Nonhomogeneous Matrix Products

Darald J. Hartfiel
Department of Mathematics
Texas A&M University, USA

World Scientific
New Jersey • London • Singapore • Hong Kong

ublished by

/orld Scientific Publishing Co. Pte. Ltd.
O Box 128, Farrer Road, Singapore 912805
'SA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661
'K office: 57 Shelton Street, Covent Garden, London WC2H 9HE

ibrary of Congress Cataloging-in-Publication Data
artfiel, D. J.
Nonhomogeneous matrix products / D.J. Hartfiel.
p. cm.
Includes bibliographical references and index.
ISBN 9810246285 (alk. paper)
1. Matrices. I. Title.

QA188 .H373 2002
512.9'434--dc21 2001052618

ritish Library Cataloguing-in-Publication Data
catalogue record for this book is available from the British Library.

'inted in Singapore by Mainland Press

Preface

A matrix product A^k is called homogeneous since only one matrix occurs as a factor. More generally, a matrix product $A_k \cdots A_1$ or $A_1 \cdots A_k$ is called a nonhomogeneous matrix product.

This book puts together much of the basic work on nonhomogeneous matrix products. Such products arise in areas such as nonhomogeneous Markov chains, Markov Set-Chains, demographics, probabilistic automata, production and manpower systems, tomography, fractals, and designing curves. Thus, researchers from various disciplines are involved with this kind of work.

For theoretical researchers, it is hoped that the reading of this book will generate ideas for further work in this area. For applied fields, this book provides two chapters: Graphics and Systems, which show how matrix products can be used in those areas. Hopefully, these chapters will stimulate further use of this material.

An outline of the organization of the book follows.

The first chapter provides some background remarks. Chapter 2 covers basic functionals used to study convergence of infinite products of matrices. Chapter 3 introduces the notion of a limiting set, the set containing all limit points of $A_1, A_2A_1, \ldots$ formed from a matrix set Σ. Various properties of this set are also studied.

Chapter 4 concerns two special semigroups that are used in studies of finite products of matrices. One of these studies, ergodicity, is covered in ıapter 5. Ergodicity concerns sequences of products $A_1, A_2A_1, \ldots$, which ›pear more like rank 1 matrices as $k \to \infty$.

Chapters 6, 7, and 8 provide material on when infinite products of ma- ices converge. Various kinds of convergence are also discussed.

Chapters 9 and 10 consider a matrix set Σ and discuss the convergence $\Sigma, \Sigma^2, \ldots$ in the Hausdorff sense. Chapter 11 shows applications of is work in the areas of graphing curves and fractals. Pictures of curves ıd fractals are done with *MATLAB**. Code is added at the end of this apter.

Chapter 12 provides results on sequences $A_1x, A_2A_1x, \ldots$ of matrix prod- :ts that vary slowly. Estimates of a product in terms of the current matrix e discussed. Chapter 13, shows how the work in previous chapters can : used to study systems. MATLAB is used to show pictures and to make lculations. Code is again given at the end of the chapter.

Finally, in the Appendix, a few results used in the book are given. This done for the convenience of the reader.

In conclusion, I would like to thank my wife, Faye Hartfiel, for typing this ›ok and for her patience in the numerous rewritings, and thus retypings, it. In addition, I would also like to thank E. H. Chionh and World :ientific Publishing Co. Pte. Ltd. for their patience and kindness while I :ote this book.

Darald J. Hartfiel

*MATLAB is a registered trademark of The MathWorks, Inc. For product infor- ıtion, please contact:

The MathWorks, Inc.
3 Apple Hill Drive
Natick, MA 01760-2098
Tel: 508 647-7000
Fax: 508-647-7001
E-mail: info@mathworks.com
Web: www.mathworks.com

Contents

1 Introduction

Let F denote either the set R of real numbers or the set C of complex numbers. Then F^n will denote the n-dimensional vector space of n-tuples over the field F.

Vector norms $\|\cdot\|$ in this book use the standard notation, $\|\cdot\|_p$ denotes the p-norm. Correspondingly induced matrix norms use the same notation. Recall, for any $n \times n$ matrix A, $\|A\|$ can be defined as

$$\|A\| = \max_{\|x\|=1} \|xA\| \quad \text{or} \quad \|A\| = \max_{\|x\|=1} \|Ax\| .$$

So, if the vector x is on the left

$$\|A\|_1 = \max_i \sum_{k=1}^{n} |a_{ik}| , \; \|A\|_\infty = \max_i \sum_{k=1}^{n} |a_{ki}|$$

while if x is on the right,

$$\|A\|_1 = \max_i \sum_{k=1}^{n} |a_{ki}| , \; \|A\|_\infty = \max_i \sum_{k=1}^{n} |a_{ik}| .$$

In measuring distances, we will use norms, except in the case where we use the projective metric. The projective metric occurs in positive vector and nonnegative matrix work.

Let M_n denote the set of $n \times n$ matrices with entries from F. By a *matrix norm* $\|\cdot\|$ on M_n, we will mean any norm on M_n that also satisfies

$$\|AB\| \leq \|A\| \, \|B\|$$

for all $A, B \in M_n$. Of course, all induced matrix norms are matrix norms.

This book is about products, called *nonhomogeneous products*, formed from M_n. An infinite product of matrices taken from M_n is an expressed product

$$\cdots A_{k+1} A_k \cdots A_1 \tag{1.1}$$

where each $A_i \in M_n$. More compactly, we write

$$\prod_{k=1}^{\infty} A_k.$$

This infinite product of matrices converges, with respect to some norm, if the sequence of products

$$A_1, A_2 A_1, A_3 A_2 A_1, \ldots$$

converges. Since norms on M_n are equivalent, convergence does not depend on the norm used.

If we want to make it clear that products are formed by multiplying on the left, as in (1.1), we can call this a *left infinite product* of matrices. A *right infinite product* is an expression

$$A_1 A_2 \cdots$$

which can also be written compactly as $\prod_{k=1}^{\infty} A_k$. Unless stated otherwise, we will work with left infinite products.

In applications, the matrices used to form products are usually taken from a specified set. In working with these sets, we use that if X is a set of $m \times k$ matrices and Y a set of $k \times n$ matrices, then

$$XY = \{AB : A \in X \text{ and } B \in Y\}.$$

And, as is customary, if X or Y is a singleton, we use the matrix, rather than the set, to indicate the product. So, if $X = \{A\}$ or $Y = \{B\}$, we write

$$AY \text{ or } XB,$$

respectively.

Any subset Σ of M_n is called a *matrix set*. Such a set is *bounded* if there is a positive constant β such that for some matrix norm $\|\cdot\|$, $\|A\| \leq \beta$ for all $A \in \Sigma$. The set Σ is *product bounded* if there is a positive constant β where $\|A_k \cdots A_1\| \leq \beta$ for all k and all $A_1, \ldots, A_k \in \Sigma$. Since matrix norms on M_n are all equivalent, if Σ is bounded or product bounded for one matrix norm, the same is true for all matrix norms.

All basic background information used on matrices can be found in Horn and Johnson (1996). The Perron-Frobenius theory, as used in this book, is given in Gantmacher (1964). The basic result of the Perron-Frobenius theory is provided in the Appendix.

2
Functionals

Much of the work on infinite products of matrices uses one functional or another. In this chapter we introduce these functionals and show some of their basic properties.

2.1 Projective and Hausdorff Metrics

The projective and Hausdorff metrics are two rather dated metrics. However, they are not well known, and there are some newer results. So we will give a brief introduction to them.

2.1.1 *Projective Metric*

Let $x \in R^n$ where $x = (x_1, \ldots, x_n)^t$. If

1. $x_i \geq 0$ for all i, then x is *nonnegative*, while if
2. $x_i > 0$ for all i, then x is *positive*.

If x and y are in R^n, we write

1. $x \geq y$ if $x_i \geq y_i$ for all i, and

2. $x > y$ if $x_i > y_i$ for all i.

The same terminology will be used for matrices.

The *positive orthant*, denoted by $(R^n)^+$, is the set of all positive vectors in R^n. The *projective metric* p introduced by David Hilbert (1895), defines a scaled distance between any two vectors in $(R^n)^+$. As we will see, if x and y are positive vectors in R^n, then

$$p(x,y) = p(\alpha x, \beta y)$$

for any positive constants α and β. Thus, the projective metric does not depend on the length of the vectors involved and so, as seen in Figure 2.1, $p(x,y)$ can be calculated by projecting the vectors to any desired position.

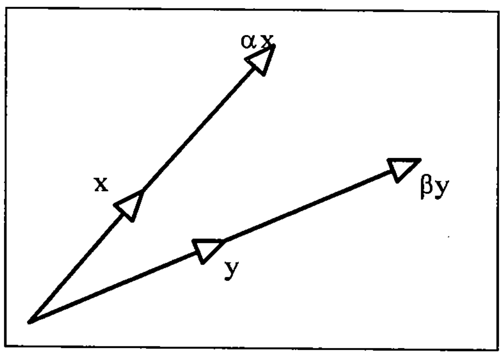

FIGURE 2.1. Various scalings of x and y.

The projective metric is defined below.

Definition 2.1 *Let x and y be positive vectors in R^n. Then*

$$p(x,y) = \ln \frac{\max_i \frac{x_i}{y_i}}{\min_j \frac{x_j}{y_j}}.$$

Other expressions for $p(x,y)$ follow.

1. $p(x,y) = \ln \max_{i,j} \frac{x_i y_j}{x_j y_i}$, the natural log of the largest cross product ratio in $[x\ y] = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \cdots & \\ x_n & y_n \end{bmatrix}$.

2. $p(x,y) = \ln\left(\max_i \frac{x_i}{y_i} \max_i \frac{y_i}{x_i}\right)$.

In working with $p(x,y)$, for notational simplicity, we define

$$M\left(\frac{x}{y}\right) = \max_i \frac{x_i}{y_i}$$

and

$$m\left(\frac{x}{y}\right) = \min_j \frac{x_j}{y_j}.$$

Some basic properties of the projective metric follow.

Theorem 2.1 *For all positive vectors x, y, and z in R^n, we have the following:*

1. $p(x,y) \geq 0$.
2. $p(x,y) = 0$ *iff* $x = \alpha y$ *for some positive constant* α.
3. $p(x,y) = p(y,x)$.
4. $p(x,y) \leq p(x,z) + p(z,y)$.
5. $p(\alpha x, \beta y) = p(x,y)$ *for any positive constants* α *and* β.

Proof. We prove the parts which don't follow directly from the definition of p.

Part (2). We prove that if $p(x,y) = 0$, then $x = \alpha y$ for some constant α. For this, if $p(x,y) = 0$,

$$\frac{M\left(\frac{x}{y}\right)}{m\left(\frac{x}{y}\right)} = 1$$

so

$$M\left(\frac{x}{y}\right) = m\left(\frac{x}{y}\right).$$

Since

$$M\left(\frac{x}{y}\right) \geq \frac{x_k}{y_k} \geq m\left(\frac{x}{y}\right)$$

for all k,

$$\frac{x_k}{y_k} = M\left(\frac{x}{y}\right)$$

for all k. Thus, setting $\alpha = M\left(\frac{x}{y}\right)$, we have that

$$x = \alpha y.$$

Part (4). We show that if x, y, and z are positive vectors in R^n, then $p(x,y) \le p(x,z) + p(z,y)$. To do this, observe that

$$x \le M\left(\frac{x}{z}\right) z \le M\left(\frac{x}{z}\right) M\left(\frac{z}{y}\right) y.$$

Thus,

$$\frac{x_j}{y_j} \le M\left(\frac{x}{z}\right) M\left(\frac{z}{y}\right)$$

for all j and so

$$M\left(\frac{x}{y}\right) \le M\left(\frac{x}{z}\right) M\left(\frac{z}{y}\right).$$

Similarly,

$$m\left(\frac{x}{y}\right) \ge m\left(\frac{x}{z}\right) m\left(\frac{z}{y}\right).$$

Putting together,

$$\begin{aligned} p(x,y) &= \ln \frac{M\left(\frac{x}{y}\right)}{m\left(\frac{x}{y}\right)} \\ &\le \ln \frac{M\left(\frac{x}{z}\right) M\left(\frac{z}{y}\right)}{m\left(\frac{x}{z}\right) m\left(\frac{z}{y}\right)} \\ &= \ln \frac{M\left(\frac{x}{z}\right)}{m\left(\frac{x}{z}\right)} + \ln \frac{M\left(\frac{z}{y}\right)}{m\left(\frac{z}{y}\right)} \\ &= p(x,z) + p(z,y). \end{aligned}$$

This inequality provides (4). ■

From property (2) of the theorem, it is clear that p is not a metric. As a consequence, it is usually called a pseudo-metric. Actually, if for each positive vector x, we define

$$ray(x) = \{\alpha x : \alpha \ge 0\}$$

then p determines a metric on these rays.

For a geometrical view of p, let x and y be positive vectors in R^n. Let α be the smallest positive constant such that

$$\alpha x \ge y.$$

Then $\alpha = M\left(\frac{y}{x}\right)$. Now let β be the smallest positive constant such that

$$\alpha x \le \beta y.$$

Thus, $\beta = M\left(\frac{\alpha x}{y}\right)$. (See Figure 2.2.) Calculation yields

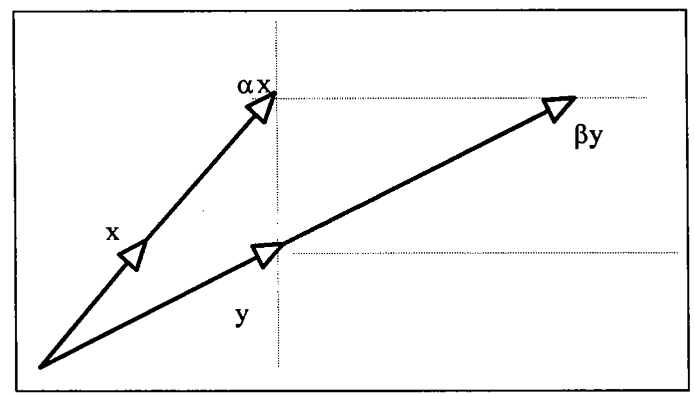

FIGURE 2.2. Geometrical view of $p(x,y)$.

$$\beta = \alpha M\left(\frac{x}{y}\right) = M\left(\frac{y}{x}\right) M\left(\frac{x}{y}\right)$$

so

$$p(x,y) = \ln \beta.$$

Thus, as a few sketches in R^2 can show, if x and y are close to horizontal or vertical, $p(x,y)$ can be large even when x and y are close in the Euclidean distance.

