T P^T P w = z_l^T w_l$. Since

$z_l^T w_l = 0$, we can apply Lemma 3.6.11 to obtain

$$z_l = \sum_{i \neq j} y_{ij} \frac{D_l^{-1}(e_i - e_j)}{\left\| D_l^{-1}(e_i - e_j) \right\|_1},$$

for $\sum_{i \neq j} y_{ij} = \|z_l\|_1$ and $D_l$ is a diagonal matrix with the elements of $w_l$ on the diagonal. Substituting this form of $z_l$ into (3.12) gives

$$\left\| A^T z \right\|_1 \leq \|z_l\|_1 \max_{i \neq j} \frac{\left\| A_l^T D_l^{-1}(e_i - e_j) \right\|_1}{\left\| D_l^{-1}(e_i - e_j) \right\|_1} + \left\| A_{m-l}^T \right\|_1 \|z_{m-l}\|_1.$$

Then

$$\begin{aligned} \frac{\left\| A^T z \right\|_1}{\|z\|_1} &\leq \frac{\|z_l\|_1}{\|z\|_1} \max_{i \neq j} \frac{\left\| A_l^T D_l^{-1}(e_i - e_j) \right\|_1}{\left\| D_l^{-1}(e_i - e_j) \right\|_1} + \left\| A_{m-l}^T \right\|_1 \frac{\|z_{m-l}\|_1}{\|z\|_1} \\ &\leq \frac{\|z_l\|_1 + \|z_{m-l}\|_1}{\|z\|_1} \max \left\{ \max_{i \neq j} \frac{\left\| A_l^T D_l^{-1}(e_i - e_j) \right\|_1}{\left\| D_l^{-1}(e_i - e_j) \right\|_1}, \left\| A_{m-l}^T \right\|_1 \right\} \end{aligned}$$

The result follows since $\left(\|z_l\|_1 + \|z_{m-l}\|_1\right) / \|z\|_1 = 1$ and from Theorem 3.6.12, $\tau_1\left(w_l, A_l\right) = \max_{i \neq j} \left\| A_l^T D_l^{-1}(e_i - e_j) \right\|_1 / \left\| D_l^{-1}(e_i - e_j) \right\|_1$. □

**An explicit form for $\tau_\infty(w, A)$**

In the same paper from 1984 (100) in which he developed an explicit form for $\tau_1(w, A)$, Seneta also presented an explicit form for $\tau_\infty(w, A)$ for real, rectangular matrices $A$.

**Theorem 3.6.14** (§3 in (100))**.** *For $A \in \mathbb{R}^{m \times n}$ and $w \in \mathbb{R}^m$, maximizing over $z \in \mathbb{R}^m$*

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

*Let $a^T$ be some row of $A^T$ from $e_i^T A^T$, so that $a_j$ is the $j^{th}$ element of $e_i^T A^T$. Assume the $w_j$'s are permuted (along with the rows of $A^T$) so that $w_j \neq 0$ for $j = 1, \ldots, k$ and $\frac{a_1}{w_1} \leq \cdots \leq \frac{a_k}{w_k}$, and choose $l$ to be the smallest integer such that $\sum_{j=1}^{l} |a_j/w_j| \geq \sum_{j=l+1}^{k} |a_j/w_j|$. Then*

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

*Proof.* Our goal in this proof is to break-down and permute the vectors $a$ and $w$ into an appropriate form so that we can determine the elements of vector $z$ that achieve the maximization of $\left|e_i^T A^T z\right|$.

For arbitrary scalar $\alpha$ and $a^T$ as a row of $A^T$

$$\left|a^T z\right| = \left|\left(a^T - \alpha w^T\right) z\right| \leq \sum_{j=1}^{n} \left|a_j - \alpha w_j\right|, \tag{3.13}$$

where $a_j$ is the $j^{th}$ element of $a^T$. This inequality holds from the triangle inequality and the fact that $\max |z_j| = 1$. Then we can re-write the right-hand side of expression (3.13) as

$$|a^T z| \leq \sum_{j=1}^{n} |a_j - \alpha w_j| = \sum_{j=1}^{k} |w_j| \left|\frac{a_j}{w_j} - \alpha\right| + \sum_{j=k+1}^{n} |a_j|, \tag{3.14}$$

The value of $\alpha$ that minimizes the right-hand side of expression (3.14) is the median of the $a_j/w_j$'s. With $l$ as stated in the theorem, $a_l/w_l$ is the median of the $a_j/w_j$'s, and

$\alpha = a_l/w_l$.

Now expressions (3.13) and (3.14) become

$$\begin{aligned}|a^T z| &\leq \sum_{j=1}^{k} |w_j| \left| \frac{a_j}{w_j} - \alpha \right| + \sum_{i=k+1}^{n} |a_j| \\ &= \sum_{j=1}^{l-1} |w_j| \left( \frac{a_l}{w_l} - \frac{a_j}{w_j} \right) + \sum_{j=l+1}^{k} |w_j| \left( \frac{a_j}{w_j} - \frac{a_l}{w_l} \right) = \sum_{j=k+1}^{n} |a_j|.\end{aligned}$$

Finally, the value of $z$ that achieves the minimum of $\left|a^T z\right|$ is

$$z_i = \begin{cases} \frac{-|w_i|}{w_i}, & \text{for } i = 1, \ldots, l-1 \\ \frac{1}{w_l} \left( \sum_{i=1}^{l-1} |w_i| - \sum_{i=l+1}^{k} |w_i| \right), & \text{for } i = l \\ \frac{|w_i|}{w_i}, & \text{for } i = l+1, \ldots, k \\ \text{sign } a_i, & \text{for } i = k+1, \ldots, n \end{cases}$$

□

**An explicit form for $\tau_2(w, A)$**

In 1984 Seneta and Tan (106) developed an upper bound on $\tau_2(w, A)$ for real, rectangular matrices, $A$ and real vectors $w$. Several years earlier, Seneta (97) investigated the special case of $\tau_2(S)$ for stochastic matrices. Seneta (97, p 586) briefly discusses his an idea for a proof that an $n \times n$ stochastic matrix $S$ has $\tau_2(S) = \lambda_1^{1/2}$ where $\lambda_1$ is the dominant eigenvalue of a positive semidefinite matrix. A similar argument to Seneta's (97, p 586) is used by Tan (110) for an ergodicity coefficient operating on a

different vector. Tan's result is presented in our Theorem 3.6.20 for $\tau_2\left(\pi, S^T\right)$, where $\pi$ denotes the stationary distribution of $S$.

In this section we present a different approach for general matrices than Seneta presented in (97). Additionally, we extend Seneta's and Tan's definition to complex matrices and vectors. First, however we recall the definition of the adjugate of a matrix. The formula relating determinants and adjugates will play a crucial role in the proof of Theorem 3.6.19.

**Definition 3.6.15** (p 479 in (76), p 137 in (78))**.** *For $A \in \mathbb{C}^{n\times n}$, the adjugate of $A$ or adjoint matrix, denoted* adj$[A]$*, is the transpose of the matrix of cofactors, and*

$$det(A) \cdot I = A \cdot \text{adj}[A].$$

We derive three different characterizations for the ergodicity coefficient $\tau_2(w, A)$, however we first begin by expressing $\tau_2(W, A)$ as an unconstrained maximization problem. We then show that $\tau_2(w, A)$, when applied to a matrix $A$ with $m$ rows, is expressed as the two-norm of a matrix with $m-1$ rows.

**Theorem 3.6.16.** *Let $A$ be a complex $m \times n$ matrix, $W \in \mathbb{C}^{m\times k}$ with orthonormal columns, and $m \times m$ unitary matrix $Q$ with $Q = \begin{pmatrix} W & Q_2 \end{pmatrix}$. Partition $A^*Q = \begin{pmatrix} A_k^* & A_{m-k}^* \end{pmatrix}$, where $A_k$ is $k \times n$ and $A_{m-k}$ is $(m-k) \times n$. Then*

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

*Proof.* Let $z$ be a $m \times 1$ vector with $\|z\|_2 = 1$ and $z^*W = 0$. Then $Q^*z = \begin{pmatrix} 0 \\ \hat{z} \end{pmatrix}$, where $\hat{z} = Q_2^*z$ is $(m-k) \times 1$, $\|\hat{z}\|_2 = 1$, so

$$A^*z = A^*QQ^*z = \begin{pmatrix} A_k^* & A_{m-k}^* \end{pmatrix} \begin{pmatrix} 0 \\ \hat{z} \end{pmatrix} = A_{m-k}^*\hat{z}.$$

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

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

□

**Corollary 3.6.17.** *Let $A$ be a complex $m \times n$ matrix, $w$ an $m \times 1$ vector with $w \neq 0$, and $Q$ a $m \times m$ unitary matrix with leading column $Qe_1 = w/\|w\|_2$. Partition $A^*Q = \begin{pmatrix} a & A_{m-1}^* \end{pmatrix}$, where $A_{m-1}$ is $(m-1) \times n$. Then*

$$\tau_2(w, A) = \|A_{m-1}\|_2.$$

*Proof.* Let $z$ be a $m \times 1$ vector with $\|z\|_2 = 1$ and $z^*w = 0$. Then $Q^*z = \begin{pmatrix} 0 \\ \hat{z} \end{pmatrix}$, where $\hat{z}$ is $(m-1) \times 1$, $\|\hat{z}\|_2 = 1$, and

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

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

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

□

With the help of Corollary 3.6.17 we derive a second characterization of $\tau_2(w, A)$: A vector $y$ that achieves the maximum for $\tau_2(w, A)$ is expressed as the solution of a linear system with right-hand side $w$.

**Theorem 3.6.18.** *In addition to the conditions of Corollary 3.6.17, let* $\tau \equiv \tau_2(w, A) = \|A^*y\|_2$ *where* $\|y\|_2 = 1$ *and* $y^*w = 0$. *Then*

$$(AA^* - \tau^2 I)y = \frac{w^*AA^*y}{\|w\|_2^2} w.$$

*Proof.* We represent the singular value problem $\|A_{m-1}\|_2$ from Corollary 3.6.17 as an eigenvalue problem,

$$Q^*AA^*Q = \begin{pmatrix} a^*a & a^*A^*_{m-1} \\ A_{m-1}a & B \end{pmatrix}, \qquad \text{w}\textit{here} \quad B \equiv A_{m-1}A^*_{m-1}.$$

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

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

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

$$\begin{aligned}(AA^* - \tau^2 I)y &= Q\left[Q^*AA^*Q - \tau^2 I\right]Q^*y = Q\begin{pmatrix} a^*a - \tau^2 & a^*A_{m-1}^* \\ A_{m-1}a & B - \tau^2 I \end{pmatrix}\begin{pmatrix} 0 \\ \hat{y} \end{pmatrix} \\ &= Q\begin{pmatrix} a^*A_{m-1}^*\hat{y} \\ (B - \tau^2 I)\hat{y} \end{pmatrix} = Q\begin{pmatrix} a^*A_{m-1}^*\hat{y} \\ 0 \end{pmatrix} = a^*A_{m-1}\hat{y}Qe_1.\end{aligned}$$

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

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

Hence

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

□

Theorem 3.6.18 in turn leads to a third characterization, which was introduced by Seneta and Tan (106). In contrast to our derivation, theirs relies on Lagrange multipliers for the solution of the constrained maximization problem

$$\tau_2(w, A) = \max_{\substack{\|z\|_2=1 \\ z^*w=0}} \|A^*z\|_2.$$

**Theorem 3.6.19** (Proposition 1 in (106)). *Let $A$ be a complex $m \times n$ matrix, $w$ an $m \times 1$ vector with $w \neq 0$. Then $(\tau_2(w, A))^2$ is the largest root of the polynomial $w^*\text{adj}[AA^* - \lambda I]w$ of degree $m - 1$ in $\lambda$.*

*Proof.* We distinguish two cases, depending on whether $y$ is an eigenvector of $AA^*$ or not.

- If $y$ is not an eigenvector of $AA^*$, then $\tau^2 = y^*AA^*y$ cannot be an eigenvalue of $AA^*$, so that the matrix $AA^* - \tau^2 I$ is nonsingular. The relation between inverse and adjugate implies

$$y = (AA^* - \tau^2 I)^{-1}w = \frac{\text{adj}[AA^* - \tau^2 I]w}{\det(AA^* - \tau^2 I)}.$$

  Multiplying the above by $w^*$ yields $0 = w^*y = w^*\text{adj}[AA^* - \tau^2 I]w$. This means that $\tau^2$ is a root of $w^*\text{adj}[AA^* - \tau^2 I]w$.