Another geometrical view can be seen by considering the curve $C = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : x_1 x_2 = 1 \right\}$ in the positive orthant of R^2. Here, $p(x,y)$ is two times the area shaded in Figure 2.3. Observe that as x and y are rotated

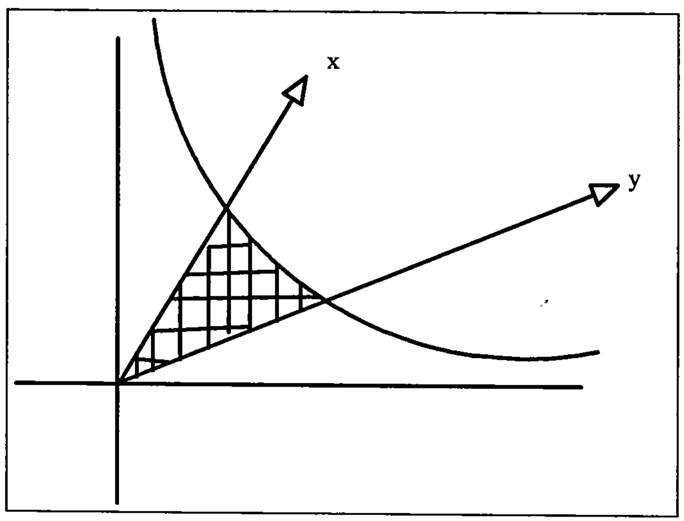

FIGURE 2.3. Another geometrical view of $p(x,y)$.

toward the x-axis or y-axis, the projective distance increases.

In the last two theorems in this section, we provide numerical results showing something of what we described geometrically above.

Theorem 2.2 *Let x and y be positive vectors in R^n.*

1. *If $p(x,y) \leq \epsilon$, $r = \min_j \frac{x_j}{y_j}$, and $m_i = \frac{x_i/y_i - r}{r}$, then we have that $m_i \leq e^{\epsilon} - 1$ and $x = r(y + My)$ where the matrix $M = \operatorname{diag}(m_1, \ldots, m_n)$.*

2. *Suppose $x = r(y + My)$, $M = \operatorname{diag}(m_1, \ldots, m_n) \geq 0$ and $r > 0$. If $m_i \leq e^{\epsilon} - 1$ for all i, then $p(x,y) \leq \epsilon$.*

Proof. We prove both parts.

Part (1). Using the definitions of r and M, we show that $m_i \leq e^{\epsilon} - 1$ and that $x = r(y + My)$. For the first part, suppose that $p(x,y) \leq \epsilon$. Then

$$\ln \max_{i,j} \frac{x_i/y_i}{x_j/y_j} \leq \epsilon$$

so we have

$$\max_{i,j} \frac{x_i/y_i}{x_j/y_j} \leq e^{\epsilon}.$$

Thus, for any i,

$$\frac{x_i/y_i}{\min_j x_j/y_j} \leq e^{\epsilon}$$

and so by subtracting 1,

$$\frac{x_i/y_i - \min_j x_j/y_j}{\min_j x_j/y_j} \leq e^{\epsilon} - 1$$

or

$$m_i \leq e^{\epsilon} - 1.$$

Now note for the second part that

$$1 + m_i = 1 + \frac{x_i/y_i - \min_j x_j/y_j}{\min_j x_j/y_j} = \frac{x_i/y_i}{\min_j x_j/y_j}.$$

Since $r = \min_j \frac{x_j}{y_j}$, we have

$$1 + m_i = \frac{1}{r}\frac{x_i}{y_i},$$

so

$$r(1 + m_i) y_i = x_i$$

or $r(y + My) = x$.

Part (2). Note by using the hypothesis, that

$$\begin{aligned} p(x,y) &= p(ry + rMy, y) = p(y + My, y) \\ &= \max_{i,j} \ln \frac{\frac{y_i + m_i y_i}{y_i}}{\frac{y_j + m_j y_j}{y_j}} \\ &= \max_{i,j} \ln \frac{1 + m_i}{1 + m_j} \leq \max_i \ln(1 + m_i) \leq \ln e^{\epsilon} = \epsilon \end{aligned}$$

the desired result. ■

Observe in the theorem, taking r and M as in (1), we have that $x = r(y + My)$. Then, r is the largest positive constant such that

$$\frac{1}{r}x - y \geq 0.$$

And

$$m_i = \frac{\frac{1}{r}x_i - y_i}{y_i}$$

for all i. Thus, viewing Figure 2.4, we estimate that $m_2 \approx 1$, so, by (2), $p(x,y) \approx \ln 2$. Turning y toward x yields a smaller projective distance.

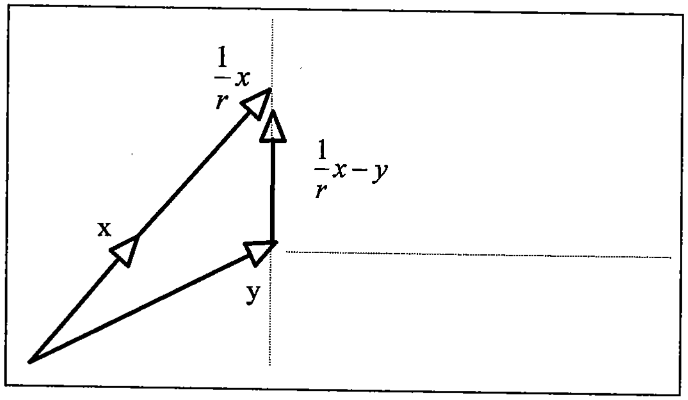

FIGURE 2.4. A view of m_1 and m_2.

In the following theorem we assume that the positive vectors x and y have been scaled so that $\|x\|_1 = \|y\|_1 = 1$. Any nonnegative vector z in R^n such that $\|z\|_1 = 1$ is called a *stochastic vector*. We also use that $e = (1, 1, \ldots, 1)^t$, the vector of 1's in R^n.

Theorem 2.3 *Let x and y be positive stochastic vectors in R^n. We have the following:*

1. $\|x - y\|_1 \leq e^{p(x,y)} - 1$.
2. $p(x,y) \leq \frac{8}{3}\frac{\|x-y\|_1}{m\left(\frac{y}{e}\right)}$ *provided that* $m\left(\frac{y}{e}\right) > 2\|x - y\|_1$.

Proof. We argue both parts.

Part (1). For any given i, if $x_i \geq y_i$, then, since $\frac{x_i}{y_i} \leq M\left(\frac{x}{y}\right)$ and $m\left(\frac{x}{y}\right) \leq 1$,

$$x_i - y_i \leq M\left(\frac{x}{y}\right) y_i - m\left(\frac{x}{y}\right) y_i.$$

And if $y_i \geq x_i$, then since $M\left(\frac{x}{y}\right) \geq 1$ and $\frac{x_i}{y_i} \geq m\left(\frac{x}{y}\right)$,

$$y_i - x_i \leq M\left(\frac{x}{y}\right) y_i - m\left(\frac{x}{y}\right) y_i.$$

Thus,

$$|x_i - y_i| \leq M\left(\frac{x}{y}\right) y_i - m\left(\frac{x}{y}\right) y_i.$$

It follows that

$$\begin{aligned} \|x - y\|_1 &\leq M\left(\frac{x}{y}\right) - m\left(\frac{x}{y}\right) \\ &= \left(\frac{M\left(\frac{x}{y}\right)}{m\left(\frac{x}{y}\right)} - 1\right) m\left(\frac{x}{y}\right) \\ &\leq e^{p(x,y)} - 1. \end{aligned}$$

Part (2). Note that

$$M\left(\frac{x}{y}\right) = \max_i \left\{1 + \frac{x_i - y_i}{y_i}\right\} \leq 1 + \frac{\|x - y\|_1}{m\left(\frac{y}{e}\right)}. \tag{2.1}$$

And, if $m\left(\frac{y}{e}\right) > 2\|x - y\|_1$, similarly we can get

$$m\left(\frac{x}{y}\right) \geq 1 - \frac{\|x - y\|_1}{m\left(\frac{y}{e}\right)} > 0. \tag{2.2}$$

Now using (2.1) and (2.2), we have

$$\frac{M\left(\frac{x}{y}\right)}{m\left(\frac{x}{y}\right)} \leq \frac{1 + \frac{\|x-y\|_1}{m\left(\frac{y}{e}\right)}}{1 - \frac{\|x-y\|_1}{m\left(\frac{y}{e}\right)}}.$$

And, using calculus

$$\begin{aligned} p(x,y) &\le \ln\left(1+\frac{\|x-y\|_1}{m\left(\frac{y}{e}\right)}\right) - \ln\left(1-\frac{\|x-y\|_1}{m\left(\frac{y}{e}\right)}\right) \\ &\le \frac{8}{3}\frac{\|x-y\|_1}{m\left(\frac{y}{e}\right)}, \end{aligned}$$

which is what we need. ■

This theorem shows that if we scale positive vectors x and y to $\frac{x}{\|x\|_1}$ and $\frac{y}{\|y\|_1}$, and $\min_i \frac{y_i}{\|y\|_1}$ is not too small, then x is close to y in the projected sense iff $\frac{x}{\|x\|_1}$ is close to $\frac{y}{\|y\|_1}$ in the 1-norm. See Figure 2.5.

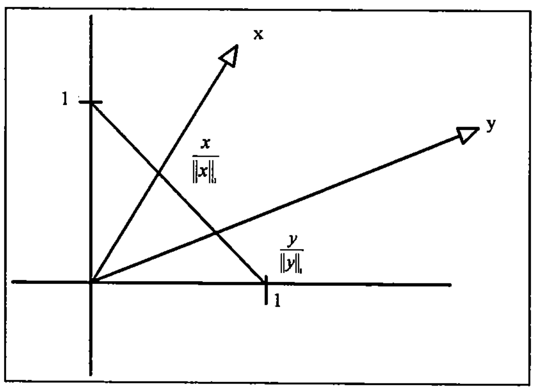

FIGURE 2.5. Projected vectors in R^2.

2.1.2 *Hausdorff Metric*

The Hausdorff metric gives the distance between two compact sets. It can be defined in a rather general setting. To see this, let (X,d) be a complete metric space where X is a subset of F^n or M_n and d a metric on X.

If K is a compact subset of X and $l \in X$, we can take a sequence $k_1, k_2, \ldots$ in K such that $d(l,k_1), d(l,k_2), \ldots$ converges to $\inf_{k\in K} d(l,k)$. And, take a subsequence $k_{i_1}, k_{i_2}, \ldots$ that converges to, say $\hat{k} \in K$ as depicted in Figure 2.6. Then,

$$\inf_{k\in K} d(l,k) = d\left(l,\hat{k}\right).$$

Hence, we can define

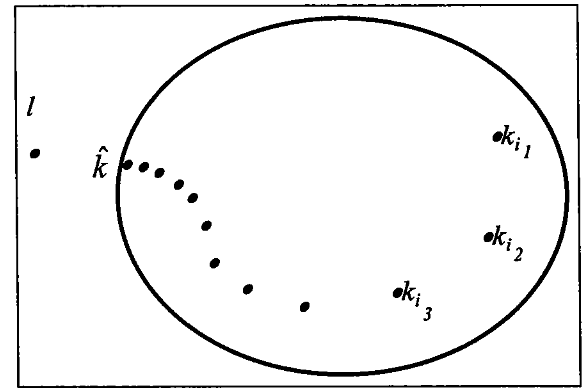

FIGURE 2.6. A view of l and $\hat{k}$.

$$d(l,K) = \min_{k\in K} d(l,k).$$

Note that if $d(l,K) \le \epsilon$, then

$$l \in K + \epsilon$$

where $K+\epsilon = \{x : d(x,k) \le \epsilon \text{ for some } k \in K\}$ as shown in Figure 2.7. We can also show that if L is a compact subset of X, then $\sup_{l\in L} d(l,K) = d\left(\hat{l},K\right)$ for some $\hat{l} \in L$. Using these observations, we define

$$\begin{aligned} \delta(L,K) &= \max_{l\in L} d(l,K) \\ &= \max_{l\in L}\min_{k\in K} d(l,k) \\ &= d\left(\hat{l},\hat{k}\right). \end{aligned}$$

If $\delta(L,K) = \epsilon$, then observe, as in Figure 2.8, that $L \subseteq K+\epsilon$.

The *Hausdorff metric* h defines the distance between two compact sets, say L and K, of X as

$$h(L,K) = \max\{\delta(L,K), \delta(K,L)\}.$$

So if $h(L,K) \le \epsilon$, then

$$L \subseteq K + \epsilon \text{ and } K \subseteq L + \epsilon$$

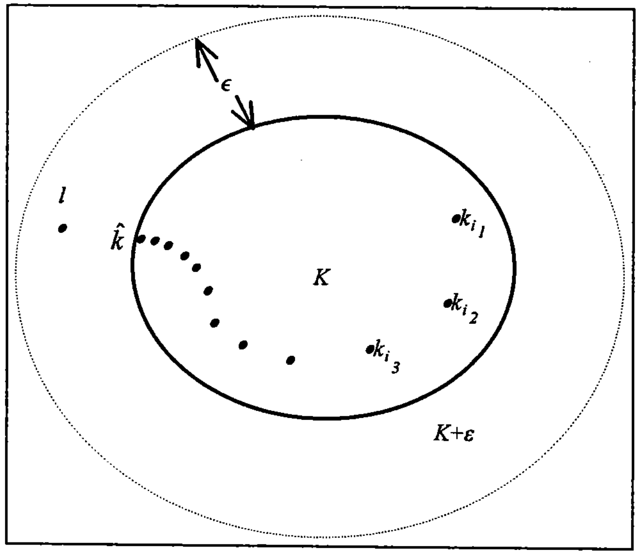

FIGURE 2.7. A sketch for $d(l, K)$.

and vice versa.

In the following we use that $H(X)$ is the set of all compact subsets of X.