- If $y$ is an eigenvector of $AA^*$ then $\tau^2 = y^*AA^*y$ is an eigenvalue of $AA^*$. Let $AA^* = V\Omega V^*$ be an eigenvalue decomposition, where $V$ is unitary, and $\Omega$ is diagonal. For any scalar $\lambda$, we obtain $\text{adj}[AA^* - \lambda I] = V\text{adj}[\Omega - \lambda I]V^*$, because $\text{adj}(XY) = \text{adj}(Y)\text{adj}(X)$ for square matrices $X$ and $Y$, and $\text{adj}(V) = \det(V)V^*$ for a unitary matrix $V$. Thus

$$w^*\text{adj}[AA^* - \lambda I]w = w^*V\text{adj}[\Omega - \lambda I]V^*w.$$

  Assume that the diagonal elements $\omega_1, \ldots, \omega_m$ of $\Omega$ are ordered so that $\omega_1 = \tau^2$ and $y = Ve_1$. From $0 = y^*w = e_1^*V^*w$ follows that the leading element of $V^*w$ is

zero,

$$V^*w = \begin{pmatrix} 0 & w_2 & \dots & w_m \end{pmatrix}^T.$$

The matrix $\Omega - \lambda I$ is a diagonal matrix, and so is its adjugate,

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

Substituting the expressions for $V^*w$ and $\text{adj}[\Omega - \lambda I]$ into $w^*\text{adj}[AA^* - \lambda I]w$ gives

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

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

Since each product $(\omega_1 - \lambda)\prod_{j=2,j\neq i}^{m}(\omega_j - \lambda)$ in $w^*\text{adj}[AA^* - \lambda I]w$ consists of $m-1$ factors, the quantity $w^*\text{adj}[AA^* - \lambda I]w$ is a polynomial of degree $m-1$ in $\lambda$. We still need to argue that $\omega_1 = \tau^2$ is the largest root of this polynomial. Applying the partitioning in Theorem 3.6.17 to the adjugate gives

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

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

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

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

Tan presents an explicit form for $\tau_2\left(\pi, S^T\right)$ where $\pi$ is the stationary distribution of the stochastic matrix $S$ (110, Theorem 2). We extend this expression to complex, rectangular matrices and arbitrary vectors. Our expression for $\tau$ is much simpler than the one presented by Tan in (110, Theorem 2), and suggests $\tau_2(w, A)$ can be interpreted as the norm of $A^*$ projected onto the subspace orthogonal to $w$.

**Theorem 3.6.20.** *Let $A \in \mathbb{C}^{m\times n}$ and $w \in \mathbb{C}^m$ with $w \neq 0$. Then*

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

*If in addition $\|w\|_2 = 1$ then $\tau_2(w, A) = \|A(I - ww^*)\|_2$.*

*Proof.* Without loss of generality we can assume that the rows of $A$ and the elements of $w$ have been permuted so that the trailing element of $w$ is nonzero, i.e. $w_m \neq 0$. Let $z \in \mathbb{C}^m$ be a vector with $\|z\|_2 = 1$ and $w^*z = 0$. The idea of the proof is to use the constraint $w^*z = 0$ to remove the trailing element from $z$, and convert the maximization into one over vectors of length $m - 1$.

Partition the rows of $A$, and the elements of $w$ and $z$ so as to distinguish the trailing

element,

$$A^* = \begin{pmatrix} A_1 & a_m \end{pmatrix}, \qquad z = \begin{pmatrix} z_1 \\ z_m \end{pmatrix}, \qquad w = \begin{pmatrix} w_1 \\ w_m \end{pmatrix}.$$

Here $A_1$ has $m-1$ columns, and $z_1$ and $w_1$ are column vectors with $m-1$ elements. From $w^*z = 0$ and $w_m \neq 0$ follows $z_m = -w_1^* z_1 / \overline{w}_m$. Substituting this into $A^*z$ gives

$$A^*z = A_1 z_1 + a_m z_m = A_1 z_1 - \frac{a_m w_1^*}{\overline{w}_m} z_1 = \left( A_1 - \frac{a_m w_1^*}{\overline{w}_m} \right) z_1.$$

Setting $R \equiv A_1 - a_m w_1^* / \overline{w}_m$ gives $A^*z = Rz_1$.

Substituting the expression for $z_m$ into $\|z\|_2 = 1$ gives

$$1 = z^*z = z_1^* z_1 + |z_m|^2 = z_1^* z_1 + z_1^* \frac{w_1 w_1^*}{|w_m|^2} z_1 = z_1^* \left( I + \frac{w_1 w_1^*}{|w_m|^2} \right) z_1.$$

Setting $P \equiv I + w_1 w_1^* / |w_m|^2$ gives $z_1^* P z_1 = 1$.

Since $z^*AA^*z = z_1^* R^* R z_1$ and $1 = \|z\|_2 = z_1^* P z_1$, the problem of maximizing $z^*AA^*z$ subject to $\|z\|_2 = 1$ and $w^*z = 0$ has been converted to the problem of maximizing $x^*R^*Rx$ subject to $x^*Px = 1$.

Because $P$ is Hermitian positive definite, it has a Cholesky factorization $P = L^*L$. With $y \equiv Lx$ we obtain $1 = x^*Px = y^*y$ and $Rx = RL^{-1}y$. Thus the problem of maximizing $x^*R^*Rx = \|Rx\|_2^2$ subject to $x^*Px = 1$ has been converted to the problem of maximizing $\|RL^{-1}y\|_2$ subject to $\|y\|_2 = 1$. In other words,

$$\tau_2(w, A)^2 = \|RL^{-1}\|_2^2 = \|RP^{-1}R^*\|_2.$$

Since $P^{-1} = I - w_1 w_1^*/\|w\|_2^2$, one can show that $RP^{-1}R^* = A(I - ww^*)A^*/\|w\|_2^2$, which gives the desired expression for $\tau_2(w, A)^2$.

If $\|w\|_2 = 1$ then $I - ww^*$ is idempotent and Hermitian, so that

$$\|A(I - ww^*)A^*\|_2 = \|A(I - ww^*)\|_2^2.$$

□

## 3.7 Ergodicity coefficients for matrices with special properties

This section is a collection of different results for special types of matrices. We begin with a subsection on matrices with a constant row sum. Results for these matrices follow by extending results for stochastic matrices, which are a specialized class of constant row sum matrices. This subsection is followed by one focusing on doubly stochastic matrices, or matrices with elements in $[0, 1]$ whose rows and columns both sum to 1.

The final subsection focuses on subdominant eigenvalues bounds for nonnegative, irreducible matrices. These matrices have been the subject of investigation because, from Perron-Frobenius theory, they have a unique dominant eigenvalue with unique, normalized positive dominant left and right eigenvectors. When a nonnegative, irreducible matrix is also stochastic, it is called ergodic or regular (see Definition 3.5.9).

### 3.7.1 Constant row sum matrices

Alpin and Gabassov (1), Pykh (84) and Seneta (97, p 584), (103, pp 189–191) have all investigated the set of constant row sum matrices, since they are a simple extension from stochastic matrices. For a matrix $A \in \mathbb{R}^{n\times n}$ with constant row sum $a$,

$$\tau_1(e, A) = \frac{1}{2}\max_{ij}\sum_{s=1}^{n}|a_{is} - a_{js}| = a - \min_{ij}\sum_{s}\min\{a_{is}, a_{js}\}$$

which are extensions of the explicit forms for $\tau_1(S)$ in Theorem 3.5.5 and Corollary 3.5.6, or a special case of Theorem 3.6.12. An eigenvalue bound follows for all eigenvalues $\lambda \neq a$ of $A$,

$$|\lambda| \leq \tau_1(e, A) \tag{3.15}$$

as an extension of Theorem 3.5.3 or a special case of Corollary 3.6.8.

In the cases where $A \in \mathbb{R}^{n\times n}$ does not have constant row sums, two options are proposed to change $A$ into an $(n+1)\times(n+1)$ matrix with equal row sums without changing the spectrum, except for adding another known eigenvalue. One option adds a column, in which each entry is the negative of the respective row sum. A final row of zeros is then added so that $a = 0$ (1), (84, p 75), (97, p 584). Seneta (97, p 584) presents a second option in which the added column adjusts the row sums to some set value $a$ (often chosen as the maximum row sum of $A$). An additional row of all zeros, except the $(n+1, n+1)$ position, which is $a$. This latter case may be of interest if $A \in \mathbb{R}^{n\times n}$ has nonnegative elements, since the new $(n+1)\times(n+1)$ matrix could also be nonnegative if $a$ is chosen appropriately.

Rothblum and Tan (91) present ergodicity coefficients and eigenvalue bounds of

scaled, square, nonnegative, irreducible matrices. We begin with a lemma illustrating that any square, nonnegative, irreducible matrix can be transformed into a constant row sum matrix.

**Lemma 3.7.1** (p 57 in (91))**.** *For nonnegative irreducible matrices $A \in \mathbb{R}^{n\times n}$ with dominant eigenvalue $\lambda_1 > 0$, there exists a diagonal matrix $D$ with positive elements, so that $D^{-1}AD$ has constant row sums $\lambda_1$.*

*Proof.* From Perron-Frobenius theory, we know that $A$ has a positive right eigenvector associated with $\lambda_1$, so $Au = \lambda_1 u$ with $u_i > 0$ for $i = 1, \ldots, n$. We can write $u = De$ so that $D$ is a diagonal matrix with $d_{ii} = u_i$ for $i = 1, \ldots, n$. Then $ADe = \lambda_1 De$. Multiplying both sides of the equation $ADe = \lambda_1 De$ by $D^{-1}$ on the right, yields $D^{-1}ADe = \lambda_1 e$, and so $D^{-1}AD$ is a nonnegative matrix whose rows sum to $\lambda_1$. □

The matrices $A$ and $D^{-1}AD$ are similar and have the same eigenvalues since $D$ is a similarity transformation. This conclusion leads to a new eigenvalue bound which may or may not be easier to compute than the eigenvalue bound in Theorem 3.7.9. This is the same result as in expression (3.15) when $p = 1$ and the matrix is nonnegative and irreducible.

**Corollary 3.7.2** (p 63 in (91))**.** *Let $A \in \mathbb{R}^{n\times n}$ be a nonnegative irreducible matrix with eigenvalues $\lambda_i$, labeled so that $\lambda_1 \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Let $D = diag(w)$, where $w > 0$ is a right eigenvector of $A$ associated with $\lambda_1$. Then*

$$|\lambda_i| \leq \tau_p^R \left(e, D^{-1}AD\right), \text{ for } i = 2, \ldots, n.$$

*Proof.* Lemma 3.7.1 implies that $D^{-1}AD$ has constant row sums and hence, $e$ is a dominant right eigenvector. Since $A$ and $D^{-1}AD$ have the same eigenvalues, we obtain from Theorem 3.7.9,

$$|\lambda_i| = \left|\lambda_i \left(D^{-1}AD\right)\right| \leq \tau_p^R \left(e, D^{-1}AD\right), \text{ for } i = 2, \ldots, n.$$

□

We present an example that applies this type of similarity transformation to stochastic matrices, to show that the transformed matrix can provide tighter bounds on the modulus of subdominant eigenvalues than the original matrix. Recall that stochastic matrices are a special case of constant row sum matrices. From his 1983 paper, Seneta (99, p 346) illustrated that one can obtain tighter bounds on the non-unit eigenvalues of a stochastic matrix with a similarity formed from the stationary distribution. Tan also used the stationary distribution in conjunction with ergodicity coefficients in (110) to examine properties and develop explicit forms. His results are mentioned in section 3.8.

Let $S \in \mathbb{R}^{n\times n}$ be a stochastic matrix with a positive stationary distribution $\pi$ ($\pi_i > 0$ for all $i$), so that $\pi^T S = \pi^T$ and $\pi^T e = 1$. Then define stochastic $\hat{S} \in \mathbb{R}^{n\times n}$ with elements $\hat{s}_{ij} = \pi_j s_{ji}/\pi_i$ for $i, j = 1, \ldots, n$. The matrix $\hat{S}$ is obtained from a similarity transformation on $S^T$,

$$\hat{S} = Q^{-1}S^T Q, \qquad \text{for } Q = \begin{pmatrix} \pi_1 & & \\ & \ddots & \\ & & \pi_n \end{pmatrix},$$

so $S$ and $\hat{S}$ have the same eigenvalues. From $Qe = \pi$, $S^T\pi = \pi$ and $Q^{-1}\pi = e$ it follows that $\hat{S}e = e$, so $\hat{S}$ is stochastic.

In some cases, the ergodicity coefficients for $\hat{S}$ may provide a better bound than the ergodicity coefficients for $S$. For example, take the following matrices from Seneta (99, p 346)

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

Since

$$Q = \begin{pmatrix} \frac{5}{13} & & & \\ & \frac{5}{13} & & \\ & & \frac{3}{26} & \\ & & & \frac{3}{26} \end{pmatrix}, \qquad \text{then } \hat{S} = Q^{-1}S^TQ = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{5}{16} & \frac{5}{16} & \frac{3}{16} & \frac{3}{16} \\ \frac{5}{16} & \frac{5}{16} & \frac{3}{16} & \frac{3}{16} \\ \frac{5}{16} & \frac{5}{16} & \frac{3}{16} & \frac{3}{16} \end{pmatrix}.$$

With eigenvalues $1, 3/16, 0, 0$ for matrices $S$ and $\hat{S}$, the ergodicity coefficients are:

$$\tau_1(S) = \frac{1}{2} \qquad \tau_1\left(S^T\right) = \frac{13}{16} \qquad \tau_1\left(\hat{S}\right) = \frac{3}{8}$$

and

$$\tau_\infty(S) = 1 \qquad \tau_\infty\left(S^T\right) = \frac{5}{8} \qquad \tau_\infty\left(\hat{S}\right) = \frac{3}{16}.$$

Thus $\tau_1\left(\hat{S}\right)$ and $\tau_\infty\left(\hat{S}\right)$ provide the tightest upper bounds on the subdominant eigenvalues of $S$, $S^T$ and $\hat{S}$.