Theorem 2.4 *Using that (X, d) is a complete metric space, we have that $(H(X), h)$ is a complete metric space.*

Proof. To show that h is a metric is somewhat straightforward. Thus, we will only show the triangular inequality. For this, let R, S, and T be in $H(X)$. Then for any $r \in R$ and $t \in T$,

$$\begin{aligned} d(r, S) &= \min_{s \in S} d(r, s) \\ &\leq \min_{s \in S} (d(r, t) + d(t, s)) \\ &= d(r, t) + \min_{s \in S} d(t, s) \\ &= d(r, t) + d(t, S). \end{aligned}$$

Since this holds for any $t \in T$,

$$d(r, S) \leq d(r, \hat{t}) + d(\hat{t}, S)$$

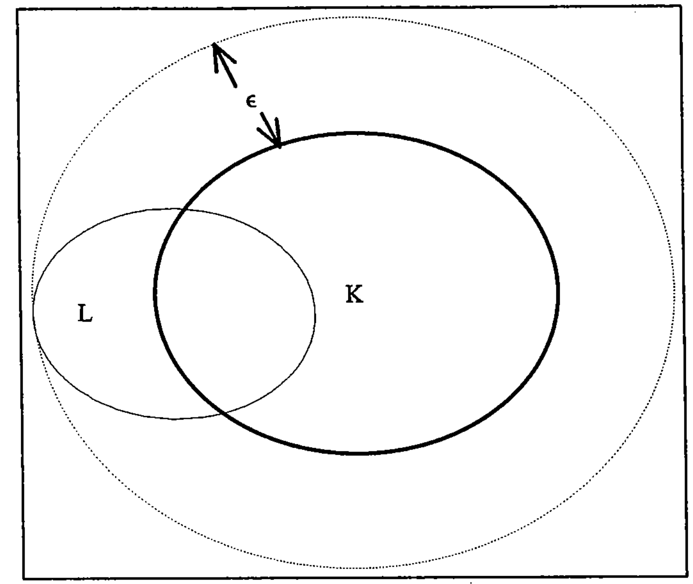

FIGURE 2.8. A sketch showing $\delta(L, K) < \epsilon$.

where $d(r, \hat{t}) = \min_{t \in T} d(r, t) = d(r, T)$. Finally,

$$\begin{aligned} \delta(R, S) &= \max_{r \in R} d(r, S) \\ &\leq \max_{r \in R} d(r, \hat{t}) + \max_{t \in T} d(t, S) \\ &= \max_{r \in R} d(r, T) + \max_{t \in T} d(t, S) \\ &= \delta(R, T) + \delta(T, S). \end{aligned}$$

Putting together, we have

$$\begin{aligned} h(R, S) &= \max\{\delta(R, S), \delta(S, R)\} \\ &\leq \max\{\delta(R, T) + \delta(T, S), \delta(S, T) + \delta(T, R)\} \\ &\leq \max\{\delta(R, T), \delta(T, R)\} + \max\{\delta(T, S), \delta(S, T)\} \\ &= h(R, T) + h(T, S). \end{aligned}$$

The proof that $(H(X), h)$ is a complete metric space is somewhat intricate as well as long. Eggleston (1969) is a source for this argument. ■

To conclude this section, we link the projective metric p and the Hausdorff metric. To do this, let S^+ denote the set of all positive stochastic vectors in R^n. As shown below, p restricted to S^+ is a metric.

Theorem 2.5 *(S^+, p) is a complete metric space.*

Proof. To show that p is a metric, we need only to show that if x and y are in S^+ and $p(x,y) = 0$, then $x = y$. This follows since if $p(x,y) = 0$, then $y = \alpha x$ for some scalar α. And since $x, y \in S^+$, their components satisfy

$$y_1 + \cdots + y_n = \alpha (x_1 + \cdots + x_n).$$

So $\alpha = 1$. Thus, $x = y$.

Finally, to show that (S^+, p) is complete, observe that if $\langle x_k \rangle$ is a Cauchy sequence from S^+, then the components of the x_k's are bounded away from 0. Then, apply Theorem 2.3. ■

As a consequence, we have the following.

Corollary 2.1 *Using that (S^+, p) is the complete metric space, we have that $(H(S^+), h)$ is a complete metric space.*

2.2 Contraction Coefficients

Let $\sum$ be a matrix set and

$$\Lambda = \Sigma \cup \Sigma^2 \cup \cdots.$$

A nonnegative function τ

$$\tau : \Lambda \to R$$

is called a *contraction coefficient* for Σ if

$$\tau(AB) \le \tau(A)\,\tau(B)$$

for all $A, B \in \Lambda$.

Contraction coefficients are used to show that a sequence of vectors or a sequence of matrices converges in some sense. In this section, we look at two kinds of contraction coefficients. And we do this in subsections.

2.2.1 Birkhoff Contraction Coefficient

The contraction coefficient for the projective metric, introduced by G. Birkhoff (1967), is defined on the special nonnegative matrices described below.

Definition 2.2 *An $m \times n$ nonnegative matrix A is row allowable if it has a positive entry in each of its rows.*

Note that if A is row allowable and x and y are positive vectors, then Ax and Ay are positive vectors. Thus we can compare $p(Ax, Ay)$ and $p(x,y)$. To do this, we use the *quotient bound result* that if $r_1, \ldots, r_n$ and $s_1, \ldots, s_n$ are positive constants, then

$$\min_i \frac{r_i}{s_i} \le \frac{r_1 + \cdots + r_n}{s_1 + \cdots + s_n} \le \max_i \frac{r_i}{s_i}. \tag{2.3}$$

This result is easily shown by induction or by using calculus.

Lemma 2.1 *Let A be an $m \times n$ row allowable matrix and x and y positive vectors. Then*

$$p(Ax, Ay) \le p(x,y).$$

Proof. Let $\hat{x} = Ax$ and $\hat{y} = Ay$. Then

$$\frac{\hat{x}_i}{\hat{y}_i} = \frac{a_{i1}x_1 + \cdots + a_{in}x_n}{a_{i1}y_1 + \cdots + a_{in}y_n}.$$

Thus using (2.3),

$$\min_k \frac{x_k}{y_k} \le \frac{\hat{x}_i}{\hat{y}_i} \le \max_k \frac{x_k}{y_k}$$

for all i. So

$$\frac{\max_i \frac{\hat{x}_i}{\hat{y}_i}}{\min_i \frac{\hat{x}_j}{\hat{y}_j}} \le \frac{\max_k \frac{x_k}{y_k}}{\min_k \frac{x_k}{y_k}}$$

and thus,

$$p(\hat{x}, \hat{y}) \le p(x,y)$$

or

$$p(Ax, Ay) \le p(x,y),$$

which yields the lemma. ■

For slightly different matrices, we can show a strict inequality result.

Lemma 2.2 *Let A be a nonnegative matrix with a positive column. If x and y are positive vectors, and $p(x,y) > 0$, then*

$$p(Ax, Ay) < p(x,y).$$

Proof. Set $\hat{x} = Ax$ and $\hat{y} = Ay$. Define $r_i = \frac{x_i}{y_i}$ and $\hat{r}_i = \frac{\hat{x}_i}{\hat{y}_i}$ for all i. Further, define $\hat{M} = \max_i \hat{r}_i$, $\hat{m} = \min_i \hat{r}_i$, $M = \max_i r_i$, and $m = \min_i r_i$.

Now

$$\hat{r}_i = \frac{\sum_{j=1}^{n} a_{ij}x_j}{\sum_{j=1}^{n} a_{ij}y_j} = \sum_{j=1}^{n}\left(\frac{a_{ij}y_j}{\sum_{j=1}^{n} a_{ij}y_j}\right)\frac{x_j}{y_j}.$$

Set $\alpha_{ij} = \frac{a_{ij}y_j}{\sum_{j=1}^{n} a_{ij}y_j} \geq 0$. Then

$$\hat{r}_i = \sum_{j=1}^{n} \alpha_{ij} r_j, \tag{2.4}$$

a convex sum.

Suppose

$$\hat{M} = \sum_{j=1}^{n} \alpha_{pj} r_j \text{ and } \hat{m} = \sum_{j=1}^{n} \alpha_{qj} r_j .$$

Using that these are convex sums, $\hat{M} \leq M$ and $\hat{m} \geq m$. Without loss of generality, assume the first column of A is positive. If $\hat{M} = M$ and $\hat{m} = m$, then since $\alpha_{i1} > 0$ for all i, by (2.4), $r_1 = M = m$, which means $p(x,y) = 0$, denying the hypothesis. Thus, suppose one of $\hat{M} < M$ or $\hat{m} > m$ holds. Then

$$\hat{M} - \hat{m} < M - m$$

and so

$$\frac{\hat{M}}{\hat{m}} - 1 < \frac{M}{m} - 1.$$

It follows that

$$p(Ax, Ay) < p(x,y),$$

the desired result. ∎

Lemma 2.1 shows that ray (Ax) and ray (Ay) are no farther apart than ray (x) and ray (y) as depicted in Figure 2.9. And, if A has a positive column, ray (Ax) and ray (Ay) are actually closer than ray (x) and ray (y).

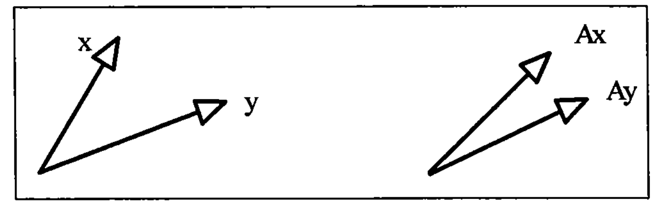

FIGURE 2.9. A ray view of $p(Ax, Ay) \leq p(x,y)$.

Define the projective coefficient, called the *Birkhoff contraction coefficient*, of an $n \times n$ row allowable matrix A as

$$\tau_B(A) = \sup \frac{p(Ax, Ay)}{p(x,y)}$$

where the sup is taken over all positive vectors in R^n. Thus,

$$p(Ax, Ay) \leq \tau_B(A)\, p(x,y)$$

for all positive x, y. And, it follows by Lemma 2.1 that

$$\tau_B(A) \leq 1.$$

Note that τ_B indicates how much ray (x) and ray (y) are drawn together when multiplying by A. A picture of this, using the area view of $p(x,y)$, is shown in Figure 2.10.

Actually, there is a formula for computing $\tau_B(A)$ in terms of the entries of A. To provide this formula, we need a few preliminary remarks.

Let A be an $n \times n$ positive matrix with $n > 1$. For any 2×2 submatrix of A, say

$$\begin{bmatrix} a_{pq} & a_{ps} \\ a_{rq} & a_{rs} \end{bmatrix}$$

the constants

$$\frac{a_{pq}a_{rs}}{a_{ps}a_{rq}}, \frac{a_{ps}a_{rq}}{a_{pq}a_{rs}}$$

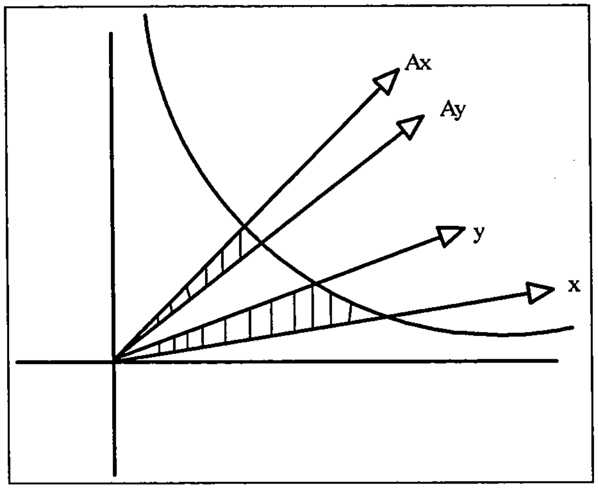

FIGURE 2.10. An area view of $\tau_B(A) = \frac{1}{2}$.

are *cross ratios*. Define

$$\phi(A) = \min \frac{a_{pq}a_{rs}}{a_{rq}a_{ps}}$$

where the minimum is over all cross ratios of A. For example, if $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then $\phi(A) = \min\left\{\frac{4}{6}, \frac{6}{4}\right\} = \frac{2}{3}$.

If A is row allowable and contains a 0 entry, define $\phi(A) = 0$. Thus for any row allowable matrix A,

$$\phi(A) \leq 1.$$

The formula for $\tau_B(A)$ can now be given. Its proof, rather intricate, can be found in Seneta (1981).

Theorem 2.6 *Let A be an $n \times n$ row allowable matrix. Then*

$$\tau_B(A) = \frac{1 - \sqrt{\phi(A)}}{1 + \sqrt{\phi(A)}}.$$

Note that this theorem implies that $\tau(A) < 1$ when A is positive and $\tau(A) = 1$ if A is row allowable and has at least one 0 entry. And that

$$p(Ax, Ay) \leq \tau_B(A)\, p(x, y)$$

for all row allowable matrices A and positive vectors x and y.

Using our previous example, where $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, we have

$$\tau_B(A) = \frac{1 - \sqrt{\frac{2}{3}}}{1 + \sqrt{\frac{2}{3}}} \approx .10$$

so for any positive vectors x and y, ray(Ax) and ray(Ay) are closer than ray(x) and ray(y).

This theorem also assures that if A is a positive $n \times n$ matrix and D_1, D_2, $n \times n$ diagonal matrices with positive main diagonals, then $\tau_B(D_1AD_2) = \tau_B(A)$. Thus, scaling the rows and columns of A does not change the contraction coefficient.

It is interesting to see what $\tau_B(A) = 0$ means about A.

Theorem 2.7 *Let A be a positive $n \times n$ matrix. If $\tau_B(A) = 0$, then A is* rank 1.

Proof. Suppose $\tau_B(A) = 0$. We will show that the i-th row of A is a scalar multiple of the 1-st row of A.

Define $\alpha = \frac{a_{i1}}{a_{11}}$. Then, since $\tau_B(A) = 0$, $\phi(A) = 1$ which assures that all cross ratios of A are 1. Thus,

$$\frac{a_{11}a_{ij}}{a_{i1}a_{1j}} = 1$$

for all j. Thus, $\frac{a_{ij}}{\alpha a_{1j}} = 1$ or $a_{ij} = \alpha a_{1j}$. Since this holds for all j, the i-th row of A is a scalar multiple of the first row of A. Since i was arbitrary, A is rank 1. ■

Probably the most useful property of τ_B follows.

Theorem 2.8 *Let A and B be $n \times n$ row allowable matrices. Then*

$$\tau_B(AB) \leq \tau_B(A)\,\tau_B(B).$$

Proof. Let x and y be positive vectors in R^n. Then Bx and By are also positive vectors in R^n. Thus

$$p(ABx, ABy) \leq \tau_B(A)\,\tau_B(B)\,p(x, y).$$

And, since this inequality holds for all positive vectors x and y in R^n,

$$\tau_B(AB) \leq \tau_B(A)\,\tau_B(B)$$

as desired. ■

We use this property as we use induced matrix norms.

Corollary 2.2 *If A is an $n \times n$ row allowable matrix and y a positive eigenvector for A, then for any positive vector x, $p\left(A^k x, y\right) \le \tau_B(A)^k p(x,y)$.*

Proof. Note that

$$\begin{aligned} p\left(A^k x, y\right) &= p\left(A^k x, A^k y\right) \\ &\le \tau_B(A)^k p(x,y) \end{aligned}$$

for all positive integers k. ■

This corollary assures that for a positive matrix A,

$$\lim_{k\to\infty} p\left(A^k x, y\right) = 0,$$

so ray $\left(A^k x\right)$ gets closer to ray (y) as k increases.