### 3.7.2 Doubly stochastic matrices

In this section we focus on doubly stochastic matrices, $P$. A doubly stochastic matrix is both row and column stochastic, so that both rows and columns sum to 1, with $0 \leq p_{ij} \leq 1$ for all $i, j$ (50, p 197).

From Theorem 3.5.2, $\tau_1(S) = 0$ if and only if $\text{rank}(S) = 1$. We present this relation for doubly stochastic matrices.

**Theorem 3.7.3** (Property 1 in (80))**.** *For doubly stochastic $P \in \mathbb{R}^{n \times n}$,*

$$\tau_1(P) = 0 \iff \text{rank}(P) = 1$$

*and*

$$\tau_1\left(P^T\right) = 0 \iff \text{rank}(P) = 1.$$

*If* $\text{rank}(P) = 1$*, then* $P = \frac{1}{n} e e^T$*.*

From Theorems 3.5.3 and 3.5.14, we know that $|\lambda| \leq \tau_p(P)$ for any eigenvalue $\lambda \neq 1$ of $P$, with $p = 1, \infty$. Since the dominant eigenvalue of $P$ is 1, the ergodicity coefficient is useful as an eigenvalue bound when $\tau_p(P) < 1$. In his paper from 1983, Seneta (99) compares the values of $\tau_1(P)$ and $\tau_\infty(P)$, in particular when $\tau_\infty(P) \leq 1$.

First, Seneta noticed that $\tau_1(P) = 1$ implies that $\tau_\infty\left(P^T\right) = 1$. To prove this we will be using the explicit form for $\tau_1(P)$ from Corollary 3.5.6,

$$\tau_1(P) = 1 - \min_{ij} \sum_{l=1}^{n} \min\left\{p_{il}, p_{jl}\right\},$$

as well as the explicit form for $\tau_\infty(P)$ from Theorem 3.5.15,

$$\tau_\infty(P) = \max_p \phi(p)$$

where $p$ is a column of $P$ and

$$\phi(p) = \begin{cases} \sum_{i=1}^{\frac{n}{2}} p_i - \sum_{i=\frac{n}{2}+1}^{n} p_i, & \text{if } n \text{ is even} \\ \sum_{i=1}^{\frac{n-1}{2}} p_i - \sum_{i=\frac{n+3}{2}}^{n} p_i, & \text{if } n \text{ is odd} \end{cases}$$

assuming the elements of $p$ have been permuted so that $p_1 \geq \cdots \geq p_n \geq 0$. For the remainder of this subsection we define the notation $k = n/2$ for even $n$, and $k = (n-1)/2$ for odd $n$.

**Theorem 3.7.4** (p 344 in (99))**.** *For doubly stochastic $P \in \mathbb{R}^{n\times n}$,*

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

*Proof.* From the explicit form for $\tau_1(P)$, we see that when $\tau_1(P) = 1$ there exists some $i, j$ so that $p_{il} = 0$ or $p_{jl} = 0$ for every $l = 1, \ldots, n$. This means there is at least one row in $P$ with at least $k$ zeros for $n$ even (or $k+1$ zeros for $n$ odd). Then, there is at least one column in $P^T$ with at least $k$ zeros for even $n$ (or $k+1$ zeros for $n$ odd).

We now use the fact that $P^T$ has at least one column with at least $k$ zeros for even $n$ (or $k+1$ zeros for $n$ odd). The function $\phi(p)$ is maximized when at least half the

elements of a column of $P$ are 0's. This is true since $\phi(p)$ is computed by subtracting the sum of the smallest $k$ values from the sum of the largest $k$. When the smallest $k$ values are 0, then $\phi(p) = \sum_{i=1}^{k} p_i = \sum_{i=1}^{n} p_i$. Since $P^T e = e$, the sum $\sum_{i=1}^{n} p_i = 1$, and $\tau_\infty\left(P^T\right) = \max_p \phi(p) = 1$. □

From $\tau_1(P) = 1$ we learn that there is at least one row in $P$ with at least $k$ zeros for $n$ even. In the following corollary, we make no assumptions on $\tau_1(P)$, but instead assume some row of $P$ has the desired property of at least $k$ zeros for $n$ even.

**Corollary 3.7.5** (p 344 in (99))**.** *If a doubly stochastic matrix $P \in \mathbb{R}^{n\times n}$ has at least one row with $k$ zeros when $n$ even (or $k+1$ zeros when $n$ is odd), then $\tau_\infty\left(P^T\right) = 1$.*

This statement for $\tau_\infty(P)$ was already proven in the proof of Theorem 3.7.4. We present a simple example to show that $\tau_1(P) < \tau_\infty\left(P^T\right) = 1$ is possible. Hence the converse of Theorem 3.7.4, $\tau_\infty\left(P^T\right) = 1 \implies \tau_1(P) = 1$, does not hold. Let

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

Then $\tau_1(P) = 5/6 < 1 = \tau_\infty\left(P^T\right)$.

A special class of doubly stochastic matrix arises in the study of the spectral theory of doubly stochastic matrices (83), in which every row contains the same elements. The elements are not required to be in the same order in each row/column. For matrices of this form, we show that $\tau_1(P) \leq \tau_\infty\left(P^T\right)$.

**Theorem 3.7.6** (p 345 in (99))**.** *Let $P \in \mathbb{R}^{n\times n}$ be a doubly stochastic matrix in which each row and each column of $P$ have some permutation of the same elements, labeled $p_1 \geq \cdots \geq p_n \geq 0$. Then $\tau_1(P) \leq \tau_\infty\left(P^T\right)$.*

*Proof.* We determine the explicit expressions for $\tau_\infty\left(P^T\right)$ and $\tau_1(P)$ for this particular matrix $P$, and exploit the fact that the rows and columns of $P$ contain the same elements.

We start with the expression for $\tau_\infty\left(P^T\right)$. Let $p$ be a row of $P$. If $n$ is even and $k = \frac{n}{2}$, then $\sum_{i=1}^n p_i = 1$ implies

$$\sum_{i=1}^{k} p_i = 1 - \sum_{i=k+1}^{n} p_i,$$

and

$$\phi(p) = \sum_{i=1}^{k} p_i - \sum_{i=k+1}^{n} p_i = 1 - 2\sum_{i=k+1}^{n} p_i.$$

Since all rows contain the same elements, $\tau_\infty\left(P^T\right) = 1 - 2\sum_{i=k+1}^n p_i$. Similarly, if $n$ is odd and $k = (n+1)/2$, then $\sum_{i=1}^k p_i = 1 - \sum_{i=k+2}^n p_i - p_{k+1}$ implies $\tau_\infty\left(P^T\right) = 1 - 2\sum_{i=k+2}^n p_i - p_{k+1}$. Thus,

$$\tau_\infty\left(P^T\right) = \begin{cases} 1 - 2\sum_{i=k+1}^n p_i, & \text{if } n \text{ is even,} \\ 1 - 2\sum_{i=k+2}^n p_i - p_{k+1}, & \text{if } n \text{ is odd.} \end{cases}$$

With regard to the expression for $\tau_1(P)$, $p_1 \geq \cdots \geq p_n \geq 0$ implies $\sum_{i=1}^k p_i \geq \sum_{i=k+1}^n p_i$ for $n$ even, and $\sum_{i=1}^k p_i + p_{k+1} \geq \sum_{i=k+2}^n p_i$ for $n$ odd. Thus, for the elements

$p_{ij}$ of $P$,

$$\sum_{l=1}^{n} \min\{p_{il}, p_{jl}\} \geq \begin{cases} 2\sum_{i=k+1}^{n} p_i, & \text{if } n \text{ is even,} \\ 2\sum_{i=k+2}^{n} p_i + p_{k+1}, & \text{if } n \text{ is odd.} \end{cases}$$

Hence,

$$\begin{aligned} \tau_1(P) &= 1 - \min_{ij} \sum_{l=1}^{n} \min\{p_{il}, p_{jl}\} \\ &\leq \begin{cases} 1 - 2\sum_{i=k+1}^{n} p_i, & \text{if } n \text{ is even,} \\ 1 - 2\sum_{i=k+2}^{n} p_i - p_{k+1}, & \text{if } n \text{ is odd.} \end{cases} \\ &= \tau_\infty\left(P^T\right) \end{aligned}$$

□

A corollary follows immediately.

**Corollary 3.7.7** (p 345 in (99))**.** *Let $P \in \mathbb{R}^{n\times n}$ be a symmetric stochastic matrix in which each row of $P$ contains the same elements, labeled $p_1 \geq \cdots \geq p_n \geq 0$. Then*

$$\tau_1(P) \leq \tau_\infty(P) \qquad \textit{and} \qquad \tau_1\left(P^T\right) \leq \tau_\infty\left(P^T\right).$$

*Proof.* If $P$ is symmetric, then $P = P^T$, and Theorem 3.7.6

$$\tau_1(P) \leq \tau_\infty\left(P^T\right) = \tau_\infty(P)$$

and

$$\tau_1\left(P^T\right) = \tau_1(P) \leq \tau_\infty\left(P^T\right).$$

□

The question remains whether we can find a symmetric stochastic $P$ with $\tau_1(P) < \tau_\infty(P)$, when each row of $P$ contains the same elements (99, p 345-346). For $n = 2, 3, 4$, $\tau_1(P) = \tau_\infty(P)$, which can be verified directly. We have already proven $\tau_1(P) = \tau_\infty(P)$ for all stochastic $2 \times 2$ matrices (see Theorem 3.5.19). We now show that when $n = 5$, $\tau_1(P) < \tau_\infty(P)$. Let

$$P = \begin{pmatrix} p_1 & p_2 & p_3 & p_4 & p_5 \\ p_2 & p_3 & p_4 & p_5 & p_1 \\ p_3 & p_4 & p_5 & p_1 & p_2 \\ p_4 & p_5 & p_1 & p_2 & p_3 \\ p_5 & p_1 & p_2 & p_3 & p_4 \end{pmatrix}$$

where $p_1 > \cdots > p_5$. Then the explicit form for $\tau_\infty\left(P^T\right) = \tau_\infty(P)$ in the proof of Corollary 3.7.5 gives, $\tau_\infty(P) = 1 - 2p_4 - 2p_5 - p_3$. The explicit form for $\tau_1(P)$ in the proof of Corollary 3.7.5 implies, $\tau_1(P) < 1 - 2p_4 - 2p_5 - p_3 = \tau_\infty(P)$. Thus, there exist symmetric stochastic matrices $P \in \mathbb{R}^{n\times n}$ for which $\tau_1(P) < \tau_\infty(P)$, so long as $n \geq 5$.

In 1984 Seneta and Tan (106) determined $\tau_2(P)$ for doubly stochastic $P$. In this theorem we present explicit forms for $\tau_2(P)$ and $\tau_2\left(P^T\right)$.

**Theorem 3.7.8** (p 3 in (106))**.** *Let $P \in \mathbb{R}^{n\times n}$ be a doubly stochastic matrix. Then*

$$\tau_2(P) = \tau_2\left(P^T\right) = \sqrt{\lambda_2\left(PP^T\right)},$$

*where* $\lambda_2\left(PP^T\right)$ *is the second largest eigenvalue of* $PP^T$.

*Proof.* We argue the case for $\tau_2(P)$ only. Since $P$ and $P^T$ are square, $\lambda_i\left(PP^T\right) = \lambda_i\left(P^TP\right)$.

The expression for the ergodicity coefficient can be written as

$$\tau_2(P) = \max_{\substack{\|z\|_2=1 \\ z^Te=0}} \left\|P^Tz\right\|_2 = \max_{\substack{z^Tz=1 \\ z^Te=0}} \sqrt{z^TPP^Tz}.$$

Choose normalized $y$ as the eigenvector for $\lambda_2\left(PP^T\right)$. Then $PP^Ty = \lambda_2\left(PP^T\right)y$ and $y^Ty = 1$. If $\lambda_1\left(PP^T\right) > \lambda_2\left(PP^T\right)$, then $y^Te = 0$. Then

$$y^TPP^Ty = \lambda_2\left(PP^T\right)y^Ty = \lambda_2\left(PP^T\right)$$

and

$$\tau_2(P) = \sqrt{\lambda_2\left(PP^T\right)}.$$

□

Clearly, $\tau_2(P)$ and $\tau_2\left(P^T\right)$ are both less than or equal to 1. If $PP^T$ is primitive (76, p 674), then $\tau_2(P) < 1$ because the second largest eigenvalue of $PP^T$ is less than 1. If $P > 0$, then $PP^T > 0$ and $P^TP > 0$, and both $\tau_2(P)$ and $\tau_2\left(P^T\right)$ are $< 1$ from Perron-Frobenius theory (76, p 667).

### 3.7.3 Nonnegative, irreducible matrices

In this section we present results of Rothblum and Tan (91) that relate ergodicity coefficients to subdominant eigenvalues. We also present bounds for ergodicity coefficients by Gross and Rothblum (38) that are based on approximations of the dominant eigenvalue and eigenvector.