We will conclude this section by extending our work to compact subsets. To do this, recall that (S^+, p) is a complete metric space. Define the Hausdorff metric on the compact subsets (closed subsets in the 1-norm) of S^+ by using the metric p. That is,

$$\delta(U,V) = \max_{u\in U}\left(\min_{v\in V} p(u,v)\right)$$

and

$$h(U,V) = \max\{\delta(U,V), \delta(V,U)\}$$

where U and V are any two compact subsets of S^+.

Let Σ be any compact subset of $n \times n$ row allowable matrices. For each $A \in \Sigma$, define the *projective map*

$$w_A : S^+ \to S^+$$

by $w_A(x) = \frac{Ax}{\|Ax\|_1}$, as shown in Figure 2.11. Define the *projective set* for Σ by

$$\Sigma_p = \{w_A : A \in \Sigma\}.$$

Then for any compact subset of U of S^+,

$$\Sigma_p U = \{w_A(x) : w_A \in \Sigma_p \text{ and } x \in U\}.$$

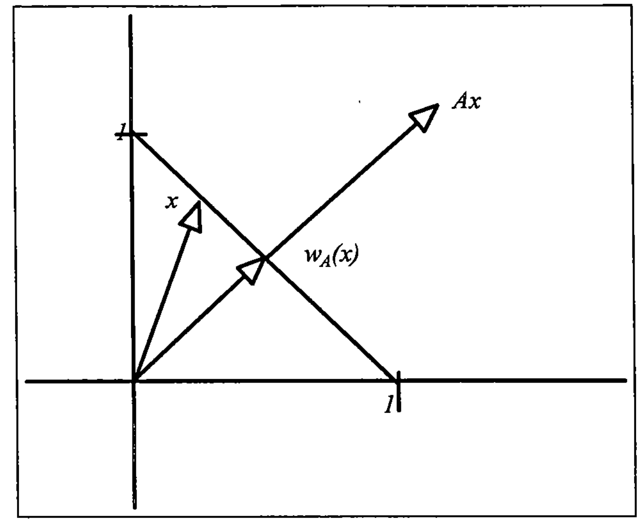

FIGURE 2.11. A view of w_A.

Thus, $\Sigma_p U$ is the projection of ΣU onto S^+. Since Σ and U are compact, so is ΣU. And thus, $\Sigma_p U$ is compact.

Now using the metric p on S^+, define

$$\begin{aligned} \tau(\Sigma) &= \sup_{U,V} \frac{h(\Sigma_p U, \Sigma_p V)}{h(U,V)} \\ &= \sup_{U,V} \frac{h(\Sigma U, \Sigma V)}{h(U,V)} \end{aligned}$$

where the sup is over all compact subsets U and V of S^+.

Using the notation described above, we have the following.

Theorem 2.9 $\tau(\Sigma) \le \max_{A\in\Sigma} \tau_B(A)$.

Proof. Let U and V be compact subsets of S^+. Then

$$\begin{aligned} \delta(\Sigma_p U, \Sigma_p V) &= \max_{Au\in\Sigma U} p(Au, \Sigma V) \\ &= p\left(\hat{A}\hat{u}, \Sigma V\right) \end{aligned}$$

for some $\hat{A}\hat{u} \in \Sigma U$. So

$$\delta(\Sigma_p U, \Sigma_p V) \le p\left(\hat{A}\hat{u}, \hat{A}\hat{v}\right)$$

where $\hat{v}$ satisfies $p(\hat{u}, \hat{v}) = p(\hat{u}, V)$. Thus,

$$\begin{aligned}\delta(\Sigma_p U, \Sigma_p V) &\leq \tau_B\left(\hat{A}\right) p(\hat{u}, \hat{v}) \\ &= \tau_B\left(\hat{A}\right) p(\hat{u}, V) \\ &\leq \tau_B\left(\hat{A}\right) \delta(U, V).\end{aligned}$$

Similarly,

$$\delta(\Sigma_p V, \Sigma_p U) \leq \tau_B\left(\tilde{A}\right) \delta(V, U)$$

for some $\tilde{A} \in \Sigma$. Thus,

$$h(\Sigma_p U, \Sigma_p V) \leq \max_{A \in \Sigma} \tau_B(A)\, h(U, V)$$

and so

$$\tau(\Sigma) \leq \max_{A \in \Sigma} \tau_B(A),$$

which is what we need to show. ■

Equality, in the theorem, need not hold. To see this, let

$$\Sigma = \quad \{A : A \text{ is a column stochastic } 2 \times 2 \text{ matrix with } \tfrac{1}{3} \leq a_{ij} \leq \tfrac{2}{3} \text{ for all } i, j\}.$$

If $x \in S^+$ and $A \in \Sigma$, then $\frac{1}{3} \leq (Ax)_i \leq \frac{2}{3}$ for all i, and so $\hat{A} = [Ax\, Ax] \in \Sigma$. Thus, for $y \in S^+, Ax = \hat{A}y$, which can be used to show $\tau(\Sigma) = 0$. Yet $\bar{A} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix} \in \Sigma$ and $\tau_B\left(\bar{A}\right) > 0$.

We define

$$\tau_B(\Sigma) = \max_{A \in \Sigma} \tau_B(A).$$

And, we have a corollary parallel to Corollary 2.2.

Corollary 2.3 *If V is a compact subset of S^+, where $\Sigma_p V = V$, then for any compact subset U of S^+,*

$$h\left(\Sigma_p^k U, V\right) \leq \tau_B(\Sigma)^k\, h(U, V).$$

This corollary shows that if we project the sequence

$$\Sigma U,\ \Sigma^2 U,\ \Sigma^3 U, \ldots$$

into S^+ to obtain

$$\Sigma_p U,\ \Sigma_p^2 U,\ \Sigma_p^3 U, \ldots$$

then if $\tau_B(\Sigma) < 1$, this sequence converges to V in the Hausdorff metric.

2.2.2 Subspace Contraction Coefficient

We now develop a contraction coefficient for a subspace of F^n. When this setting arises in applications, row vectors rather than column vectors are usually used. Thus, in this subsection F^n will denote row vectors.

To develop this contraction coefficient, we let A be an $n \times n$ matrix and E an $n \times k$ full column rank matrix. Further, we suppose that there is a $k \times k$ matrix M such that

$$AE = EM.$$

Now extend the columns of E to a basis and use this basis to form

$$P = [E\ G].$$

Partition

$$P^{-1} = \begin{bmatrix} H \\ J \end{bmatrix}$$

where H is $k \times n$. Then we have

$$\begin{aligned} AP &= A[E\ G] \\ &= \begin{bmatrix} E\ G \end{bmatrix} \begin{bmatrix} M & C \\ 0 & N \end{bmatrix} \end{aligned} \tag{2.5}$$

where

$$\begin{bmatrix} C \\ N \end{bmatrix} = P^{-1}AG.$$

Now set

$$W = \{x \in F^n : xE = 0\}.$$

Then W is a subspace and if $x \in W$, $xA \in W$ as well. Thus, for any vector norm $\|\cdot\|$, we can define

$$\begin{aligned}\tau_W(A) &= \max_{\substack{x\in W\\ \|x\|=1}} \|xA\| \\ &= \max_{\substack{x\in W\\ x\neq 0}} \frac{\|xA\|}{\|x\|}.\end{aligned}$$

Notice that from the definition,

$$\|xA\| \le \tau_W(A)\,\|x\|$$

for all $x \in W$. Thus, if $\tau_W(A) = \frac{1}{2}$, then A contracts the subspace W by at least $\frac{1}{2}$. So, a circle of radius r ends up in a circle of radius $\frac{1}{2}r$, or less, as shown in Figure 2.12.

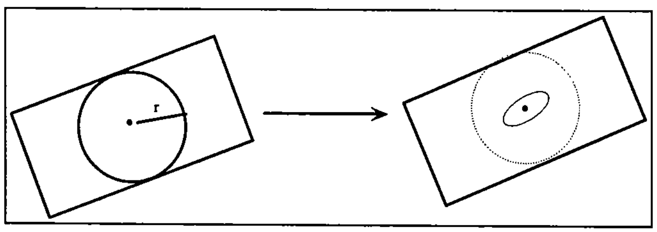

FIGURE 2.12. A view of $\tau_W(A) = \frac{1}{2}$.

If B is an $n \times n$ matrix such that $BE = E\hat{M}$ for some $k \times k$ matrix $\hat{M}$, then for any $x \in W$,

$$\|xAB\| \le \tau_W(B)\,\|xA\| \le \tau_W(A)\,\tau_W(B)\,\|x\|.$$

Thus,

$$\tau_W(AB) \le \tau_W(A)\,\tau_W(B).$$

We now link $\tau_W(A)$ to N given in (2.5). To do this, define on F^{n-k}

$$\|z\|_J = \|zJ\|.$$

It is easily seen that $\|\cdot\|_J$ is a norm.

We now show that $\tau_W(A)$ is actually $\|N\|_J$.

Theorem 2.10 *Using the partition in (2.5), $\tau_W(A) = \|N\|_J$.*

Proof. We first show that $\tau_W(A) \le \|N\|_J$. For this, let x be a vector such that $xE = 0$. Then

$$\begin{aligned}\|xA\| &= \left\| x\ [E\ G] \begin{bmatrix} M & C \\ 0 & N \end{bmatrix} \begin{bmatrix} H \\ J \end{bmatrix} \right\| \\ &= \left\| [0\ xG] \begin{bmatrix} M & C \\ 0 & N \end{bmatrix} \begin{bmatrix} H \\ J \end{bmatrix} \right\| \\ &= \left\| [0\ xGN] \begin{bmatrix} H \\ J \end{bmatrix} \right\| \\ &= \|xGNJ\| \\ &= \|xGN\|_J \\ &\le \|xG\|_J\,\|N\|_J. \end{aligned} \tag{2.6}$$

Now

$$\begin{aligned}\|xG\|_J = \|xGJ\| &= \left\| [0\ xG] \begin{bmatrix} H \\ J \end{bmatrix} \right\| \\ &= \left\| x\,[EG] \begin{bmatrix} H \\ J \end{bmatrix} \right\| \\ &= \|x\|.\end{aligned}$$

Thus, plugging into (2.6) yields

$$\|xA\| \le \|N\|_J\,\|x\|.$$

And, since this holds for all $x \in W$,

$$\tau_W(A) \le \|N\|_J.$$

We now show that $\|N\|_J \le \tau_W(A)$. To do this, let $z \in F^{n-k}$ be such that $\|z\|_J = 1$ and $\|N\|_J = \|zN\|_J$.

Now,

$$\begin{aligned}\|N\|_J &= \|zN\|_J \\ &= \|zNJ\| \\ &= \left\| (0,z) \begin{bmatrix} M & C \\ 0 & N \end{bmatrix} \begin{bmatrix} H \\ J \end{bmatrix} \right\| \\ &= \|zJA\| \\ &\le \|zJ\|\,\tau_W(A) \\ &\le \|z\|_J\,\tau_W(A) \\ &= \tau_W(A),\end{aligned}$$

which gives the theorem. ■

A converse of this theorem follows.

Theorem 2.11 *Let A be an $n \times n$ matrix and P an $n \times n$ matrix such that*

$$P^{-1}AP = \begin{bmatrix} M & C \\ 0 & N \end{bmatrix},$$

where M is $k \times k$. Let $\|\cdot\|$ be any norm on F^{n-k}. Then there is a norm $\|\cdot\|_G$ on F^n and thus on W, such that $\tau_W(A) = \|N\|$.

Proof. We assume P and P^{-1} are partitioned as in (2.5) and use the notation given there. We first find a norm on

$$W = \{x : xE = 0\}.$$

For this, if $x \in W$, define

$$\|x\|_G = \|xG\|.$$

To see that $\|\cdot\|_G$ is a norm, let $\|x\|_G = 0$. Then $\|xG\| = 0$, so $xG = 0$. Since $x \in W$, $xP = 0$ and so $x = 0$. The remaining properties assuring that $\|\cdot\|_G$ is a norm are easily established.

Now, extend $\|\cdot\|_G$ to a norm, say $\|\cdot\|_G$, on F^n. We show the contraction coefficient τ_W, determined from this norm, is such that $\tau_W(A) = \|N\|$. Using the norm and part of the proof of the previous theorem, recall that if $z \in F^{n-k}$,

$$\|z\|_J = \|zJ\|_G = \|zJG\| = \|zI\| = \|z\|.$$

Thus, $\|N\|_J = \|N\|$ and hence

$$\tau_W(A) = \|N\|,$$

as required. ■

Formulas for computing $\tau_W(A)$ depend on the vector norm used as well as on E. We restrict our work now to R^n so that we can use convex polytopes. If the vector norm, say $\|\cdot\|$, has a unit ball which is a convex polytope, that is

$$K = \{x \in R^n : xE = 0 \text{ and } \|x\| \le 1\}$$

is a convex polytope, then a formula, in terms of the vertices of this convex polytope, can be found. Using this notation, we have the following.

Theorem 2.12 *Let A be an $n \times n$ matrix and $\|\cdot\|$ a vector norm on R^n that produces a unit ball which is a convex polytope K in W. If $\{v_1, \dots, v_s\}$ are the vertices of K, then*

$$\tau_W(A) = \max_i \{\|v_iA\|\}.$$

Proof. Let $x \in K$ where $\|x\| = 1$. Write

$$x = \alpha_1 v_1 + \cdots + \alpha_s v_s$$

a convex combination of $v_1, \dots, v_s$. Then it follows that

$$\begin{aligned} \|xA\| &= \left\| \sum_{i=1}^{s} \alpha_i v_i A \right\| \\ &\le \sum_{i=1}^{s} \alpha_i \|v_iA\| \\ &\le \max_i \{\|v_iA\|\}. \end{aligned}$$

Thus,

$$\tau_W(A) \le \max_i \{\|v_iA\|\}.$$

That equality holds can be seen by noting that no vertex can be interior to the unit ball. Thus, $\|v_i\| = 1$ for all i, so $\max_{\substack{x \in W \\ \|x\|=1}} \|xA\|$ is achieved at a vertex. ■

We will give several examples of computing $\tau_W(A)$, for various W, in Chapter 11. For now, we look at a classical result.