For nonnegative irreducible matrices Rothblum and Tan (91) bound the absolute value of the subdominant eigenvalues. Rothblum and Tan stray from convention and define an ergodicity coefficient (91, p 57) as the maximum modulus of the subdominant eigenvalues. They also define two ergodicity coefficients, one that operates on real vectors, and one that operates on complex vectors. We introduce the following notation, which differs from that of the original authors (91, p 59). For nonnegative irreducible $A \in \mathbb{R}^{n \times n}$ and dominant right eigenvector $w$ of $A$,

$$\tau_p^R(w, A) = \max_{\substack{\|z\|_p = 1 \\ z^T w = 0}} \left\| A^T z \right\|_p, \text{ where } z \in \mathbb{R}^n$$

and

$$\tau_p(w, A) = \max_{\substack{\|z\|_p = 1 \\ z^T w = 0}} \left\| A^T z \right\|_p, \text{ where } z \in \mathbb{C}^n$$

Obviously $\tau_p^R(w, A) \leq \tau_p(w, A)$ since $\mathbb{R}^n \subset \mathbb{C}^n$. We also note here, that unlike other ergodicity coefficients for general matrices (see Definition 3.6.1), Rothblum and Tan require the vector in $\tau_p(w, A)$ and $\tau_p^R(w, A)$ to be the positive dominant right eigenvector of the nonnegative, irreducible matrix $A$.

Rothblum and Tan show that the ergodicity coefficient over all real vectors bounds

the modulus of both real and complex subdominant eigenvalues.

**Theorem 3.7.9** (Theorem 3.1 (91)). *Let $A \in \mathbb{R}^{n\times n}$ be nonnegative irreducible matrix with dominant eigenvalue $\lambda_1$, and dominant right eigenvector $w$. For all eigenvalues $\lambda \neq \lambda_1$ of $A$,*

$$|\lambda| \leq \tau_p^R(w, A).$$

This is proven by Bauer, et al (6, §5.2) as well as Rothblum and Tan (91), all of whom construct a seminorm that can represent a complex eigenvector as a real vector.

Rothblum and Tan (91) exploited a known norm relation to develop a computable upper bound on the modulus of all eigenvalues of a matrix, based on the Frobenius norm $\|B\|_F^2 = \sum_{ij} |b_{ij}|^2$ for any $B \in \mathbb{C}^{m\times n}$, since $\|B\|_F$ is easy to compute.

**Lemma 3.7.10** (Lemma 5.4 in (91)). *For any matrix $B \in \mathbb{R}^{n\times n}$,*

$$|\lambda| \leq \|B\|_2 \leq \|B\|_F$$

*for all eigenvalues $\lambda$ of $B$.*

*Proof.* In subsection 3.9.2, Theorem 3.9.6, we state the well known theorem that $|\lambda| \leq \|B\|_p$ for all eigenvalues $\lambda$ of $B$, with $B \in \mathbb{C}^{n\times n}$. So the first inequality, $|\lambda| \leq \|B\|_2$ follows directly. It remains to prove the inequality $\|B\|_2 \leq \|B\|_F$.

With $b_i^T$ denoting the $i^{th}$ row of $B$, the Cauchy-Schwarz inequality implies

$$\|Bx\|_2^2 = \sum_{i=1}^{n} \left|b_i^T x\right|^2 \leq \sum_{i=1}^{n} \|b_i\|_2^2 \|x\|_2^2 = \|B\|_F^2 \|x\|_2^2.$$

Hence, $\dfrac{\|Bx\|_2}{\|x\|_2} \leq \|B\|_F$ for all $x \neq 0$. □

We now present computable upper bounds on the maximum modulus of subdominant eigenvalues using the ergodicity coefficient and the expression $|\lambda| \leq \|B\|_2 \leq \|B\|_F$ in Lemma 3.7.10.

**Theorem 3.7.11** (Theorem 5.5 in (91))**.** *Let $A \in \mathbb{R}^{n\times n}$ be a nonnegative, irreducible matrix with dominant eigenvalue $\lambda_1 > 0$ and dominant right eigenvector $w$. For all $x \in \mathbb{R}^n$, and all eigenvalues $\lambda \neq \lambda_1$ of $A$,*

$$|\lambda| \leq \tau_p^R(w, A) \leq \left\| \left(A - wx^T\right)^T \right\|_p \tag{3.16}$$

*and in particular for $p = 2$,*

$$|\lambda| \leq \tau_2^R(w, A) \leq \left\| \left(A - wx^T\right)^T \right\|_2 \leq \left\| \left(A - wx^T\right)^T \right\|_F . \tag{3.17}$$

*The right hand side of expression (3.17) attains a minimum when $x = A^T w/\|w\|_2^2$.*

*Proof.* We have already proven the first inequality in expression (3.16). It remains to prove the second inequality. If $z^T w = 0$, then $\left(A - wx^T\right)^T z = A^T z - xw^T z = A^T z$, and

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

Expression (3.17) follows from expression (3.16) and Lemma 3.7.10.

We still need to find the value of $x$ that minimizes $\left\|(A - wx^T)^T\right\|_F$. The Frobenius norm can be obtained by summing 2-norms of columns,

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

Thus, $\min_x \left\|\left(A - wx^T\right)^T\right\|_F^2$ corresponds to $n$ independent minimization problems $\|Ae_i - wx_i\|_2$. We can view $\|wx_i - Ae_i\|_2$ as a least squares problems with $n \times 1$ coefficient matrix $w$ of full column rank and right hand side $Ae_i$. The unique solution is

$$x_i = \left(w^T w\right)^{-1} w^T Ae_i = \frac{1}{\|w\|_2^2} e_i^T A^T w.$$

Thus $x = A^T w / \|w\|_2^2$. Therefore,

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

where $I - \frac{ww^T}{\|w\|_2^2}$ is the orthogonal projector onto the subspace orthogonal to $w$. □

We now turn our attention to the results of Gross and Rothblum (38), who are concerned with computable approximations to the normalized dominant eigenvector of a nonnegative, irreducible matrix. We begin with a relation between two ergodicity coefficients with different vectors.

**Lemma 3.7.12** (Lemma 4.1 in (38))**.** *Let $B \in \mathbb{R}^{n\times n}$ and $f \in \mathbb{R}^n$. For every pair of real vectors $u$ and $v$ such that $f^T u = f^T v = 1$,*

$$\tau_p^R(u, B) \leq \tau_p^R(v, B) + 2 \|B\|_q \|f\|_p \|v - u\|_q$$

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

*Proof.* There exists a $y$ such that $\tau_p^R(u, B) = \left\|B^T y\right\|_p$, with $\|y\|_p = 1$ and $y^T u = 0$. If $B^T y = 0$, then the expression in the lemma is trivial because $\tau_p^R(u, B) \geq 0$ and norms are nonnegative.

Suppose $y - v^T y f = 0$, then $y = v^T y f$ and $u^T y = v^T y u^T f = v^T y$ because $u^T f = 1$. Thus, $v^T y = 0$, and hence $y = 0$. But $\|y\|_p = 1$, and we have proved by contradiction that $y - v^T y f \neq 0$.

If $B^T y \neq 0$, we define the vector $z = \left(y - v^T y f\right) / \left\|y - v^T y f\right\|_p$ which satisfies $\|z\|_p = 1$ and $z^T v = 0$. Then

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

Because $y^T u = 0$, we can rewrite the above expression as

$$\tau_p^R(v, B) \geq \frac{\left\|B^T y - (v - u)^T y B^T f\right\|_p}{\left\|y - (v - u)^T y f\right\|_p}.$$

From the triangle inequality and the fact that the norm of a product is less than or equal to the product of the norms, we obtain

$$\tau_p^R(v, B) \geq \frac{\left\|B^T y\right\|_p - \left\|(v - u)^T\right\|_p \|y\|_p \left\|B^T\right\|_p \|f\|_p}{\|y\|_p + \left\|(v - u)^T\right\|_p \|y\|_p \|f\|_p}.$$

For any matrix $A \in \mathbb{R}^{n \times n}$, $\left\|B^T\right\|_p = \|B\|_q$ if $1/p + 1/q = 1$. We also know that

$\tau_p^R(u,B) = \left\|B^T y\right\|$ and $\|y\|_p = 1$, which allows us to write

$$\tau_p^R(v,B) \geq \frac{\tau_p^R(u,B) - \|v-u\|_q \|B\|_q \|f\|_p}{1 + \|v-u\|_q \|f\|_p}.$$

This expression can be rearranged as

$$\tau_p^R(u,B) \leq \tau_p^R(v,B)\left[1 + \|v-u\|_q \|f\|_p\right] + \|v-u\|_q \left\|B^T\right\|_p \|f\|_p,$$

from which it follows that

$$\tau_p^R(u,B) \leq \tau_p^R(v,B) + 2\|v-u\|_q \left\|B^T\right\|_p \|f\|_p$$

because Theorem 3.6.3 implies $\tau_p^R(v,B) \leq \left\|B^T\right\|_p$. □

We apply the results of this lemma, along with the eigenvalue bound in Corollary 3.6.8 to obtain a new upper bound on subdominant eigenvalues.

**Theorem 3.7.13** (Theorem 4.3 in (38))**.** *Let $A \in \mathbb{R}^{n\times n}$ be a nonnegative irreducible matrix with dominant eigenvalue $\lambda_1 > 0$ and dominant, positive eigenvector $w$. Choose $f \in \mathbb{R}^n$ so that $f^T w = 1$. Then for all eigenvalues $\lambda \neq \lambda_1$ of $A$,*

$$|\lambda| \leq \tau_p^R(w,A) \leq \tau_p^R(v,A) + 2\|A\|_q \|f\|_p \|v-w\|_q,$$

*for each $v \in \mathbb{R}^n$ satisfying $f^T v = 1$.*

The proof follows directly from Theorem 3.7.9, which states $|\lambda| \leq \tau_p^R(w,A)$ for all eigenvalues $\lambda \neq \lambda_1$ of $A$ when $w$ is a dominant eigenvector, and Lemma 3.7.12 which

states $\tau_p^R(w, A) \leq \tau_p^R(v, A) + 2 \|A\|_q \|f\|_p \|v - w\|_q$ when $f^T w = f^T v = 1$ for $f \in \mathbb{R}^n$.

## 3.8 Ergodicity coefficients with the dominant left eigenvector

In this section, we examine an ergodicity coefficient that varies from those presented previously. In 1983, Tan (110) defined a new ergodicity coefficient $\tau_p^R\left(\pi, S^T\right)$ where $\pi$ is a dominant left eigenvector of $S$. The following year, Tan (111) extended this ergodicity coefficient to nonnegative, irreducible matrices $A$.

We proceed with results derived by Tan in 1983 (110), whose purpose was to develop a "computable bound on the geometric convergence rate to ergodicity" of an ergodic stochastic matrix, although he also presented results for non-ergodic, or reducible stochastic matrices.

**Definition 3.8.1** (p 278 in (110))**.** *A stochastic matrix $S$ is ergodic if $\lim_{k\to\infty} S^k = e\pi^T$, where $\pi$ is the stationary distribution of $S$.*

We now state Tan's ergodicity coefficient, defined to operate on a dominant left eigenvector of a stochastic matrix $S$.

**Definition 3.8.2** (p 278 in (110))**.** *For a stochastic matrix $S \in \mathbb{R}^{n\times n}$,*

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

*where $\pi$ is a dominant left eigenvector of $S$, and the maximization ranges over $z \in \mathbb{R}^n$.*

When $S$ is an ergodic (irreducible) matrix, all elements of $\pi$ are positive.

Like previous coefficients, it is straightforward to argue that $\tau_p^R\left(\pi, S^T\right)$ is continuous and bounded. We also present several additional results similar to those proven for $\tau_p^R(S)$ in Theorem 3.5.18.

**Theorem 3.8.3** (p 279 in (110))**.** *For stochastic matrices* $S, S_1, S_2$

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

The proof of these items are identical to previous proofs. Since $\tau_p^R\left(\pi, S^T\right)$ is just a special case of $\tau_p(w, A)$ we also know that the results from section 3.6 hold.

In addition to these properties, Tan focused on computation of explicit forms for stochastic $S \in \mathbb{R}^{n \times n}$ in the cases of the 1-, 2-, and $\infty$-norms. When $S$ is not ergodic, then $S$ is reducible and can be permuted into a specific block form.

**Lemma 3.8.4** (p 223 in (9))**.** *If stochastic* $S \in \mathbb{R}^{n \times n}$ *is reducible, then there exist a permutation matrix* $P$*, so that* $S$ *can be permuted to*

$$PSP^T = \begin{pmatrix} A & 0 \\ C & B \end{pmatrix} \tag{3.18}$$

*where $A \in \mathbb{R}^{k\times k}$, $k < n$, is irreducible and a block diagonal matrix with square matrices on the diagonal, $B \in \mathbb{R}^{(n-k)\times(n-k)}$, and each row of $C \in \mathbb{R}^{(n-k)\times k}$ has a positive entry.*

The matrix $PSP^T$ is a Markov chain, whose first $k$ states are recurrent, and last $n-k$ states are transient.

We now present two lemmas for the values of the stationary distribution when $S$ is in the form of (3.18).

**Lemma 3.8.5** (Lemma (3.20), p 223 in (9))**.** *When a reducible stochastic matrix is in the form of (3.18), then $|\lambda(B)| < 1$, for all eigenvalues $\lambda(B)$ of $B$.*

*Proof.* The eigenvalue of $S$ and $PSP^T$ are the same. Additionally, since $A$ and $B$ are square and on the diagonal of $PSP^T$, the combined eigenvalues of $A$ and $B$ are the eigenvalues of $S$. The matrix $S$ being a stochastic matrix implies that the elements of $B$ are nonnegative and the row sums are less than or equal to one. However, because each row of $C$ has a positive entry, then each row of $B$ has a row sums less than 1. Thus $|\lambda(B)| < 1$, for all eigenvalues $\lambda(B)$ of $B$. □

**Lemma 3.8.6.** *Let $S$ be an $n \times n$ reducible stochastic matrix, permuted as in (3.18). Then the stationary distribution of $S$ equals*

$$\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} = \begin{pmatrix} \pi_1 \\ 0 \end{pmatrix},$$

*for $\pi_1 \in \mathbb{R}^k$, with $\pi_i > 0$ for $i = 1, \ldots, k$.*

*Proof.* Since $S$ is stochastic, $Se = e$, or

$$\begin{pmatrix} A & 0 \\ C & B \end{pmatrix} \begin{pmatrix} e \\ e \end{pmatrix} = e.$$

Hence $Ae = e$, so clearly $A$ is a stochastic matrix. Since $A$ is irreducible, Perron-Frobenius implies $A$ has a dominant positive left eigenvector $\pi_1^T$.

Additionally from $\pi^T S = \pi^T$, we obtain

$$\begin{pmatrix} \pi_1^T & \pi_2^T \end{pmatrix} \begin{pmatrix} A & 0 \\ C & B \end{pmatrix} = \begin{pmatrix} \pi_1^T & \pi_2^T \end{pmatrix}$$

which yields $\pi_2^T B = \pi_2^T$. Since the dominant eigenvalue of $B$ is strictly less than 1 from Lemma 3.8.5, then $\pi_2^T B = \pi_2^T$ can only hold if $\pi_2^T = 0$. Then

$$\pi = \begin{pmatrix} \pi_1 \\ 0 \end{pmatrix}.$$

□

Similar to the results in section 3.6.2, we will present bound for $\tau_1^R$, $\tau_\infty^R$, and finally $\tau_2^R$. We begin with an explicit form for $\tau_1^R\left(\pi, S^T\right)$ when $S$ is irreducible. This theorem is a generalization of Theorem 3.6.12 with $A = S^T$ and $w = \pi$.

**Theorem 3.8.7** (Remark p 284 in (110))**.** *If $S$ is irreducible with dominant left eigen-*

*vector $\pi$, such that $\pi_i > 0$ for $i = 1, \ldots, n$, then*

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

*where $D$ is a diagonal matrix with $D(i,i) = \pi_i$ for $i = 1, \ldots, n$.*

The situation where $S$ is reducible and of the form in (3.18) is analogous to Corollary 3.6.13 since any dominant left eigenvector of $S$ has at least one zero entry.

**Corollary 3.8.8** (Theorem 3 in (110))**.** *Let $S \in \mathbb{R}^{n \times n}$ be a stochastic reducible matrix as in (3.18), with stationary distribution $\pi^T = \begin{pmatrix} \pi_1^T \\ 0 \end{pmatrix}$. Then*

$$\tau_1^R\left(\pi, S^T\right) = \max\left\{\tau_1^R\left(\pi_1, A^T\right), \|B\|_1\right\}.$$

We now present an explicit form for $\tau_\infty^R\left(\pi, S^T\right)$. This theorem is a special case of Theorem 3.6.14 again with $A = S^T$ and $w = \pi$.

**Theorem 3.8.9** (Theorem 1 in (110))**.** *Let $S \in \mathbb{R}^{n \times n}$ be a stochastic, reducible matrix as in (3.18) with stationary distribution $\pi \in \mathbb{R}^n$. Then*

$$\tau_\infty^R\left(\pi, S^T\right) = \max_{\substack{\|z\|_\infty = 1 \\ \pi^T z = 0}} \|Sz\|_\infty = \max_i \max_{\substack{\|z\|_\infty = 1 \\ \pi^T z = 0}} \left|e_i^T S z\right|,$$

*where the maximum ranges over $z \in \mathbb{R}^n$. Let $s^T = e_i^T S$ be a row of $S$, so that $s_j$ is the $j^{th}$ element of $e_i^T S$. We assume the rows (and columns) of $S$ and $\pi$ have been permuted, so that $s_1/\pi_1 \leq \cdots \leq s_k/\pi_k$. From Lemma 3.8.6 it follows that $\pi_i > 0$ for*

$i = 1, \ldots, k$. *Define* $l$ *as the smallest integer so that* $\sum_{i=1}^{l} \pi_i \geq \frac{1}{2}$. *Then*

$$z_i = \begin{cases} -1, & \text{for } i = 1, \ldots, l-1 \\ \frac{1}{\pi_l} \left( \sum_{j=1}^{l-1} \pi_j - \sum_{j=l+1}^{k} \pi_j \right), & \text{for } i = l \\ 1, & \text{for } i = l+1, \ldots, k \\ \text{sign } s_i, & \text{for } i = k+1, \ldots, n. \end{cases}$$

When the matrix is irreducible, the above theorem applies with $k = n$.

## 3.9 Applications

In the following subsections, ergodicity coefficients are utilized as condition numbers, eigenvalue and singular value bounds, and in the case of specific eigenvalue bounds, Lehmann bounds.

### 3.9.1 Condition Numbers

In 2001, Cho and Meyer (22) compared eight different condition numbers for the stationary distribution of a stochastic matrix, two of which involved an ergodicity coefficient. More recently, Kirkland, et al., (64) expand on the analysis of condition numbers by Cho and Meyer (22), and introduce a new form of an ergodicity coefficient.

We begin with a result from Seneta (101) that illustrates how the ergodicity coefficient of a stochastic matrix can bound the condition number for the stationary distribution of stochastic linear systems $\pi^T S = \pi^T$. This theorem shows that the normwise

absolute error in the perturbed solution is bounded by $\|E\|_\infty$ when $\tau_1(S) \ll 1$.

**Theorem 3.9.1** (§2 in (101))**.** *Let $S$ and $S+E$ be stochastic, irreducible matrices with $\pi^T S = \pi^T$, $\widehat{\pi}^T(S+E) = \widehat{\pi}^T$ and $\pi^T e = \widehat{\pi}^T e = 1$. If $\tau_1(S) < 1$, then*

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

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

$$(\widehat{\pi} - \pi)^T (I - S) = \widehat{\pi}^T E. \tag{3.19}$$

Because $S$ is irreducible and $\tau_1(S) < 1$, the dominant eigenvalue 1 is simple and we can write $S = e\pi^T + Q$, where the magnitude of the eigenvalues of $Q$ is less than 1. Substituting this expression for $S$ in (3.19) and using $\pi^T e = \widehat{\pi}^T e = 1$ 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 we can solve for $(\widehat{\pi} - \pi)^T$,

$$(\widehat{\pi} - \pi)^T = \widehat{\pi}^T E (I - Q)^{-1}.$$

Then write $(I-Q)^{-1} = \sum_{i=0}^{\infty} Q^i$ (76, p 126), and let $y^T = \widehat{\pi}^T E$. Transposing both sides, then taking norms, we obtain

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

We now need to relate the above expression to $\tau_1(Q)$. From the fact that $S$ and $S+E$ are stochastic, we know that $Se = e$ and $(S+E)e = e$, so $Ee = 0$. Thus $y^T e = 0$, and

$$\|\widehat{\pi} - \pi\|_1 \leq \sum_{i=0}^{\infty} \left\| \left(Q^i\right)^T y \right\|_1 \leq \|y\|_1 \max_{y^T e = 0} \sum_{i=0}^{\infty} \frac{\left\| \left(Q^i\right)^T y \right\|_1}{\|y\|_1} = \|y\|_1 \sum_{i=0}^{\infty} \tau_1\left(Q^i\right)$$

The submultiplicative property, #5, of Theorem 3.5.18 implies

$$\sum_{i=0}^{\infty} \tau_1\left(Q^i\right) \leq \sum_{i=0}^{\infty} \left[\tau_1\left(Q\right)\right]^i.$$

Then

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

Since $y^T e = 0$ we can obtain $y^T Q = y^T S$, and $\tau_1(Q)$ can be replaced by $\tau_1(S)$. Finally, bound $\|y\|_1 = \left\|E^T \widehat{\pi}\right\|_1 \leq \|E\|_\infty \|\widehat{\pi}\|_1 = \|E\|_\infty$. The fact that $\|\widehat{\pi}\|_1 = 1$ comes from the fact that $\widehat{\pi}$ is the dominant eigenvector of an irreducible matrix, so $\widehat{\pi}_i > 0$ for all $i$, and by assumption, $\widehat{\pi}^T e = \|\widehat{\pi}\|_1$. We obtain the result

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

□

In order to discuss the next form of an ergodicity coefficient as a condition number, we define a new ergodicity coefficient. The coefficient, $\tau$ has two subscripts that denote the two different norms being used.

**Definition 3.9.2** ((1.10) in (64)). *For any $1 \leq p \leq \infty$ and $A \in \mathbb{R}^{n\times n}$ define*

$$\tau_{p,1}^{R}(e, A) = \max_{z^T e=0} \frac{\left\|A^T z\right\|_p}{\|z\|_1}$$

*where the maximization ranges over $z \in \mathbb{R}^n$.*

We also need to recall the definition of a group inverse of a matrix.

**Definition 3.9.3** (p 641 in (76)). *The group inverse of a matrix $A \in \mathbb{R}^{n\times n}$, if it exists, is the unique matrix denoted $A^{\#}$ that satisfies*

$$AA^{\#}A = A, \qquad A^{\#}AA^{\#} = A^{\#}, \qquad AA^{\#} = A^{\#}A.$$

*The group inverse of a matrix exists if and only if its range and nullspaces are complimentary subspaces in $\mathbb{R}^n$.*

We now show that the new coefficient defined in Definition 3.9.2 is also a condition number.

**Theorem 3.9.4** (Proposition 2.1 in (64)). *Let $S$ and $\widehat{S} = S + E$ be stochastic, irreducible matrices with $\pi^T S = \pi^T$ and $\widehat{\pi}^T(S + E) = \widehat{\pi}^T$, and $\pi^T e = \widehat{pi}^T e = 1$. If $A = I - S$, then*

$$\|\pi - \widehat{\pi}\|_p \leq \tau_{p,1}^{R}\left(e, A^{\#}\right) \|E\|_\infty$$

*where $A^{\#} = \left(I - S + e\pi^T\right)^{-1} - e\pi^T$ denotes the group inverse of $A$ from Definition 3.9.3.*

*Proof.* Since $A = I - S$ is an $M$-matrix, we know its generalized inverse exists. It is easy to verify that $A^{\#} = \left(I - S + e\pi^T\right)^{-1} - e\pi^T$.

From $\pi^T S = \pi^T$ and $\widehat{\pi}^T(S+E) = \widehat{pi}^T$ we write

$$\pi^T - \widehat{pi}^T = \pi^T S - \widehat{pi}^T(S+E) = \left(\pi^T - \widehat{pi}^T\right) S - \widehat{pi}^T E.$$

The above expression for $\pi^T - \widehat{pi}^T$, combined with $A = I - S$ gives

$$\left(\pi^T - \widehat{pi}^T\right) A = -\widehat{pi}^T E.$$

Multiplying by the generalized inverse $A^{\#} = \left(I - S + e\pi^T\right)^{-1} + e\pi^T$ and taking transposes of both sides, yields

$$\pi - \widehat{pi} = -\left(A^{\#}\right)^T E^T \widehat{pi},$$

because $\pi^T e = \widehat{pi}^T e = 1$. Taking the $p$-norm of both sides, $\left\|\pi - \widehat{pi}\right\|_p = \left\|\left(A^{\#}\right)^T E^T \widehat{pi}\right\|_p$. If $y = E^T\widehat{pi}$, then $y^T e = 0$ because $Ee = \left(\widehat{S} - S\right)e = 0$ since $S$ and $\widehat{S}$ are stochastic. Then because $y^T e = 0$

$$\left\|\pi - \widehat{pi}\right\|_p = \|y\|_1 \frac{\left\|\left(A^{\#}\right)^T y\right\|_p}{\|y\|_1} \leq \|y\|_1 \max_{z^T e = 0} \frac{\left\|\left(A^{\#}\right)^T z\right\|_p}{\|z\|_1} = \|y\|_1 \tau_{p,1}\left(e, A^{\#}\right)$$

As a final step,

$$\|y\|_1 = \left\|E^T\widehat{pi}\right\|_1 \leq \left\|E^T\right\|_1 \left\|\widehat{pi}\right\|_1 = \|E\|_\infty$$

because $\widehat{pi}^T e = 1$ and since $\widehat{pi}_i \geq 0$ for all $i$, then $\left\|\widehat{pi}\right\|_1 = 1$. Thus, we have acheived the inequality as stated in the theorem,

$$\left\|\pi - \widehat{pi}\right\|_p \leq \tau_{p,1}^R\left(e, A^{\#}\right)\|E\|_\infty.$$

□

Seneta (102) presents a third example where ergodicity coefficients serve as condition numbers. His result follows directly from result by Kirkland, et al., in Theorem 3.9.4 with $p = 1$. Notice when $p = 1$, the new ergodicity coefficient $\tau_{p,1}^{R}\left(e, A^{\#}\right)$ reduces to $\tau_1\left(A^{\#}\right)$.