An $n \times n$ nonnegative matrix A is stochastic if each of its rows is stochastic. Note that in this case

$$Ae = e,$$

where $e = (1, 1, \dots, 1)^t$, so we can set $E = e$. Then using the 1-norm,

$$K = \{x \in R^n : xe = 0 \text{ and } \|x\|_1 \le 1\}.$$

The vertices of this set are those vectors having precisely two nonzero entries, namely $\frac{1}{2}$ and $-\frac{1}{2}$. Thus,

$$\tau_W(A) = \max_{i \ne j} \left\| \frac{1}{2}a_i - \frac{1}{2}a_j \right\|_1$$

where a_k denotes the k-th row of A. Written in the classical way,

$$\tau_1(A) = \frac{1}{2}\max_{i \neq j} \|a_i - a_j\|_1,$$

is called the *coefficient of ergodicity for stochastic matrices.*

To conclude this subsection, we show how to describe subspace coefficients on the compact subsets. To do this we suppose that Σ is a compact matrix set and that if $A \in \Sigma$, then A has partitioned form as given in (2.5). Let $S \subseteq F^n$ such that if $x, y \in S$, then $x - y \in W$ (a subset of a translate of W). We define

$$\tau(\Sigma) = \max_{R \neq S} \frac{h(R\Sigma, T\Sigma)}{h(R,T)}$$

where the maximum is over all compact subsets R and T in S.

A bound on $\tau(\Sigma)$ follows.

Theorem 2.13 $\tau(\Sigma) \leq \max_{A \in \Sigma} \tau_W(A)$.

Proof. The proof is as in Theorem 2.9. ■

We now define

$$\tau_W(\Sigma) = \max_{A \in \Sigma} \tau_W(A).$$

2.2.3 *Blocking*

In applications of products of matrices, we need the required contraction coefficient to be less than 1. However, we often find a larger coefficient. How this is usually resolved is to use products of matrices of a specified length, called *blocks.* For any matrix set Σ, an *r-block* is defined as any product π in Σ^r.

We now prove a rather general, and useful, theorem.

Theorem 2.14 *Let τ be a contraction coefficient (either τ_B or τ_W) for a matrix set Σ. Suppose $\tau(\pi) \leq \tau_r$ for some constant $\tau_r < 1$ and all r-blocks π of Σ, and that $\tau(\Sigma) \leq \beta$ for some constant β.*

When all products are taken from Σ, we have the following. If $\tau = \tau_B$, then $\tau(A_1), \tau(A_2A_1), \tau(A_3A_2A_1), \ldots$ converges to 0. If $\tau = \tau_W$, then $\tau(A_1), \tau(A_1A_2), \tau(A_1A_2A_3), \ldots$ converges to 0. And, both have rate of convergence $\tau_r^{\frac{1}{r}}$.

Proof. We prove the result for τ_B. Let $A_1, A_2A_1, A_3A_2A_1, \ldots$ be a sequence of products taken from Σ. Partition, as possible, each product $A_k \ldots A_1$ in the sequence into r-blocks,

$$\pi_s \cdots \pi_1 A_t \cdots A_1$$

where $k = sr + t, t < r$, and $\pi_1, \ldots, \pi_s$ are r-blocks. Then

$$\begin{aligned} &\tau(\pi_s \cdots \pi_1 A_t \cdots A_1) \\ &\leq \tau(\pi_s) \cdots \tau(\pi_1)\tau(A_t) \cdots \tau(A_1) \\ &\leq \tau_r^s \beta^t. \end{aligned}$$

Thus $\tau(A_k \cdots A_1) \to 0$ as $k \to \infty$.

Concerning the geometric rate, note that for $\tau_r > 0$,

$$\begin{aligned} \tau_r^s &= \tau_r^{\frac{k}{r} - \frac{t}{r}} \\ &\leq \tau_r^{\frac{k}{r}} \tau_r^{-\frac{t}{r}} \\ &\leq \tau_r^{\frac{k}{r}} \tau_r^{-1} \\ &= \tau_r^{-1} \left(\tau_r^{\frac{1}{r}}\right)^k. \end{aligned}$$

Thus,

$$\tau(A_k \cdots A_1) \leq \tau_r^{-1} \left(\tau_r^{\frac{1}{r}}\right)^k,$$

which shows that the rate is geometric. ■

2.3 Measures of Irreducibility and Full Indecomposability

Measures give an indication of how the nonzero entries in a matrix are distributed within that matrix. In this section, we look at two such measures.

For the first measure, let A be an $n \times n$ nonnegative matrix. We say that A is *reducible* if there is a 0-submatrix, say in rows numbered $r_1, \ldots, r_s$ and columns numbered $c_1, \ldots, c_{n-s}$, where $r_1, \ldots, r_s, c_1, \ldots, c_{n-s}$ are distinct. (Thus, a 1×1 matrix A is reducible iff $a_{11} = 0$.) For example

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 4 & 0 \\ 3 & 0 & 2 \end{bmatrix}$$

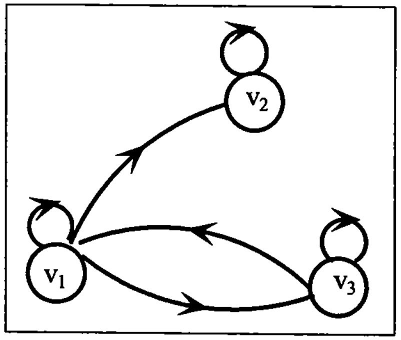

FIGURE 2.13. The graph of A.

is reducible since in row 2 and columns 1 and 3, there is a 0-submatrix.

If P is a permutation matrix that moves rows $r_1, \ldots, r_s$ into rows $1, \ldots, s$, then

$$PAP^t = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}$$

where A_{11} is $s \times s$. In the example above,

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

An $n \times n$ nonnegative matrix A is *irreducible*, if it is not reducible.

As shown in Varga (1962), a *directed graph* can be associated with A by using vertices $v_1, \ldots, v_n$ and defining an arc from v_i to v_j if $a_{ij} > 0$. Thus for our example, we have Figure 2.13. And, A is irreducible if and only if there is a path (directed), of positive length, from any v_i to any v_j. Note that in our example, there is no path from v_2 to v_3, so A is reducible.

A measure, called a *measure of irreducibility*, is defined on an $n \times n$ nonnegative matrix A, $n > 1$, as

$$u(A) = \min_R \left(\max_{\substack{i \in R \\ j \in R'}} a_{ij} \right)$$

where R is a nonempty proper subset of $\{1, \ldots, n\}$ and R' its compliment. This measure tells how far A is from being reducible.

For the second measure, we say that an $n \times n$ nonnegative matrix A is *partly decomposable* if there is a 0-submatrix, say in rows numbered $r_1, \ldots, r_s$ and columns numbered $c_1, \ldots, c_{n-s}$. (Thus a 1×1 matrix A is partly decomposable iff $a_{11} = 0$.) For example,

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 4 & 0 & 3 \\ 2 & 0 & 1 \end{bmatrix}$$

is partly decomposable since there is a 0-submatrix in rows 2 and 3, and column 2.

If we let P and Q be $n \times n$ permutation matrices such that P permutes rows $r_1, \ldots, r_s$ into rows $1, \ldots, s$ and Q permutes columns $c_1, \ldots, c_{n-s}$ into columns $s+1, \ldots, n$, then

$$PAQ = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}$$

where A_{11} is $s \times s$.

An $n \times n$ nonnegative matrix A is *fully indecomposable* if it is not partly decomposable. Thus, A is fully indecomposable iff whenever A contains a $p \times q$ 0-submatrix, then $p + q \leq n - 1$.

There is a link between irreducible matrices and fully indecomposable matrices. As shown in Brualdi and Ryser (1991), A is irreducible iff $A + I$ is fully indecomposable.

A *measure of full indecomposability* can be defined as

$$U(A) = \min_{S,T} \left(\max_{\substack{S,T \\ i \in S, j \in T}} a_{ij} \right)$$

where $S = \{r_1, \ldots, r_s\}$ and $T = \{c_1, \ldots, c_{n-s}\}$ are nonempty proper subsets of $\{1, \ldots, n\}$.

We now show a few basic results about fully indecomposable matrices.

Theorem 2.15 *Let A and B be $n \times n$ nonnegative fully indecomposable matrices. Suppose that the largest 0-submatrices in A and B are $s_A \times t_A$ and $s_B \times t_B$, respectively. If*

$$s_A + t_A = n - k_A$$
$$s_B + t_B = n - k_B,$$

then the largest 0-submatrix, say a $p \times q$ submatrix, in AB satisfies

$$p + q \leq n - k_A - k_B.$$

Proof. Suppose P and Q are $n \times n$ permutations such that

$$P(AB)Q = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

where C_{12} is $p \times q$ and the largest 0-submatrix in AB.

Let R be an $n \times n$ permutation matrix such that

$$PAR = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}$$

where A_{11} is $p \times s$ and has no 0 columns.

Partition

$$R^t BQ = \begin{bmatrix} B_{11} & B_{12} \\ B_{12} & B_{22} \end{bmatrix}$$

where B_{11} is $s \times (n-q)$. Thus, we have

$$\begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{bmatrix}.$$

Now,

$$A_{11}B_{12} = 0$$

and since A_{11} has no 0 columns

$$B_{12} = 0.$$

Thus, $s + q \leq n - k_B$. And, using A, $p + (n-s) \leq n - k_A$, so

$$\begin{aligned} p + q &\leq (s - k_A) + (n - k_B - s) \\ &\leq n - k_A - k_B, \end{aligned}$$

the desired result. ■

Several corollaries are immediate.

Corollary 2.4 *Let A and B be $n \times n$ fully indecomposable matrices. Then AB is fully indecomposable.*

Proof. If AB contains a $p \times q$ 0-submatrix, then by the theorem, $p + q \leq n - 1 - 1 = n - 2$. Thus, AB is fully indecomposable, as was to be shown. ■

Corollary 2.5 *Let $A_1, \ldots, A_{n-1}$ be $n \times n$ fully indecomposable matrices. Then $A_1 \cdots A_{n-1}$ is positive.*

Proof. Note that $k_{A_1 A_2} \geq k_{A_1} + k_{A_2}$. And

$$\begin{aligned} k_{A_1 \cdots A_{n-1}} &\geq k_{A_1} + \cdots + k_{A_{n-1}} \\ &\geq 1 + \cdots + 1 \\ &= n - 1. \end{aligned}$$

Thus, if $A_1 \cdots A_{n-1}$ has a $p \times q$ 0-submatrix, then

$$\begin{aligned} p + q &\leq n - k_{A_1 \ldots A_{n-1}} \\ &\leq n - (n-1) \\ &= 1. \end{aligned}$$

This inequality cannot hold, hence $A_1 \cdots A_{n-1}$ can contain no 0-submatrix. The result follows. ■

The measure of full indecomposability can also be seen as giving some information about the distribution of the sizes of the entries in a product of matrices.

Theorem 2.16 *Let A and B be $n \times n$ fully indecomposable matrices. Then*

$$U(AB) \geq U(A)\,U(B).$$

Proof. Construct $\hat{A} = [\hat{a}_{ij}]$ where

$$\hat{a}_{ij} = \begin{cases} 0 \text{ if } a_{ij} < U(A) \\ a_{ij} \text{ otherwise.} \end{cases}$$

Construct $\hat{B}$ in the same way. Then both $\hat{A}$ and $\hat{B}$ are fully indecomposable. Thus $\hat{A}\hat{B}$ is fully indecomposable, and so we have that $U\left(\hat{A}\hat{B}\right) > 0$.

If $\left(\hat{A}\hat{B}\right)_{ij} > 0$, then $\left(\hat{A}\hat{B}\right)_{ij} \geq U(A)\,U(B)$. Hence,

$$U(AB) \geq U(A)\,U(B),$$

the indicated result. ■

An immediate corollary follows.

Corollary 2.6 *If $A_1, \ldots, A_{n-1}$ are $n \times n$ fully indecomposable matrices, then*

$$(A_1 \cdots A_{n-1})_{ij} \geq U(A_1) \cdots U(A_{n-1})$$

for all i and j.

2.4 Spectral Radius

Recall that for an $n \times n$ matrix A, the *spectral radius* $\rho(A)$ of A is

$$\rho(A) = \max_{\lambda} \{|\lambda| : \lambda \text{ is an eigenvalue of } A\}.$$

It is easily seen that

$$\rho(A) = \left(\rho\left(A^k\right)\right)^{\frac{1}{k}} \tag{2.7}$$

and that for any matrix norm $\|\cdot\|$

$$\rho(A) = \lim_{k\to\infty} \left\|A^k\right\|^{\frac{1}{k}}. \tag{2.8}$$

In this section, we use both (2.7) and (2.8) to generalize the notion of spectral radius to a bounded matrix set Σ.

To generalize (2.7), let

$$\rho_k(\Sigma) = \sup\left\{\rho\left(\prod_{i=1}^{k} A_i\right) : A_i \in \Sigma \text{ for all } i\right\}.$$

The *generalized spectral radius* of Σ is

$$\rho(\Sigma) = \lim_{k\to\infty} \sup\left(\rho_k(\Sigma)\right)^{\frac{1}{k}}.$$

To generalize (2.8), let $\|\cdot\|$ a matrix norm and define

$$\hat{\rho}_k(\Sigma, \|\cdot\|) = \sup\left\{\left\|\prod_{i=1}^{k} A_i\right\| : A_i \in \Sigma \text{ for all } i\right\}.$$

The *joint spectral radius* is

$$\hat{\rho}(\Sigma, \|\cdot\|) = \lim_{k\to\infty} \sup\left\{\hat{\rho}_k(\Sigma, \|\cdot\|)^{\frac{1}{k}}\right\}.$$

Note that if $\|\cdot\|_a$ is another matrix norm, then since norms are equivalent, there are positive constants α and β such that

$$\alpha \|A\|_a \leq \|A\| \leq \beta \|A\|_a$$

for all $n \times n$ matrices A. Thus,

$$\alpha^{\frac{1}{k}} \left\|\prod_{i=1}^{k} A_i\right\|_a^{\frac{1}{k}} \leq \left\|\prod_{i=1}^{k} A_i\right\|^{\frac{1}{k}} \leq \beta^{\frac{1}{k}} \left\|\prod_{i=1}^{k} A_i\right\|_a^{\frac{1}{k}}$$

and so

$$\hat{\rho}(\Sigma, \|\cdot\|_a) = \hat{\rho}(\Sigma, \|\cdot\|).$$

Hence, the value $\hat{\rho}(\Sigma, \|\cdot\|)$ does not depend on the matrix norm used, and we can write $\hat{\rho}(\Sigma)$ for $\hat{\rho}(\Sigma, \|\cdot\|)$. In addition, if the set Σ used in $\hat{\rho}(\Sigma)$ is clear from context, we simply write $\hat{\rho}$ for $\hat{\rho}(\Sigma)$.

We can also show that if P is an $n \times n$ nonsingular matrix and we define

$$P\Sigma P^{-1} = \left\{PAP^{-1} : A \in \Sigma\right\}$$

then

$$\rho\left(P\Sigma P^{-1}\right) = \rho(\Sigma).$$

Further, for any matrix norm $\|\cdot\|$, $\|A\|_P = \left\|PAP^{-1}\right\|$ is a matrix norm. Thus $\hat{\rho}\left(P\Sigma P^{-1}\right) = \hat{\rho}(\Sigma)$. So both ρ and $\hat{\rho}$ are invariant under similarity transformations.