**Corollary 3.9.5** (Theorem 2 in (102)). *Let $S$ and $\widehat{S} = S + E$ be stochastic, irreducible matrices with $\pi^T S = \pi^T$ and $\widehat{\pi}^T(S + E) = \widehat{pi}^T$, and $\pi^T e = \widehat{pi}^T e = 1$. If $A = I - S$, then*

$$\left\|\pi - \widehat{\pi}\right\|_1 \leq \tau_1\left(A^{\#}\right)\|E\|_\infty$$

*where $A^{\#} = \left(I - S + e\pi^T\right)^{-1} - e\pi^T$ denotes the group inverse of $A$ from Definition 3.9.3.*

### 3.9.2 Eigenvalue bounds

We present new and existing eigenvalue bounds and show their relationship to the ergodicity coefficient. We begin with the well-known statement that the eigenvalues of a matrix are bounded above by the norm of the matrix.

**Theorem 3.9.6** (p 497 in (76)). *For $A \in \mathbb{C}^{n\times n}$ we have*

$$|\lambda| \leq \|A\|_p, \text{ for all eigenvalues } \lambda \text{ of } A.$$

**Lemma 3.9.7** (p 518 in (76)). *For $A \in \mathbb{C}^{n\times n}$ with eigenvalues $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$, choose vectors $w$ and $v$ so that $v^* w = 1$ where $w$ and $v$ are right and left eigenvectors*

*of the dominant eigenvalue of $A$, $\lambda_1$. Then $wv^*$ is a spectral projector and $A - wv^*$ has dominant eigenvalue $\lambda_2$.*

We now present an eigenvalue bound related to ergodicity coefficients, that makes use of Theorem 3.9.6 and Lemma 3.9.7.

**Theorem 3.9.8.** *Let $A \in \mathbb{C}^{n\times n}$ have eigenvalues $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Let $w$ and $v$ be right and left eigenvectors of the dominant eigenvalue $\lambda_1$, chosen so that $v^*w = 1$. Then*

$$\tau_p(w, A) \leq \|A^* - vw^*\|_p .$$

*Proof.* Recall $\tau_p(w, A) = \max_{\|z\|_p=1, w^*z=0} \|A^*z\|_p$ for $z \in \mathbb{C}^n$. From $w^*z = 0$ we get $(A - wv^*)^* z = A^*z - vw^*z = A^*z$ and

$$\tau_p(w, A) = \max_{\substack{\|z\|_p=1 \\ w^*z=0}} \|(A - wv^*)^* z\|_p = \tau_p (u, A - wv^*)$$

and from Theorem 3.6.3 we know $\tau_p (w, A - wv^*) \leq \|(A - wv^*)^*\|_p$. □

This result is shown for $p = 1, \infty$ in (100). From Lemma 3.9.7 we know that $A - wv^*$ has dominant eigenvalue $\lambda_2$, and from Theorem 3.9.6 we know that $\|A - wv^*\|_p$ is an upper bound on all eigenvalues of $A - uv^*$, including the dominant eigenvalue $\lambda_2$.

A similar result for stochastic matrices follows directly.

**Corollary 3.9.9.** *For stochastic $S \in \mathbb{R}^{n\times n}$ with dominant right eigenvector $e$ and dominant left eigenvector (stationary distribution) $\pi^T$, we have*

$$\tau_p(S) \leq \left\|S^T - \pi e^T\right\|_p .$$

By assuming certain properties on $A$, we now present upper and lower bounds and equivalence statements for $\tau_p(W, A)$ when $W$ spans an invariant subspace associated with the largest eigenvalues in magnitude. We begin with an upper and lower bound on the ergodicity coefficient of a diagonalizable matrix.

**Theorem 3.9.10.** *Let $A \in \mathbb{C}^{n\times n}$ be a diagonalizable matrix and define $W \in \mathbb{C}^{n\times k}$ to have columns spanning an invariant subspace associated with eigenvalues of $A$, when $A$ has eigenvalues $|\lambda_1| \geq \cdots \geq |\lambda_k| > |\lambda_{k+1}| \geq \cdots |\lambda_n|$. Then*

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

*where $A = X\Lambda X^{-1}$ with $X = \begin{pmatrix} W & X_1 \end{pmatrix}$, $X^{-1} = \begin{pmatrix} V \\ X_2 \end{pmatrix}$. Since $X \in \mathbb{C}^{n\times n}$ and $W \in \mathbb{C}^{n\times k}$, then $X_1 \in \mathbb{C}^{n\times(n-k)}$, $V \in \mathbb{C}^{k\times n}$ and $X_2 \in \mathbb{C}^{(n-k)\times n}$. Additionally, $\Lambda = \begin{pmatrix} \Lambda_1 & \\ & \Lambda_2 \end{pmatrix}$, where $\Lambda_1 = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{pmatrix} \in \mathbb{C}^{k\times k}$ and $\Lambda_2 = \begin{pmatrix} \lambda_{k+1} & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \in \mathbb{C}^{(n-k)\times(n-k)}$.*

*Proof.* The lower bound was proven in Theorem 3.6.7. It remains to prove the upper bound.

Choose $y$ to achieve the maximum of $\tau_p(W, A)$, so that $\tau_p(W, A) = \left\|A^*y\right\|_p$ with $y^*W = 0$ and $\|y\|_p = 1$. Then

$$A^*y = X^{-*}\Lambda^*X^*y = \begin{pmatrix} V^* & X_2^* \end{pmatrix} \begin{pmatrix} \Lambda_1^* & \\ & \Lambda_2^* \end{pmatrix} \begin{pmatrix} W^* \\ X_1^* \end{pmatrix} y.$$

However, since $W^*y = 0$,

$$A^*y = \begin{pmatrix} 0 & X_2^* \end{pmatrix} \begin{pmatrix} 0 & \\ & \Lambda_2^* \end{pmatrix} \begin{pmatrix} 0 \\ X_1^* \end{pmatrix} y.$$

Then

$$\tau_p(W, A) \leq \|A^*y\|_p \leq \|X_2^*\|_p \|\Lambda_2^*\|_p \|X_1^*\|_p \|y\|_p = |\lambda_{k+1}| \|X_1^*\|_p \|X_2^*\|_p$$

because the 2-norm of a diagonal matrix is the magnitude of the largest value, and we defined $\|y\|_p = 1$. □

We now present a corollary to the previous result in which $A$ is assumed normal.

**Corollary 3.9.11.** *Let $A \in \mathbb{C}^{n\times n}$ be a normal matrix and define $W \in \mathbb{C}^{n\times k}$ to have normalized columns spanning an invariant subspace associated with eigenvalues of $A$, when $A$ has eigenvalues $|\lambda_1| \geq \cdots \geq |\lambda_k| > |\lambda_{k+1}| \geq \cdots |\lambda_n|$. Then*

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

*Proof.* From Theorem 3.9.10 we know that

$$|\lambda_{k+1}| \leq \tau_2(W, A) \leq |\lambda_{k+1}| \|X_1^*\|_2 \|X_2^*\|_2,$$

where $A = X\Lambda X^{-1}$ with $X = \begin{pmatrix} W & X_1 \end{pmatrix}$, $X^{-1} = \begin{pmatrix} V \\ X_2 \end{pmatrix}$.

Since $A$ is a normal matrix, it is unitarily similar to a diagonal matrix (76, p 547). Then $X^* = X^{-1}$ and $\|X\|_2 = \|X^*\|_2 = 1$. Thus, $\|X_1^*\|_2 = \|X_2^*\|_2 = 1$ and

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

□

We now prove an upper bound on the ergodicity coefficient using the Schur decomposition of a matrix.

**Theorem 3.9.12.** *Let $A \in \mathbb{C}^{n\times n}$ have eigenvalues $|\lambda_1| \geq \cdots \geq |\lambda_k| > |\lambda_{k+1}| \geq \cdots |\lambda_n|$ and let $W \in \mathbb{C}^{n\times k}$ have normalized columns spanning an invariant subspace associated with eigenvalues of $A$. Then*

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

*where $N$ is the departure from normality for $A$.*

*Proof.* The lower bound was proven in Theorem 3.6.7. It remains to prove the upper bound.

To prove the upper bound, we need the Schur decomposition $A = VTV^*$ where $V$ is a unitary matrix, written $V^* = \begin{pmatrix} W^* \\ V_2 \end{pmatrix}$ and $T = \Lambda + N$ is an upper triangular matrix with $\Lambda = \begin{pmatrix} \Lambda_1 & \\ & \Lambda_2 \end{pmatrix}$, where $\Lambda_1 = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{pmatrix} \in \mathbb{C}^{k\times k}$ and $\Lambda_2 =$

$\begin{pmatrix} \lambda_{k+1} & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \in \mathbb{C}^{(n-k)\times(n-k)}$ and strictly upper triangular matrix $N$.

Choose $y$ to achieve the maximum of $\tau_p(W,A)$, so that $\tau_p(W,A) = \|A^*y\|_p$ with $y^*W = 0$ and $\|y\|_p = 1$. Then

$$A^*y = VTV^*y = V\left[\begin{pmatrix} \Lambda_1 & \\ & \Lambda_2 \end{pmatrix} + N\right]\begin{pmatrix} W^* \\ V_2 \end{pmatrix} y.$$

From $W^*y = 0$,

$$A^*y = V\left[\begin{pmatrix} 0 & \\ & \Lambda_2 \end{pmatrix} + N\right]\begin{pmatrix} 0 \\ V_2 \end{pmatrix} y.$$

Then

$$\tau_2(W,A) = \|A^*y\|_2 \le \|V\|_2 \left\| \begin{pmatrix} 0 & \\ & \Lambda_2 \end{pmatrix} + N \right\|_2 \|V_2\|_2$$

but since $V$ is unitary, $\|V\|_2 = \|V_2\|_2 = 1$. The from the triangle inequality and because $\|\Lambda_2\|_2 = |\lambda_{k+1}|$,

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

□

We now present some results from Walker (112) that bound subdominant eigenvalues of positive matrices using a similar idea to Theorem 3.9.8. The difference is that the deflated matrix is also a positive matrix. We begin by recalling part of Perron's Theorem.

**Theorem 3.9.13** (p 667 in (76))**.** *Define $A \in \mathbb{R}^{n\times n}$ as a matrix with strictly positive entries. Then for dominant eigenvalue $\lambda_1$ of $A$ we know the following:*

- *$\lambda_1$ is the unique dominant eigenvalue, and $\lambda_1 > 0$*
- *There exists real eigenvectors $u, v > 0$, where $Au = \lambda_1 u$ and $v^T A = \lambda_1 v^T$.*

**Theorem 3.9.14** (pp. 133–134 in (112))**.** *Define $A \in \mathbb{R}^{n\times n}$ with strictly positive entries and a unique dominant eigenvalue $\lambda_1$, and dominant right and left eigenvectors $u$ and $v$ chosen so that $v^T u = 1$. Then the matrix*

$$A_c = A - \frac{\lambda_1 u v^T}{1+c}$$

*is positive for any $c$ with $1 + c > \lambda_1 \max_{i,j} \frac{u_i v_j}{a_{ij}}$. The eigenvalues of $A_c$ are*

$$\frac{c\lambda_1}{1+c}, \lambda_2(A), \ldots, \lambda_n(A).$$

*Proof.* We begin by proving that $A_c$ is a positive matrix by examining the $(k, l)$ element of $A_c$

$$\begin{aligned}
(A_c)_{kl} &= a_{kl} - \frac{\lambda_1 u_k v_l}{1+c} \\
&> a_{kl} - \frac{\lambda_1 u_k v_l}{\lambda_1 \max_{i,j} \frac{u_i v_j}{a_{ij}}} \\
&= a_{kl} - u_k v_l \min_{i,j} \frac{a_{ij}}{u_i v_j} \\
&\geq a_{kl} - u_k v_l \frac{a_{kl}}{u_k v_l} \\
&= 0
\end{aligned}$$

since $u_k > 0$ and $v_l > 0$.

We know that $A_c$ has dominant eigenvalue $Re\left(\lambda_2\left(A\right)\right)$ when $c\lambda_1/(1+c) < Re\left(\lambda_2\left(A\right)\right)$, because in this situation, $Re\left(\lambda_2\left(A\right)\right) = \lambda_2(A)$ follows from Perron's Theorem applied to $A_c$. □

We now present another corollary to Theorem 3.9.8.

**Corollary 3.9.15.** *If $A$ and $A_c$ are positive matrices and if there exists a scalar $c$ such that*

$$c\lambda_1/(1+c) < Re\left(\lambda_2\left(A\right)\right)$$

*and $1+c > \lambda_1 \max_{i,j} \frac{u_i v_j}{a_{ij}}$, then, $Re\left(\lambda_2\left(A\right)\right) = \lambda_2(A)$ is the unique dominant eigenvalue of $A_c$, and*

$$\tau_p(u, A) \leq \left\|A_c^T\right\|_p$$

*for dominant right and left eigenvectors $u$ and $v$ of $A$, with $v^T u = 1$.*

*Proof.* This follows as in the proof of Theorem 3.9.8 from $\tau_p(u, A) = \max_{\substack{\|z\|_p=1 \\ u^T z=0}} \left\|A^T z\right\|_p$ for $z \in \mathbb{R}^n$. If $u^T z = 0$ then

$$A^T z = \left(A^T - \frac{\lambda_1 v u^T}{1+c}\right) z = \left(A - \frac{\lambda_1 u v^T}{1+c}\right)^T z = A_c^T z.$$

Hence $A^T z = A_c^T z$ whenever $u^T z = 0$.

Theorem 3.6.3 implies

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

□

### 3.9.3 Connection of $\tau_2$ to Singular Values

Without loss of generality, let $A \in \mathbb{C}^{m\times n}$, with $m \geq n$, have singular values ordered in decreasing magnitude,

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n.$$

The variational characterization for the singular values of a matrix $A$ (50, Thm 7.3.10) implies that the $i^{th}$ largest singular value of $A$ is

$$\sigma_i = \min_{w_1,\ldots,w_{i-1}\in\mathbb{C}^n} \max_{\substack{z\in\mathbb{C}^n, \|z\|_2=1 \\ z\perp w_1,\ldots,w_{i-1}}} \|Az\|_2, \qquad i = 1,\ldots,n, \tag{3.20}$$

where the minimum ranges over all $(n-i+1)-$dimensional subspaces $\mathcal{P}$ of $\mathbb{C}^n$. The extreme values are $\sigma_1 = \max_{\|z\|_2=1} \|Az\|_2$ and $\sigma_n = \min_{\|z\|_2=1} \|Az\|_2$. From expression (3.20) we relate singular values to the ergodicity coefficient $\tau_p(W, A)$ from Definition 3.6.2. We employ the fact that the singular values of $A$ are the same as the singular values of $A^*$.

We show that $\tau_2(W, A)$ is bounded above and below by singular values of $A$.

**Theorem 3.9.16.** *For $A \in \mathbb{C}^{m\times n}$ with $m \geq n$, and $W \in \mathbb{C}^{m\times k}$, $n \geq k$,*

$$\sigma_{k+1} \leq \tau_2(W, A) \leq \sigma_1.$$

*Proof.* Denote the singular values of $A$ as $\sigma(A)$, and the singular values of $A^*$ as $\sigma(A^*)$.

From $\sigma_1(A) = \sigma_1(A^*)$ and the variational characterization, we know that

$$\max_{\|z\|_2=1} \|A^* z\|_2 = \sigma_1(A^*) = \sigma_1(A) = \max_{\|z\|_2=1} \|Az\|_2.$$

Then

$$\sigma_1(A) = \sigma_1(A^*) = \max_{\|z\|_2=1} \|A^* z\|_2 \geq \max_{\substack{\|z\|_2=1 \\ z^*W=0}} \|A^* z\|_2 = \tau_2(W, A).$$

Additionally, from the variational characterization we obtain,

$$\sigma_{k+1}(A) = \sigma_{k+1}(A^*) = \min_{w_1,\ldots,w_k \in \mathbb{C}^n} \max_{\substack{z \in \mathbb{C}^n, \|z\|_2=1 \\ z \perp w_1,\ldots,w_k}} \|A^* z\|_2,$$

and so,

$$\min_{w_1,\ldots,w_k \in \mathbb{C}^n} \max_{\substack{z \in \mathbb{C}^n, \|z\|_2=1 \\ z \perp w_1,\ldots,w_k}} \|A^* z\|_2 \leq \max_{\substack{\|z\|_2=1 \\ z^*W=0}} \|A^* z\|_2 = \tau_2(W, A).$$

We also obtain

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

□

The ergodicity coefficient $\tau_p(w, A)$ from Definition 3.6.1 is a special case of $\tau_p(W, A)$ from Definition 3.6.2 with $W \in \mathbb{C}^{n \times 1}$. Therefore we now present similar bounds on the ergodicity coefficient $\tau_2(w, A)$ in terms of the singular values of $A$.

**Corollary 3.9.17.** *For $A \in \mathbb{C}^{m \times n}$ with $m \geq n$, and $w \in \mathbb{C}^{m \times 1}$,*

$$\sigma_2 \leq \tau_2(w, A) \leq \sigma_1$$

*and*

$$\sigma_2 = \min_{w \in \mathbb{C}^m} \tau_2(w, A).$$

We extend the result in Theorem 3.9.16 and Corollary 3.9.17 to eigenvalue bounds by adding the restriction that $A$ be a normal matrix. This result differs from Corollary 3.9.11, because in this corollary we have no restrictions on the columns of $W$.

**Corollary 3.9.18.** *If $A$ is an $n \times n$ normal matrix with ordered eigenvalues $|\lambda_1| \geq \cdots \geq |\lambda_n|$ and $W \in \mathbb{C}^{n \times k}, n \geq k$,*

$$|\lambda_{k+1}(A)| \leq \tau_2(W, A) \leq |\lambda_1(A)|.$$

*If $w \in \mathbb{C}^{n \times 1}$*

$$|\lambda_2| \leq \tau_2(w, A) \leq |\lambda_1|.$$

*Proof.* If $A$ is normal, then $\sigma_i = |\lambda_i|$ for $i = 1, \ldots, n$. □

### 3.9.4 Connection of $\tau_2$ to Lehmann Bounds

In this section we illustrate why the ergodicity coefficients defined in Definitions 3.6.1 and 3.6.2 can be viewed as Lehmann bounds.

For normal matrices $A$ we show that $\tau_2(W, A^*)$ is a special case of a Lehmann bound. In general, Lehmann bounds provide eigenvalue inclusion regions in terms of a matrix restricted to a subspace. Parlett (78, §10.5) presents Lehmann bounds for Hermitian matrices. Beattie and Ipsen extend Lehmann bounds to general matrices (7). We first define a Lehmann bound in terms of the singular value of a matrix.

**Definition 3.9.19** (Corollary 2.3 in (7))**.** *Let $A \in \mathbb{C}^{n\times n}$ be a normal matrix, $\alpha \in \mathbb{C}$, and $X$ an $n \times m$ matrix with orthonormal columns. Then each disk*

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

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

**Theorem 3.9.20.** *For $n \times n$ normal matrix $A$ and $\alpha \in \mathbb{C}$, the ergodicity coefficient $\tau_2\left(W, (A - \alpha I)^*\right)$ is a Lehman bound. Specifically, for every matrix $W \in \mathbb{R}^{n\times k}$, $k \leq n$, with rank$(W) = k$, the disk*

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

*contains at least $n - k$ eigenvalues of $A$.*

*Proof.* In Definition 3.9.19, let $i = 1$ and $m = n - k$. Choose $X \in \mathbb{C}^{n\times(n-k)}$ with orthonormal columns and range$(X) = \{z : z^*W = 0\}$. Then we obtain the relationship between $\tau_2$ and $\sigma_1$ from the definition of $\tau_2$ and the variational characterization

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

We have now shown that the ergodicity coefficient in the 2-norm is a singular value of a matrix restricted to a subspace. Further, $\tau_2\left(W, (A - \alpha I)^*\right)$ defines a Lehmann bound. □

**Corollary 3.9.21.** *If $A$ is normal then the coefficient $\tau_2\left(w, (A - \alpha I)^*\right)$ is a Lehmann*

*bound based on Definition 3.9.19. Specifically, for every vector* $w \in \mathbb{C}^{n\times 1}$*, the disk*

$$\{\lambda : |\lambda - \alpha| \leq \tau_2 \left(w, (A - \alpha I)^*\right)\}$$

*contains at least* $n - 1$ *eigenvalues of* $A$*.*

*Proof.* The proof follows directly from Theorem 3.9.20 with $k = 1$. □

# Chapter 4

# Conclusions and Future Work

We have presented two applications for stochastic matrices. Chapter 2 describes an application to ranking with respect to Google and the PageRank vector. The material in Chapter 2 was published in December 2007 in the SIAM Journal on Matrix Analysis and Applications (54). Chapter 3 contains results on the coefficient of ergodicity, which can be thought of as a bound on the subdominant eigenvalues of a matrix.

## 4.1 Contributions from Chapter 2

The Google matrix is a stochastic matrix whose entries contain information about the probability of moving from one webpage to another webpage. Associated with this matrix is the PageRank vector, or dominant left eigenvector of the Google matrix. This vector provides a ranking of all webpages, regardless of search topic. The values of PageRank can aid in the search engine process, but find those values is a computationally demanding process, due to the size of the Google matrix (more than 11.5

billion rows and columns in 2005 (39)).

We present a new approach to computing the PageRank vector by exploiting the structure of the Google matrix. Specifically, there are webpages called dangling nodes that have no outlinks to other webpages (some examples include image, audio, and pdf files). These webpages associated with dangling nodes are "lumped" into a single node. We showed that the lumping can be expressed as a similarity transformation, and that we can compute the PageRank of the nondangling nodes separately from that of the dangling nodes. Our PageRank algorithm applies the power method to the smaller lumped matrix, but the convergence rate is the same as that of the power method applied to the full Google matrix. The efficiency of the lumping algorithm increases as the number of dangling nodes increases. We also extend the expression for PageRank and the algorithm to more general Google matrices that have several different dangling node vectors, when it is required to distinguish among different classes of dangling nodes. We analyze the effect of the dangling node vector on the PageRank and show that the PageRank of the dangling nodes depends strongly on that of the nondangling nodes but not vice versa.

## 4.2 Contributions and Future work for Chapter 3

Ergodicity coefficients were originally introduced to estimate the rate at which a product of stochastic matrices converges to a rank-1 matrix. An ergodicity coefficient for a

stochastic matrix $S \in \mathbb{R}^{n\times n}$ can be defined as

$$\tau_p(S) = \max_{\substack{\|z\|_p=1 \\ z^T e=0}} \left\| S^T z \right\|_p \qquad \text{for } z \in \mathbb{R}^n.$$

It is the $p-$norm of a matrix restricted to a subspace. Throughout Chapter 3, we sought to present unified and consistent notation. Our goal was to give new proofs and explanations for a numerical linear algebra audience.

There is still much to be done in terms of ergodicity coefficients. The main goal of Chapter 3 was to present a new, or often forgotten, topic to the numerical linear algebra community. We believe this coefficient could be useful for many applications, but first, people need to understand what the ergodicity coefficient is and how it has been used in the past.

Much has been done on coefficients of ergodicity for stochastic matrices, specifically in the cases of $p = 1$ (Section 3.5.1) and $p = \infty$ (Section 3.5.2). Rhodius (85), Seneta (97), and Tan (110) focused on $\tau_p(S)$ for all $p$-norms. They were able to prove several properties, including continuity and submultiplicativity, as well as showing that that $\tau_p(S)$ is proper, has a computable upper bound, and is an eigenvalue bound. Additionally Rhodius (85, p 145) was able to derive a maximum value for $\tau_p(S)$ for all stochastic matrices $S$ (Section 3.5.3).

Even though so many contributions have been made for ergodicity coefficients of stochastic matrices, we still believe there is room for more results. It could be useful to compare the value of the ergodicity coefficient to the value of the second largest eigenvalue. Although we know $\tau_p(S)$ is an upper bound on subdominant eigenvalues,

no one has quantified their relationship.

Following Seneta (100), Rothblum and Tan (91, §7), and Hartfiel (46, §3), we expand on previous work and present properties in Section 3.6 for a more general ergodicity coefficient defined for $A \in \mathbb{C}^{m\times n}$ and $W \in \mathbb{C}^{m\times k}$ as

$$\tau_p(W, A) = \max_{\substack{\|z\|_p=1 \\ z^*W=0}} \|A^*z\|_p \qquad \text{for } z \in \mathbb{C}^m.$$

In addition to showing that $\tau_p(W, A)$ is continuous, subadditive, and weakly submultiplicative, we also presented relationships between this ergodicity coefficient and both eigenvalue bounds, and singular value bounds.

There are many areas where work could still be done on $\tau_p(W, A)$. We have developed several explicit forms for $\tau_2(w, A)$, one of which extends to $\tau_2(W, A)$ (see Section 3.6.2). However, we have not made progress deriving explicit forms for any other $p$-norms for $\tau_p(w, A)$ or $\tau_p(W, A)$. Like ergodicity coefficients for stochastic matrices, it would be interesting to see how tight the eigenvalue and singular value bounds are for $\tau_p(w, A)$ and $\tau_p(W, A)$.

Another possible area of interest is ergodicity coefficients as condition numbers (Section 3.9.1). Seneta (101; 102) and Kirkland (64) show how an ergodicity coefficient of a stochastic matrix can bound the condition number for the stationary distribution of stochastic linear systems $\pi^T S = \pi^T$. It is possible than an ergodicity coefficient can be used in place of a condition number in other situations as well.