Our first result links the generalized spectral radius and the joint spectral radius.

Lemma 2.3 *For any matrix norm, on a bounded matrix set Σ,*

$$\rho_k(\Sigma)^{\frac{1}{k}} \leq \rho(\Sigma) \leq \hat{\rho}(\Sigma) \leq \hat{\rho}_k(\Sigma)^{\frac{1}{k}}.$$

Proof. To prove the first inequality, note that for any positive integer m,

$$\rho_k(\Sigma)^m \leq \rho_{mk}(\Sigma).$$

Thus, taking the mk-th roots,

$$\rho_k(\Sigma)^{\frac{1}{k}} \leq \rho_{mk}(\Sigma)^{\frac{1}{mk}}.$$

Now computing lim sup, as $m \to \infty$, of the right side, we have the first inequality.

The second inequality follows by observing that

$$\rho\left(A_{i_k} \cdots A_{i_1}\right) \leq \left\|A_{i_k} \cdots A_{i_1}\right\|$$

for any matrices $A_{i_1}, \ldots, A_{i_k}$.

For the third inequality, let l be a positive integer and write

$$l = kq + r$$

where $0 \le r < k$. Note that for any product of l matrices from Σ,

$$\begin{aligned}\|A_l \cdots A_1\| &\le \|A_{kq+r} \cdots A_{kq+1} A_{kq} \cdots A_1\| \\ &\le \beta^r \left\|A_{kq} \cdots A_{(k-1)q+1} \cdots A_q \cdots A_1\right\| \\ &\le \beta^r \hat{\rho}_k \left(\Sigma^q\right)\end{aligned}$$

where β is a bound on the matrices in Σ. Thus,

$$\begin{aligned}\hat{\rho}_l (\Sigma)^{\frac{1}{l}} &\le \beta^{\frac{r}{l}} \hat{\rho}_k (\Sigma)^{\frac{q}{l}} \\ &= \beta^{\frac{r}{l}} \hat{\rho}_k (\Sigma)^{\frac{-r}{kl}} \hat{\rho}_k (\Sigma)^{\frac{1}{k}} \\ &= \left(\beta^r \hat{\rho}_k (\Sigma)^{-\frac{r}{k}}\right)^{\frac{1}{l}} \hat{\rho}_k (\Sigma)^{\frac{1}{k}}.\end{aligned}$$

Now, computing lim sup as $l \to \infty$, we have

$$\hat{\rho} (\Sigma) \le \hat{\rho}_k (\Sigma)^{\frac{1}{k}}$$

as required. ■

Using this lemma, we have simpler expressions for $\rho (\Sigma)$ and $\hat{\rho} (\Sigma)$.

Theorem 2.17 *If Σ is a bounded matrix set, then we have that $\rho (\Sigma) = \lim_{k\to\infty} \rho_k (\Sigma)^{\frac{1}{k}}$ and $\hat{\rho} (\Sigma) = \lim_{k\to\infty} \hat{\rho}_k (\Sigma)^{\frac{1}{k}}$.*

Proof. We prove the second inequality. By the lemma, we have

$$\hat{\rho} (\Sigma) \le \hat{\rho}_k (\Sigma)^{\frac{1}{k}}$$

for all k. Thus, for any k,

$$\hat{\rho} (\Sigma) \le \inf_{j\ge k} \hat{\rho}_j (\Sigma)^{\frac{1}{j}} \le \sup_{j\ge k} \hat{\rho}_j (\Sigma)^{\frac{1}{j}}$$

from which it follows that

$$\hat{\rho} (\Sigma) \le \lim_{k\to\infty} \inf \hat{\rho}_k (\Sigma)^{\frac{1}{k}} \le \lim_{k\to\infty} \sup \hat{\rho}_k (\Sigma)^{\frac{1}{k}} = \hat{\rho} (\Sigma).$$

So,

$$\lim_{k\to\infty} \hat{\rho}_k (\Sigma)^{\frac{1}{k}} = \hat{\rho} (\Sigma),$$

the desired result. ■

Berger and Wang (1995), in a rather long argument, showed that for bounded sets Σ, $\rho (\Sigma) = \hat{\rho} (\Sigma)$. However, we will not develop this relationship since we use the traditional $\hat{\rho}$, over ρ, in our work.

We now give a few results on the size of $\hat{\rho}$. A rather obvious such bound follows.

Theorem 2.18 *If Σ is a bounded set of $n \times n$ matrices, then $\hat{\rho} (\Sigma) \le \sup_{A\in\Sigma} \|A\|$.*

For the remaining result, we observe that if Σ is a product bounded matrix set, then a vector norm $\|\cdot\|_v$ can be defined from any vector norm $\|\cdot\|$ by

$$\|x\|_v = \sup_{l\ge 0} \left\{\|A_{i_l} \dots A_{i_1} x\| : A_{i_l} \dots A_{i_1} \in \Sigma\right\}$$

(when $l = 0$, $\|A_{i_l} \dots A_{i_1} x\| = \|x\|$). Using this vector norm, we can see that if $A \in \Sigma$, then

$$\|Ax\|_v \le \|x\|_v$$

for all x. Thus we have the following.

Lemma 2.4 *If Σ is a product bounded matrix set, then there is a vector norm $\|\cdot\|_v$ such that for the induced matrix norm,*

$$\|A\|_v \le 1.$$

This lemma provides a last result involving an expression for $\hat{\rho} (\Sigma)$.

Theorem 2.19 *If Σ is a bounded matrix set,*

$$\hat{\rho} (\Sigma) = \inf_{\|\cdot\|} \sup_{A\in\Sigma} \|A\|.$$

Proof. Let $\epsilon > 0$ and define

$$\hat{\Sigma} = \left\{\frac{1}{\hat{\rho} + \epsilon} A : A \in \Sigma\right\}.$$

Then $\hat{\Sigma}$ is product bounded since, if $B \in \hat{\Sigma}^k$,

$$\|B\| \le \frac{1}{(\hat{\rho} + \epsilon)^k} \hat{\rho}_k.$$

Note that $\frac{1}{(\hat{\rho}+\epsilon)^k} \hat{\rho}_k \to 0$ as $k \to \infty$.

Thus, by Lemma 2.4, there is a norm $\|\cdot\|_b$ such that

$$\|C\|_b \leq 1$$

for all $C \in \hat{\Sigma}$.

Now, if $A \in \Sigma$, $\frac{1}{\hat{\rho}+\epsilon} A \in \hat{\Sigma}$, so

$$\|A\|_b \leq \hat{\rho} + \epsilon .$$

Thus,

$$\inf_{\|\cdot\|} \sup_{A \in \Sigma} \|A\| \leq \hat{\rho} + \epsilon ,$$

and since ϵ was arbitrary,

$$\inf_{\|\cdot\|} \sup_{A \in \Sigma} \|A\| \leq \hat{\rho}.$$

Finally by Theorem 2.18,

$$\inf_{\|\cdot\|} \sup_{A \in \Sigma} \|A\| \geq \hat{\rho}.$$

The result follows from the last two inequalities. ■

2.5 Research Notes

Some material on the projective metric can be found in Bushell (1973), Golubinsky, Keller and Rothchild (1975) and in the book by Seneta (1981). Geometric discussions can be found in Bushell (1973) and Golubinsky, Keller and Rothchild. Artzrouni (1996) gave the inequalities that appeared in Theorem 2.2.

A source for basic work on the Hausdorff metric is a book by Eggleston (1969).

Birkhoff (1967) developed the expression for τ_B. A proof can also be found in Seneta (1981). Arzrouni and Li (1995) provided a 'simple' proof for this result. Bushell (1973) showed that $\left((R^n)^+ \cap U, p\right)$, where U is the unit sphere, was a complete metric space. Altham (1970) discussed measurements in general.

Much of the work on τ_W in subsection 2 is based on Hartfiel and Rothblum (1998). However, special such topics have been studied by numerous

authors. Seneta (1981) as well as Rothblum and Tan (1985) showed that for a positive stochastic matrix A, $\tau_B(A) \geq \tau_1(A)$ where $\tau_1(A)$ is the subspace contractive coefficient with $E = (1, 1, \ldots, 1)^t$. More recently, Rhodius (2000) considered contraction coefficients for infinite stochastic matrices. It should be noted that these authors call contraction coefficients, *coefficients of ergodicity*.

General work on measures for irreducibility and full indecomposability were given by Hartfiel (1975). Christian (1979) also contributed to that area. Hartfiel (1973) used measures to compute bounds on eigenvalues and eigenvectors.

Rota and Strang (1960) introduced the joint spectral radius $\hat{\rho}$, while Daubechies and Lagaries (1992) gave the generalized spectral radius ρ. Lemma 2.3 was also done by those authors. Berger and Wang (1992) proved that $\rho(\Sigma) = \hat{\rho}(\Sigma)$, as long as Σ is bounded. This theorem was also proved by Elsner (1995) by different techniques. Beyn and Elsner (1997) proved Lemma 2.4 and Theorem 2.19. Some of this work is implicit in the paper by Rota and Strang.

3

Semigroups of Matrices

Let Σ be a product bounded matrix set. A *matrix sequence* of the sequence $\langle \Sigma^k \rangle$ is a sequence $\pi_1, \pi_2, \ldots$ of products taken from $\Sigma, \Sigma^2, \ldots$, respectively. A *matrix subsequence* is a subsequence of a *matrix sequence*.

The *limiting set*, Σ^∞, of the sequence $\langle \Sigma^k \rangle$ is defined as

$$\Sigma^\infty = \{A : A \text{ is the limit of a matrix subsequence of } \langle \Sigma^k \rangle\}.$$

Two examples may help with understanding these notions.

Example 3.1 *Let* $\Sigma = \left\{ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}$. *Then* $\lim_{k \to \infty} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^k$ *does not exist. However,* $\Sigma^\infty = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}$.

Example 3.2 *Let* $\Sigma = \left\{ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \right\}$. *Then we can show that* $\Sigma^\infty = \left\{ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \right\}$.

As we will see in Chapter 11, limiting sets can be much more complicated.

3.1 Limiting Sets

This section describes various properties of limiting sets.

3.1.1 *Algebraic Properties*

Some algebraic properties of a limiting set Σ^∞ are given below.

Theorem 3.1 *If Σ is product bounded, then Σ^∞ is a compact semigroup.*

Proof. To show that Σ^∞ is a semigroup, let $A, B \in \Sigma^\infty$. Then there are matrix subsequences of $\langle \Sigma^k \rangle$, say

$$\pi_{i_1}, \pi_{i_2}, \ldots$$
$$\hat{\pi}_{j_1}, \hat{\pi}_{j_2}, \ldots$$

that converge to A and B, respectively. The sequence

$$\pi_{i_1}\hat{\pi}_{j_1}, \pi_{i_2}\hat{\pi}_{j_2}, \ldots$$

is a matrix subsequence of $\langle \Sigma^k \rangle$ which converges to AB. Thus, $AB \in \Sigma^\infty$ and since A and B were arbitrary, Σ^∞ is a semigroup.

The proof that Σ^∞ is topologically closed is a standard proof, and since Σ is product bounded, Σ^∞ is bounded. Thus, Σ^∞ is a compact set. ■

A product result about Σ^∞ follows.

Theorem 3.2 *If Σ is product bounded, then $\Sigma^\infty\Sigma^\infty = \Sigma^\infty$.*

Proof. Since Σ^∞ is a semigroup, $\Sigma^\infty\Sigma^\infty \subseteq \Sigma^\infty$. To show equality holds, let $A \in \Sigma^\infty$ and $\pi_1, \pi_2, \ldots$ a matrix subsequence of $\langle \Sigma^k \rangle$ that converges to A. If π_k has l_k factors, $k > 1$, factor

$$\pi_k = B_k C_k$$

where B_k contains the first $[l_k/2]$ factors of π_k and C_k the remaining factors.

Since Σ is product bounded, the sequence $B_1, B_2, \ldots$ has a convergent matrix subsequence $B_{i_1}, B_{i_2}, \ldots$ which converges to, say, B. Since $C_{i_1}, C_{i_2}, \ldots$ is bounded, it has a convergent subsequence, say, $C_{j_1}, C_{j_2}, \ldots$ which converges to, say, C. Thus $\pi_{j_1}, \pi_{j_2}, \ldots$ converges to BC. Since B and C are in Σ^∞ and $A = BC$, it follows that $\Sigma^\infty \subseteq \Sigma^\infty\Sigma^\infty$, and so $\Sigma^\infty\Sigma^\infty = \Sigma^\infty$. ■

Actually, multiplying Σ^∞ by Σ doesn't change that set.

Theorem 3.3 *If Σ is product bounded and compact, it follows that $\Sigma\Sigma^\infty = \Sigma^\infty = \Sigma^\infty\Sigma$.*

Proof. We only show that $\Sigma^\infty \subseteq \Sigma\Sigma^\infty$. To do this, let $B \in \Sigma^\infty$. Since $B \in \Sigma^\infty$, there is a matrix subsequence $\pi_{i_1}, \pi_{i_2}, \ldots$ that converges to B.

Factor, for $k > 1$,

$$\pi_{i_k} = A_{i_k} C_{i_k}$$

where $A_{i_2}, A_{i_3}, \ldots$ are in Σ. Now, since Σ is compact, this sequence has a subsequence, say

$$A_{j_2}, A_{j_3}, \ldots$$

which converges to, say, A. And, likewise $C_{j_2}, C_{j_3}, \ldots$ has a subsequence, say

$$C_{k_2}, C_{k_3}, \ldots$$

which converges to, say, C. Thus $A_{k_2}C_{k_2}, A_{k_3}C_{k_3}, \ldots$ converges to AC.

Noting that $A \in \Sigma$, $C \in \Sigma^\infty$ and that

$$AC = B,$$

we have that $\Sigma^\infty \subseteq \Sigma\Sigma^\infty$ and the result follows. ■

When $\Sigma = \{A\}$, multiplying Σ^∞ by any matrix in Σ^∞ doesn't change that set.

Theorem 3.4 *If $\Sigma = \{A\}$ is product bounded, then for any $B \in \Sigma^\infty$, $B\Sigma^\infty = \Sigma^\infty = \Sigma^\infty B$.*

Proof. We prove that $B\Sigma^\infty = \Sigma^\infty$.

Since Σ^∞ is a semigroup, $B\Sigma^\infty \subseteq \Sigma^\infty$. Thus, we need only show that equality holds. For this, let $C \in \Sigma^\infty$. Then we have by Theorem 3.3

$$A^k\Sigma^\infty = \Sigma^\infty$$

for all k. Thus, there is a sequence $C_1, C_2, \ldots$, in Σ^∞ such that

$$A^k C_k = C$$

for all k.

Now suppose the sequence

$$A^{k_1}, A^{k_2}, \ldots$$

converges to B. Since Σ^∞ is bounded, there is a subsequence of $C_{k_1}, C_{k_2}, \ldots$, say

$$C_{j_1}, C_{j_2}, \ldots$$

that converges to, say, $\hat{C}$. Thus

$$B\hat{C} = C.$$

And, as a consequence $\Sigma^\infty \subseteq B\Sigma^\infty$. ■

Using this theorem, we can show that, for $\Sigma = \{A\}$, Σ^∞ is actually a group.

Theorem 3.5 *If $\Sigma = \{A\}$ and Σ is product bounded, then Σ^∞ is a commutative group.*

Proof. We know that Σ^∞ is a semigroup. Thus, we need only prove the additional properties that show Σ^∞ is a commutative group.

To show that Σ^∞ is commutative, let B and C be in Σ^∞. Suppose the sequence

$$A^{i_1}, A^{i_2}, \ldots$$

and

$$A^{j_1}, A^{j_2}, \ldots$$

converge to B and C, respectively. Then

$$\begin{aligned} BC &= \lim_{k\to\infty} A^{i_k} \lim_{k\to\infty} A^{j_k} \\ &= \lim_{k\to\infty} A^{i_k + j_k} \\ &= \lim_{k\to\infty} A^{j_k} \lim_{k\to\infty} A^{i_k} \\ &= CB. \end{aligned}$$

Thus, Σ^∞ is commutative.

To show Σ^∞ has an identity, let $B \in \Sigma^\infty$. Then by using Theorem 3.4, we have that

$$CB = B$$

for some C in Σ^∞. We show that C is the identity in Σ^∞.

For this, let $D \in \Sigma^\infty$. Then by using Theorem 3.4, we can write

$$D = BT$$

for some $T \in \Sigma^\infty$. Now

$$\begin{aligned} CD &= CBT \\ &= BT \\ &= D. \end{aligned}$$

And, by commutivity, $DC = D$. Thus, C is the identity in Σ^∞.

For inverses, let $H \in \Sigma^\infty$. Then, by Theorem 3.4, there is a $E \in \Sigma^\infty$, such that

$$HE = C,$$

so $E = H^{-1}$.

The parts above show that Σ^∞ is a group. ■

3.1.2 Convergence Properties

In this subsection, we look at the convergence properties of

$$\Sigma, \Sigma^2, \ldots$$

where Σ is a product bounded matrix set. Recall that in this case, by Theorem 3.1, Σ^∞ is a compact set.

Concerning the long run behavior of products, we have the following.

Theorem 3.6 *Suppose Σ is product bounded and $\epsilon > 0$. Then, there is a constant N such that if $k > N$ and $\pi_k \in \Sigma^k$, then*

$$d(\pi_k, \Sigma^\infty) < \epsilon.$$

Proof. The proof is by contradiction. Thus, suppose for some $\epsilon > 0$, there is a matrix subsequence from the sequence $\langle \Sigma^k \rangle$, no term of which is in $\Sigma^\infty + \epsilon$. Since Σ is product bounded, these products have a subsequence that converges to, say, A, $A \in \Sigma^\infty$. This implies that $d(A, \Sigma^\infty) \geq \epsilon$, a contradiction. ■

Note that this theorem provides the following corollary.

Corollary 3.1 *Using the hypothesis of the theorem, for some constant N,*

$$\Sigma^k \subseteq \Sigma^\infty + \epsilon$$

for all $k > N$.

In the next result we show that if the sequence $\langle \Sigma^k \rangle$ converges in the Hausdorff sense, then it converges to Σ^∞.

Theorem 3.7 *Let Σ be a product bounded compact set. If $\hat{\Sigma}$ is a compact subset of M_n and $h\left(\Sigma^k, \hat{\Sigma}\right) \to 0$, then $\hat{\Sigma} = \Sigma^\infty$.*

Proof. By the previous corollary, we can see that $\hat{\Sigma} \subseteq \Sigma^\infty$.

Now, suppose $\hat{\Sigma} \neq \Sigma^\infty$; then there is an $A \in \Sigma^\infty$ such that $A \notin \hat{\Sigma}$. Thus,

$$d\left(A, \hat{\Sigma}\right) = \epsilon,$$

where ϵ is a positive constant.

Let $\pi_{k_1}, \pi_{k_2}, \ldots$ be a matrix subsequence of $\langle \Sigma^k \rangle$ that converges to A. Then there is a positive constant N such that for $i > N$,

$$d\left(\pi_i, \hat{\Sigma}\right) > \frac{\epsilon}{2}.$$

Thus,

$$\delta\left(\Sigma^{k_i}, \hat{\Sigma}\right) > \frac{\epsilon}{2}$$

for all $i > N$. This contradicts that $h\left(\Sigma^{k_i}, \hat{\Sigma}\right) \to 0$ as $i \to \infty$. Thus $\hat{\Sigma} = \Sigma^\infty$. ■

In many applications of products of matrices, the matrices are actually multiplied by subsets of F^n. Thus, if

$$W \subseteq F^n,$$

we have the sequence

$$W, \Sigma W, \Sigma^2 W, \ldots.$$

In this sequence, we use the words vector sequence, vector subsequence, and limiting set W_∞ with the obvious meanings.

A way to calculate W_∞, in terms of Σ^∞, follows.

Theorem 3.8 *If Σ is a product bounded set and W a compact set, then $W_\infty = \Sigma^\infty W$.*

Proof. Let $w_0 \in W_\infty$. Then w_0 is the limit of a vector subsequence, say, $\pi_1 w_1, \pi_2 w_2, \ldots$ of $\langle \Sigma^k W \rangle$. Since W is compact and Σ product bounded, we can find a subsequence $\pi_{i_1} w_{i_1}, \pi_{i_2} w_{i_2}, \ldots$ of our sequence such that $w_{i_1}, w_{i_2}, \ldots$ converges to, say, w and $\pi_{i_1}, \pi_{i_2}, \ldots$ converges to, say, $\pi \in \Sigma^\infty$. Thus

$$w_0 = \pi w$$

and we can conclude that $W_\infty \subseteq \Sigma^\infty W$.

Now let $\pi_0 w_0 \in \Sigma^\infty W$, where $w_0 \in W$ and $\pi_0 \in \Sigma^\infty$. Then there is a matrix subsequence $\pi_{i_1}, \pi_{i_2}, \ldots$ that converges to π_0. And we have that $\pi_{i_1} w_0, \pi_{i_2} w_0, \ldots$ is a vector subsequence of $\langle \Sigma^k W \rangle$, which converges to $\pi_0 w_0$. Thus, $\pi_0 w_0 \in W_\infty$, and so $\Sigma^\infty W \subseteq W_\infty$. ■

Theorem 3.9 *If Σ is a product bounded compact set, W a compact set, and $h\left(\Sigma^k, \Sigma^\infty\right) \to 0$ as $k \to \infty$, then $h\left(\Sigma^k W, W_\infty\right) \to 0$ as $k \to \infty$.*

Proof. By Theorem 3.8, we have that $W_\infty = \Sigma^\infty W$, and so we will show that $h\left(\Sigma^k W, \Sigma^\infty W\right) \to 0$ as $k \to \infty$.

Since W is compact, it is bounded by, say, β. We now show that, for all k,

$$h\left(\Sigma^k W, \Sigma^\infty W\right) \leq \beta h\left(\Sigma^k, \Sigma^\infty\right) \tag{3.1}$$

from which the theorem follows.

To do this, let $\pi_k w_0 \in \Sigma^k W$ where $w_0 \in W$ and $\pi_k \in \Sigma^k$. Let $\pi \in \Sigma^\infty$ be such that $d(\pi_k, \pi) \leq h\left(\Sigma^k, \Sigma^\infty\right)$. Then

$$\begin{aligned} d(\pi_k w_0, \pi w_0) &\leq \beta d(\pi_k, \pi) \\ &\leq \beta h\left(\Sigma^k, \Sigma^\infty\right). \end{aligned}$$

And since $\pi_k w_0$ was arbitrary,

$$\delta\left(\Sigma^k W, \Sigma^\infty W\right) \leq \beta h\left(\Sigma^k, \Sigma^\infty\right).$$

Similarly,

$$\delta\left(\Sigma^\infty W, \Sigma^k W\right) \leq \beta h\left(\Sigma^k, \Sigma^\infty\right)$$

from which (3.1) follows. ■

A result, which occurs in applications rather often, follows.

Corollary 3.2 *Suppose Σ is a product bounded compact set and W a compact set. If $\Sigma W \subseteq W$, then $h\left(\Sigma^k W, W_\infty\right) \to 0$ as $k \to 0$.*

Proof. It is clear that $\Sigma^{k+1}W \subseteq \Sigma^k W$, so $W_\infty \subseteq \Sigma^k W$ for all k. Thus, we need only show that if $\epsilon > 0$, there is a constant N such that for all $k > N$

$$\Sigma^k W \subseteq W_\infty + \epsilon.$$

This follows as in the proof of Theorem 3.6. ■

We conclude this subsection by showing a few results for the case when Σ is finite.

If $\Sigma = \{A_1, \ldots, A_m\}$ is product bounded and W compact, then since $\Sigma\Sigma^\infty = \Sigma^\infty$,

$$W_\infty = \Sigma W_\infty = A_1 W_\infty \cup \cdots \cup A_m W_\infty.$$

For $m = 3$, this is somewhat depicted in Figure 3.1. Thus, although each A_i

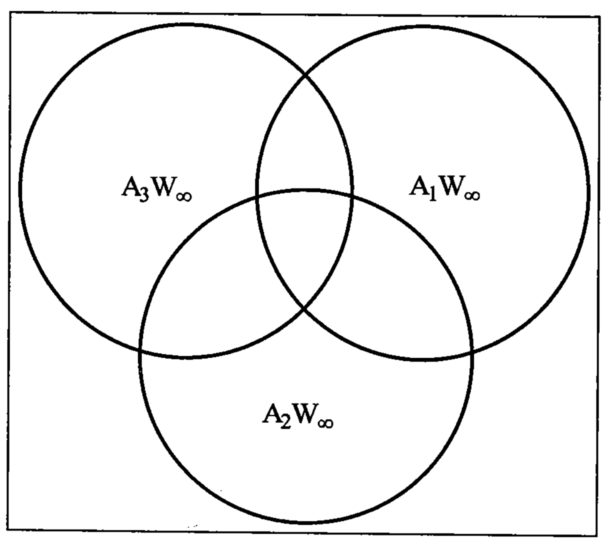

FIGURE 3.1. A view of ΣW_∞.

may contract W_∞ into W_∞, the union of those contractions reconstructs W_∞.

When $\Sigma = \{A\}$, we can give something of an ϵ-view of how $A^k W$ tends to W_∞. We need a lemma.

Lemma 3.1 *Suppose $\{A\}$ is product bounded and W a compact set. Given an $\epsilon > 0$ and any $B \in \Sigma^\infty$, there is a constant N such that*

$$d\left(A^N w, Bw\right) < \epsilon$$

for all $w \in W$.

The theorem follows.

Theorem 3.10 *Using the hypothesis of the lemma, suppose that $L(x) = Ax$ is nonexpansive. Given $\epsilon > 0$ and $B \in \Sigma^\infty$, there is a constant N such that for $k \geq 1$,*

$$h\left(A^{N+k}W, A^k BW\right) < \epsilon.$$

Proof. By the lemma, there is a constant N such that

$$d\left(A^N w, Bw\right) < \epsilon$$

for all $w \in W$. Since L is nonexpansive,

$$d\left(A^{N+k}w, A^k Bw\right) < \epsilon$$

for all $w \in W$ and $k \geq 1$. From this inequality, the theorem follows. ■

Since $\Sigma^\infty W = W_\infty$, $BW \subseteq W_\infty$. Thus this theorem says that $A^{N+k}W$ stays within ϵ of $A^k BW \subseteq W_\infty$ for all k.

We now show that on W_∞, $L(x) = Ax$ is an isometry, so there can be no more collapsing of W_∞.

Theorem 3.11 *Let $\Sigma = \{A\}$ be product bounded and W a compact set. If $L(x) = Ax$ is nonexpansive, then $L(x) = Ax$ is an isometry on W_∞.*

Proof. We first give a preliminary result. For it, recall that $W_\infty = \Sigma^\infty W$ and that Σ^∞ is a group. Let $B \in \Sigma^\infty$ and $A^{i_1}, A^{i_2}, \ldots$ a matrix subsequence of $\langle \Sigma^k \rangle$ that converges to B.

Let $\bar{x}, \bar{y} \in \Sigma^\infty W$. Then, since $L(x) = Ax$ is nonexpansive, we have

$$\begin{aligned} 0 &\leq d\left(A^{i_k}\bar{x}, A^{i_k}\bar{y}\right) - d\left(AA^{i_k}\bar{x}, AA^{i_k}\bar{y}\right) \\ &\leq d\left(A^{i_k}\bar{x}, A^{i_k}\bar{y}\right) - d\left(A^{i_{k+1}}\bar{x}, A^{i_{k+1}}\bar{y}\right). \end{aligned}$$

Taking the limit at $k \to \infty$, we get

$$0 \leq d(B\bar{x}, B\bar{y}) - d(AB\bar{x}, AB\bar{y}) = 0$$

or

$$d(B\bar{x}, B\bar{y}) = d(AB\bar{x}, AB\bar{y}). \tag{3.2}$$

Now let $x, y \in W_\infty$. Since $W_\infty = \Sigma^\infty W = \Sigma^\infty \Sigma^\infty W = \Sigma^\infty W_\infty$, $x = B_1\bar{x}$ and $y = B_2\bar{y}$ where $\bar{x}, \bar{y} \in W_\infty$ and $B_1, B_2 \in \Sigma^\infty$. Using (3.2), and that Σ^∞ is a group,

$$\begin{aligned} d(x, y) &= d(B_1\bar{x}, B_2\bar{y}) \\ &= d\left(B_1\bar{x}, B_1B_1^{-1}B_2\bar{y}\right) \\ &= d\left(AB_1\bar{x}, AB_1\left(B_1^{-1}B_2\bar{y}\right)\right) \\ &= d(Ax, Ay), \end{aligned}$$

the desired result. ■

3.2 Bounded Semigroups

In this section, we give a few results about product bounded matrix sets Σ.