There are surely other areas where the coefficient of ergodicity might be useful. These ideas for future work are just a starting point and places where obvious contri-

butions could be made. We believe that other applications exist which could benefit from inclusion of the ergodicity coefficient, but will not begin to list what they might include.

# References


[1] Y. A. Al'pin and N. Z. Gabassov, *A remark on the localization of the characteristic values of real matrices*, Izvestiya VUZ. Matematika, 20 (1976), pp. 98–100.

[2] A. Arasu, J. Novak, A. Tomkins, and J. Tomlin, *PageRank computation and the structure of the web: Experiments and algorithms*, in Proceedings of the Eleventh International World Wide Web Conference (WWW2002), New York, 2002, ACM Press. Available online at http://www2002.org/CDROM/poster/173.pdf.

[3] M. Artzrouni, *The local coefficient of ergodicity of a nonnegative matrix*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 507–516.

[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.

[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.

[6] F. L. Bauer, E. Deutsch, and J. Stoer, *Abschätzungen für Eigenwerte positiver linearer Operatoren*, Linear Algebra Appl., 2 (1969), pp. 275–301.

[7] C. Beattie and I. C. Ipsen, *Inclusion regions for matrix eigenvalues*, Linear Algebra Appl., 358 (2003), pp. 281–291.

[8] P. Berkhin, *A survey on PageRank computing*, Internet Math., 2 (2005), pp. 73–120.

[9] A. Berman and R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, Academic Press, New York, 1979.

[10] S. N. Bernštein, *Sur le théorème limite du calcul des probabilités*, Math. Ann., 85 (1922), pp. 237–241.

[11] ——, *Sur les sommes de quantités dépendantes*, Izv. Akad. Nauk SSSR, 15-17 (1926), pp. 1459–1478.

[12] ——, *Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes*, Math. Ann., 97 (1926), pp. 1–59. (Russian translation in Uspekhi Mat. Nauk, 10 (1944), pp. 65–144).

[13] ——, *Sur les sommes de quantités dépendantes*, Dokl. Akad. Nauk SSSR., A, 4 (1928), pp. 55–60.

[14] ——, *Teoriya Veroiatnostei*, Gostehizdat, Moscow-Leningrad, fourth ed., 1946. Relevant portions are republished in (15).

[15] ——, *Classification of Markov chains and their matrices*, Sobranie Sochineniy: Tom IV, Teoriya Veroiatnostei i Matematicheskaia Statistika [1911-1946], (1964), pp. 455–483.

[16] M. Bianchini, M. Gori, and F. Scarselli, *Inside PageRank*, ACM Transactions on Internet Technology, 5 (2005), pp. 92–128.

[17] C. Brezinski and M. Redivo-Zaglia, *The PageRank vector: Properties, computation, approximation, and acceleration*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 551–575.

[18] C. Brezinski, M. Redivo-Zaglia, and S. Serra-Capizzano, *Extrapolation methods for pagerank computations*, C. R. Math. Acad. Sci. Paris, 340 (2005), pp. 393–397.

[19] S. Brin and L. Page, *The anatomy of a large-scale hypertextual web search engine*, Comput. Networks and ISDN Systems, 30 (1998), pp. 107–117.

[20] A. Z. Broder, R. Lempel, F. Maghoul, and J. Pedersen, *Efficient pagerank approximation via graph aggregation*, in Proceedings of the Thirteenth International World Wide Web Conference (WWW2004), New York, 2004, ACM Press, pp. 484–485.

[21] S. Chatterjee and E. Seneta, *Towards consensus: some convergence theorems on repeated averaging*, J. Appl. Probability, 14 (1977), pp. 89–97.

[22] 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.

[23] H. Cohn, *Finite non-homogeneous Markov chains: Asymptotic behaviour*, Adv. Appl. Probab., 8 (1976), pp. 502–516.

[24] ——, *On a paper by Doeblin on non-homogeneous Markov chains*, Adv. Appl. Probab., 13 (1981), pp. 388–401.

[25] T. Dayar and W. J. Stewart, *Quasi lumpability, lower-bounding coupling matrices, and nearly completely decomposable Markov chains*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 482–498.

[26] G. M. DelCorso, A. Gullí, and F. Romani, *Fast PageRank computation via a sparse linear system*, Internet Math., 2 (2005), pp. 251–273. Available online at http://www. internetmathematics.org/volumes/2/3/DelCorso.pdf.

[27] R. L. Dobrushin, *Central limit theorem for non-stationary Markov chains*, Dokl. Akad. Nauk SSSR., 102: 1 (1955), pp. 5–8.

[28] ——, *Central limit theorem for non-stationary Markov chains, I, II*, Theory Prob. Appl., 1 (1956), pp. 65–80.

[29] W. Doeblin, *Le cas discontinu des probabilités en chaîne*, Publ. Fac. Sci. Univ. Masaryk (Brno), 236 (1937), pp. 3–13.

[30] E. B. Dynkin, *On certain limit theorems for Markov chains*, Ukrain. Mat. Zh., 6: 1 (1954), pp. 21–27.

[31] N. Eiron, K. S. McCurley, and J. A. Tomlin, *Ranking the web frontier*, in Proceedings of the Thirteenth International World Wide Web Conference (WWW2004), New York, 2004, ACM Press, pp. 309–318.

[32] L. Eldén, *The eigenvalues of the Google matrix*, Tech. Rep. LiTH-MAT-R-04-01, Department of Mathematics, Linköping University, Linköping, Sweden, 2004.

[33] N. Gastinel, *Linear numerical analysis*, Academic Press, Inc., Publishers, Paris, France, 1970.

[34] D. Gleich, L. Zhukov, and P. Berkhin, *Fast parallel PageRank: A linear system approach*, tech. rep., Yahoo!, Sunnyvale, CA, 2004.

[35] G. H. Golub and C. Greif, *An Arnoldi-type algorithm for computing PageRank*, BIT Numerical Mathematics, 46 (2006), pp. 759–771.

[36] G. H. Golub and C. F. V. Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1983.

[37] W. Greub, *Linear Algebra*, Springer-Verlag, New York, fourth ed., 1975.

[38] 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.

[39] A. Gulli and A. Signorini, *The indexable web is more than* 11.5 *billion pages*, in Proceedings of the Fourteenth International World Wide Web Conference (WWW2005), New York, 2005, ACM Press, pp. 902–903.

[40] L. Gurvits and J. Ledoux, *Markov property for a function of a Markov chain: a linear algebra approach*, Linear Algebra Appl., 404 (2005), pp. 85–117.

[41] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen, *Combating web spam with TrustRank*, in Proceedings of the Thirtieth VLDB Conference, New York, 2004, ACM Press, pp. 576–587.

[42] J. Hajnal, *The ergodic properties of non-homogeneous finite Markov chains*, Proc. Cambridge Philos. Soc., 52 (1956), pp. 67–77.

[43] ——, *Weak ergodicity in nohomogeneous Markov chains*, Proc. Cambridge Philos. Soc., 54 (1958), pp. 233–246.

[44] ——, *On products of non-negative matrices*, Math. Proc. Cambridge Philos. Soc., 79 (1976), pp. 521–530.

[45] D. Hartfiel, *Coefficients of ergodicity for imprimitive matrices*, Commun. Stat. – Stochastic Models, 15 (1999), pp. 81–88.

[46] D. Hartfiel and U. G. Rothblum, *Convergence of inhomogeneous products of matrices and coefficients of ergodicity*, Linear Algebra Appl., 277 (1998), pp. 1–9.

[47] T. H. Haveliwala and S. D. Kamvar, *The second eigenvalue of the Google matrix*, tech. rep., Computer Science Department, Stanford University, Palo Alto, CA, 2003.

[48] T. H. Haveliwala, S. D. Kamvar, D. Klein, C. D. Manning, and G. H. Golub, *Computing PageRank using power extrapolation*, Tech. Rep. 2003-45, Stanford University, Palo Alto, CA, 2003. Available online at http://dbpubs.stanford.edu/pub/2003-45.

[49] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM: Society for Industrial and Applied Mathematics, Philadelphia, second ed., 2002.

[50] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.

[51] R. A. Horn and S. Serra-Capizzano, *A general setting for the parametric google matrix*, Internet Math., (to appear).

[52] M. Iosifescu, *On two recent papers on ergodicity in non-homogeneous Markov chains*, Ann. Math. Statist., 43 (1972), pp. 1732–1736.

[53] ——, *Finite Markov Processes and their Applications*, John Wiley, 1980.

[54] I. C. Ipsen and T. M. Selee, *Pagerank computation, with special attention to dangling nodes*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 1281–1296.

[55] 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.

[56] I. C. F. Ipsen and R. S. Wills, *Mathematical properties and analysis of Google's PageRank*, Bol. Soc. Esp. Mat. Apl., 34 (2006), pp. 191–196.

[57] D. Isaacson and R. Madsen, *Positive columns for stochastic matrices*, J. Appl. Prob., 11 (1974), pp. 829–835.

[58] D. L. Isaacson and R. Madsen, *Markov Chains*, Wiley, 1976.

[59] R. W. Jernigan and R. H. Baran, *Testing lumpability in Markov chains*, Statist. Probab. Lett., 64 (2003), pp. 17–23.

[60] S. Kamvar, T. Haveliwala, and G. Golub, *Adaptive methods for the computation of PageRank*, Linear Algebra Appl., 386 (2004), pp. 51–65.

[61] S. D. Kamvar, T. H. Haveliwala, C. D. Manning, and G. H. Golub, *Extrapolation methods for accelerating PageRank computations*, in Proceedings of the Twelfth International World Wide Web Conference (WWW2003), Toronto, New York, 2003, ACM Press, pp. 261–270.

[62] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, Van Nostrand Co., Princeton, NJ, 1960.

[63] J. F. C. Kingman, *Geometrical aspects of the theory of non-homogeneous Markov chains*, Math. Proc. Cambridge Philos. Soc., 77 (1975), pp. 171–183.

[64] S. J. Kirkland, M. Neumann, and N.-S. Sze, *On optimal condition numbers for Markov chains*, Numer. Math., 110 (2008), pp. 521 – 537.

[65] A. N. Kolmogorov, *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung*, Math. Ann., 104 (1931), pp. 415–458.

[66] A. N. Langville and C. D. Meyer, *Deeper inside PageRank*, Internet Math., 1 (2004), pp. 335–380.

[67] ——, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, NJ, 2006.

[68] ——, *A reordering for the PageRank problem*, SIAM J. Sci. Comput., 27 (2006), pp. 2112–2120.

[69] ——, *Updating Markov chains with an eye on Google's PageRank*, SIAM J. Matrix Anal. Appl., 27 (2006), pp. 968–987.

[70] C. P. Lee, G. H. Golub, and S. A. Zenios, *A fast two-stage algorithm for computing PageRank and its extensions*, tech. rep., Stanford University, Palo Alto, CA, 2003.

[71] 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.

[72] A. Lešanovský, *Coefficient of ergodicity generated by non-symmetrical vector norms*, Czechoslovak Mathematical Journal, 40 (1990), pp. 284–294.

[73] A. A. Markov, *Extension of the law of large numbers to dependent quantities (in Russian)*, Izvestiia Fiz.-Matem. Obsch. Kazan Univ., 2 (1906), pp. 135–156. English translation (75).

[74] ——, *Izbrannie Trudy (Selected Works)*, A.N.S.S.S.R., Leningrad, 1951.

[75] ——, *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 (73) into English.

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

[77] L. Page, S. Brin, R. Motwani, and T. Winograd, *The PageRank citation ranking: Bringing order to the web*, 1999. Available online at http://dbpubs.stanford. edu/pub/1999-66.

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

[79] U. Păun, *A class of ergodicity coefficients, and applications*, Math. Rep. (Bucur.), 4 (54) (2002), pp. 225–232.

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

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

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

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

[84] 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.

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

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

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

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

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

[90] U. G. Rothblum, *Explicit solutions to optimization problems on the intersections of the unit ball of the $l_1$ and $l_\infty$ norms with a hyperplane*, SIAM J. Matrix Anal. Appl., 5 (1984), pp. 619–632.

[91] 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.

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

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

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

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

[96] ——, *Review of Pykh's "The characteristic numbers of stochastic matrices," #11167*, Math. Rev., 48 (1974), p. 1296.

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

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

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

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

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

[102] ——, *Sensitivity analysis, ergodicity coefficients and rank-one updates for finite Markov chains*, in Numerical solution of Markov chains, W. J. Stewart, ed., Marcel Dekker, New York, 1991, pp. 121 – 129.

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

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

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

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

[107] 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.

[108] J. Stewart, *Calculus: Concepts and Contexts 3*, Thomson Brooks/Cole, 2005.

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

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

[111] ——, *Spectrum localization of an ergodic stochastic matrix*, Bull. Inst. Math. Acad. Sinica, 12 (1984), pp. 147–151.

[112] S. G. Walker, *Improving bounds for nonmaximal eigenvalues of positive matrices*, Linear Algebra Appl., 397 (2005), pp. 133–139.

[113] R. S. Wills, *Google's PageRank: The math behind the search engine*, Math. Intelligencer, 28 (2006), pp. 6–11.

[114] C. Zenger, *A comparison of some bounds for the non-trivial eigenvalues of stochastic matrices*, Numer. Math., 19 (1972), pp. 209–211.