Theorem 3.12 *Let Σ be product bounded matrix set. Then there is a norm, say $\|\cdot\|$, such that $\|A\| \leq 1$ for all $A \in \Sigma$.*

Proof. Let $\Lambda = \Sigma \cup \Sigma^2 \cup \cdots$. Define a vector norm on F^n by

$$\|x\| = \sup\{\|x\|_2, \|\pi x\|_2 : \pi \in \Lambda\}.$$

Then if $A \in \Sigma$,

$$\|Ax\| = \sup\{\|Ax\|_2, \|\pi Ax\|_2 : \pi \in \Lambda\},$$

and since $A, \pi A \in \Lambda$,

$$\begin{aligned} &\leq \sup\{\|x\|_2, \|\pi x\|_2 : \pi \in \Lambda\} \\ &= \|x\|. \end{aligned}$$

Since this inequality holds for all x,

$$\|A\| \leq 1,$$

which is what we need. ■

A special such result follows.

Theorem 3.13 *Suppose that Σ is a bounded matrix set. Suppose further that each matrix $M \in \Sigma$ has partition form*

$$M = \begin{bmatrix} A & B_{12} & B_{13} & \cdots & B_{1k} \\ 0 & C_1 & B_{23} & \cdots & B_{2k} \\ 0 & 0 & C_2 & \cdots & B_{3k} \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & C_k \end{bmatrix}$$

where all matrices on the block main diagonal are square. If there are vector norms $\|\cdot\|_a, \|\cdot\|_{c_1}, \ldots, \|\cdot\|_{c_k}$ such that

$$\|A\|_a \leq 1 \quad \text{and} \quad \|C_i\|_{c_i} \leq \alpha < 1$$

for all $M \in \Sigma$, there is a vector norm $\|\cdot\|$ such that $\|M\| \leq 1$ for all $M \in \Sigma$.

Proof. We prove the result for $k = 1$. For this, we drop subscripts, using $M = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$. The general proof then follows by induction.

For all $x \in F^n$, partition $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ compatible with M. Now, for any constant $K > 0$, we can define a vector norm $\|\cdot\|$ by

$$\|x\| = \|x_1\|_a + K\|x_2\|_c.$$

Then we have, for any $M \in \Sigma$,

$$\begin{aligned} \|Mx\| &= \|Ax_1 + Bx_2\|_a + K\|Cx_2\|_c \\ &\leq \|Ax_1\|_a + \|Bx_2\|_a + K\|Cx_2\|_c \\ &\leq \|x_1\|_a + \|B\|\|x_2\|_c + K\|Cx_2\|_c \end{aligned}$$

where

$$\|B\| = \max_{x_2 \neq 0} \frac{\|Bx_2\|_a}{\|x_2\|_c}.$$

Thus,

$$\begin{aligned} \|Mx\| &\leq \|x_1\|_a + (\|B\| + K\|C\|_c)\|x_2\|_c \\ &= \|x_1\|_a + \left(\frac{\|B\|}{K} + \|C\|_c\right) K\|x_2\|_c. \end{aligned}$$

Since Σ is bounded, $\|B\|$, over all choices of M, is bounded by, say, β. So we can choose K such that

$$\frac{\beta}{K} + \alpha < 1.$$

Then,

$$\begin{aligned} \|Mx\| &\le \|x_1\|_a + K\|x_2\|_c \\ &= \|x\|. \end{aligned}$$

This shows that $\|M\| \le 1$, and since M was arbitrary

$$\|M\| \le 1$$

for all $M \in \Sigma$. ■

In the same way we can prove the following.

Corollary 3.3 *If $\|A\|_a \le 1$ is changed to $\|A\|_a \le \alpha$, then for any δ, $\alpha < \delta < 1$, there is a norm $\|\cdot\|$ such that $\|M\| \le \delta$ for all $M \in \Sigma$.*

Our last result shows that convergence and product bounded are connected. To do this, we need the Uniform Boundedness Lemma.

Lemma 3.2 *Suppose X is a subspace of F^n and*

$$\sup \|\pi x\| < \infty$$

where the sup *is over all π in $\Lambda = \Sigma \cup \Sigma^2 \cup \cdots$ and the inequality holds for any $x \in X$, where $\|x\| = 1$. Then Σ is product bounded on X.*

Proof. We prove the result for the 2-norm. We let $\{x_1, \ldots, x_r\}$ be an orthonormal basis for X. Then, if $x \in X$ and $\|x\|_2 = 1$,

$$x = \alpha_1 x_1 + \cdots + \alpha_r x_r$$

where $|\alpha_1|^2 + \cdots + |\alpha_r|^2 = 1$.

Now let β be such that

$$\sup \|\pi x_i\|_2 \le \beta$$

for $i = 1, \ldots, r$ and all $\pi \in \Lambda$. It follows that if $\pi \in \Lambda$ and $\|x\|_2 = 1$, then

$$\begin{aligned} \|\pi x\|_2 &\le |\alpha_1| \|\pi x_1\|_2 + \cdots + |\alpha_r| \|\pi x_r\|_2 \\ &\le n\beta. \end{aligned}$$

Thus, since x was arbitrary $\|\pi\| \le n\beta$. Since π was arbitrary, Σ is product bounded. ■

Theorem 3.14 *If all infinite products from a matrix set Σ converge, then Σ is product bounded.*

Proof. Suppose all infinite products from Σ converge. Further let $\Lambda = \Sigma \cup \Sigma^2 \cup \cdots$ and define

$$X = \{x \in F^n : \Lambda x \text{ is bounded}\}.$$

Then X is a subspace and $\pi : X \to X$ for all $\pi \in \Lambda$. By the Uniform Boundedness Lemma, there is a constant β such that

$$\sup_{x \in X} \frac{\|\pi x\|}{\|x\|} = \beta < \infty,$$

for all $\pi \in \Lambda$.

If $X = F^n$, then Σ is product bounded. Thus, we suppose $X \ne F^n$. We now show that given an $x \notin X$ and $c > 1$, there are matrices $A_1, \ldots, A_k \in \Sigma$ such that

$$\|A_k \cdots A_1 x\| \ge c \tag{3.3}$$

and

$$A_k \cdots A_1 x \notin X.$$

Since $x \notin X$, there are matrices $A_1, \ldots, A_k \in \Sigma$, such that

$$\|A_k \cdots A_1 x\| > \max(1, \beta \|\Sigma\|)\, c. \tag{3.4}$$

If $A_k \cdots A_1 x \notin X$, we are through. Otherwise, there is a t, $t < k$, such that $A_t \cdots A_1 x \notin X$, while $A_{t+1}(A_t \cdots A_1 x) \in X$. Thus, we have

$$\begin{aligned} \|A_k \cdots A_{t+2}(A_{t+1} \cdots A_1 x)\| &\le \beta \|A_{t+1} \cdots A_1 x\| \\ &\le \beta \|\Sigma\| \|A_t \cdots A_1 x\|. \end{aligned}$$

Thus, by (3.4),

$$c \le \|A_t \cdots A_1 x\|$$

which gives (3.3).

Now, applying result (3.3), suppose $x \notin X$. Then there are $\pi_1, \pi_2, \ldots$ in Λ such that

$$\begin{aligned} &\|\pi_1 x\| \ge 1 \text{ and } \pi_1 x \notin X \\ &\|\pi_2 \pi_1 x\| \ge 2 \text{ and } \pi_2 \pi_1 x \notin X \\ &\cdots \end{aligned}$$

'hus,

$$\pi = \ldots \pi_k \ldots \pi_1$$

; not convergent. This contradicts the hypothesis, and so it follows that $\zeta = F^n$. And, Σ is product bounded. ■

Putting Theorem 3.12 and Theorem 3.14 together, we obtain the following *norm-convergence result.*

Corollary 3.4 *If all infinite products from a matrix set Σ converge, then here is a vector norm $\|\cdot\|$ such that $\|A\| \leq 1$ for all $A \in \Sigma$.*

3.3 Research Notes

Section 1 was developed from Hartfiel (1981, 1991, 2000). Limiting sets, under different names, such as attractors, have been studied elsewhere.

Theorem 3.12 appears in Elsner (1993). Also see Beyn and Elsner (1997). Theorem 3.14 was proved by G. Schechtman and published in Berger and Wang (1992).

4
Patterned Matrices

In this chapter we look at matrix sets Σ of nonnegative matrices in M_n. We find conditions on Σ that assure that contraction coefficients τ_B and τ_W are less than 1 on r-blocks, for some r, of Σ.

4.1 Scrambling Matrices

The contraction coefficient τ_B is less than 1 on any positive matrix in M_n. The first result provides a set Σ in which $(n-1)$-blocks are all positive.

In Corollary 2.5, we saw the following.

Theorem 4.1 *If each matrix in Σ is fully indecomposable, then every $(n-1)$-block from Σ is positive.*

For another such result, we describe a matrix which is somewhat like a fully indecomposable matrix. An $n \times n$ nonnegative matrix A is *primitive* if $A^k > 0$ for some positive integer k.

Instead of computing powers, a matrix can sometimes be checked for primitivity by inspecting its graph. As shown in Varga (1962), if the graph of A is connected (There is a path of positive length from any vertex i to any vertex j.) and

$$k_i = \text{gcd of the lengths of all paths from vertex } i \text{ to itself,}$$

ıen A is primitive if and only if $k_1 = 1$. (Actually, k_1 can be replaced by ıy k_i.)

For example, the Leslie matrix

$$A = \begin{bmatrix} 1 & 2 & 1 \\ .3 & 0 & 0 \\ 0 & .4 & 0 \end{bmatrix}$$

as the graph shown in Figure 4.1 and is thus primitive.

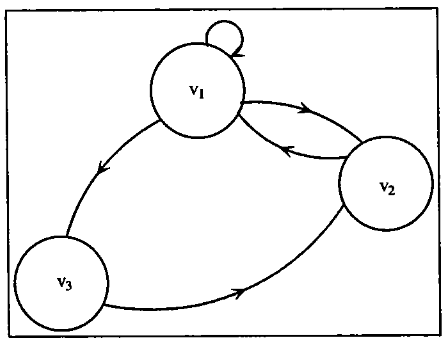

FIGURE 4.1. The graph of A.

A rather well known result on matrix sets and primitive matrices follows. To give it, we need the following notion. For a given matrix A, define $A^* = [a^*_{ij}]$, called the *signum matrix*, by

$$a^*_{ij} = \begin{cases} 1 & \text{if } a_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Theorem 4.2 *Let Σ be a matrix set. Suppose that for all $k = 1, 2, \ldots$, each k-block taken from Σ is primitive. Then there is a positive integer r such that each r-block from Σ is positive.*

Proof. Let

$$p = \text{number of } (0,1)\text{-primitive } n \times n \text{ matrices}$$

and

$$q = \text{the smallest exponent } k \text{ such that } A^k > 0 \text{ for all } (0,1)\text{-primitive matrices } A.$$

Let $r = p + 1$ and $A_{i_1}, \ldots, A_{i_r}$ matrices in Σ. Then, by hypothesis $A_{i_1}, A_{i_2}A_{i_1}, \ldots, A_{i_r} \cdots A_{i_1}$ is a sequence of primitive matrices. Since there are r such matrices, the sequence $A^*_{i_1}, (A_{i_2}A_{i_1})^*, \ldots, (A_{i_r} \cdots A_{i_1})^*$ has a duplication, say

$$(A_{i_s} \cdots A_{i_1})^* = (A_{i_t} \cdots A_{i_1})^*$$

where $s > t$. Thus

$$(A_{i_s} \cdots A_{i_{t+1}})^* (A_{i_t} \cdots A_{i_1})^* = (A_{i_t} \cdots A_{i_1})^*$$

where the matrix arithmetic is Boolean.

Set

$$B = (A_{i_s} \cdots A_{i_{t+1}})^* \text{ and } A = (A_{i_t} \cdots A_{i_1})^*.$$

So we have

$$BA = A.$$

From this it follows that since $B^q > 0$,

$$B^q A = A > 0;$$

thus, $A_{i_t} \cdots A_{i_1} > 0$, and so $A_{i_r} \cdots A_{i_1} > 0$, the result we wanted. ■

A final result of this type uses the following notion. If B is an $n \times n$ $(0,1)$-matrix and $A^* \geq B$, then we say that A has *pattern* B.

Theorem 4.3 *Let B be a primitive $n \times n$ $(0,1)$-matrix. If each matrix in Σ has pattern B, then for some r, every r-block from Σ is positive.*

Proof. Since B is primitive, $B^r > 0$ for some positive integer. Thus, since $(A_{i_r} \cdots A_{i_1})^* \geq (B^r)^*$, the result follows. ■

In the remaining work in this chapter, we will not be interested in r-blocks that are positive but in r-blocks that have at least one positive column. Recall that if A has a positive column, then

$$p(Ax, Ay) < p(x, y)$$

for any positive vectors x and y. Thus, there is some contraction. We can obtain various results of this type by looking at the graph of a matrix.

Let A be an $n \times n$ nonnegative matrix. Then the ij-th entry of A^s is

$$\sum a_{ik_1} a_{k_1 k_2} \cdots a_{k_{s-1} j}$$

where the sum is over all $k_1, \ldots, k_{s-1}$. This entry is positive iff in the graph of A, there is a path, say $v_i, v_{k_1}, v_{k_2}, \ldots, v_{k_{s-1}}, v_j$ from v_i to v_j.

In terms of graphs, we have the following.

Theorem 4.4 *Let A be an $n \times n$ nonnegative matrix in the partitioned form*

$$A = \begin{bmatrix} P & 0 \\ B & C \end{bmatrix} \tag{4.1}$$

where P is an $m \times m$ primitive matrix.

If, in the graph of A, there is a path from each vertex from C to some vertex from P, then there is a positive integer s such that A^s has its first m columns positive.

Proof. Since P is primitive, there is a positive integer k such that $P^{k+t} > 0$ for all $t \geq 0$. Thus, there is a path from any vertex of P to any vertex of P having length $k + t$.

Let t_i denote the length of a path from v_i, a vertex from C, to a vertex of P. Let $t = \max t_i$. Then, using the remarks in the previous paragraph, if v_i is a vertex of C, then there is a path of length $k + t$ to any vertex in P. Thus, A^s, where $s = k + t$, has its first m columns positive. ■

An immediate consequence follows.

Corollary 4.1 *Let A be an $n \times n$ nonnegative matrix as given in (4.1). If each matrix in Σ has pattern A, then for some r, every r-block from Σ has a positive column.*

We extend the theorem, a bit, as follows. Let A be an $n \times n$ nonnegative matrix. As shown in Gantmacher (1964), there is an $n \times n$ permutation matrix P such that

$$PAP^t = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ A_{21} & A_2 & \cdots & 0 \\ & & \cdots & \\ A_{s1} & A_{s2} & \cdots & A_s \end{bmatrix} \tag{4.2}$$

where each A_k is either an $n_k \times n_k$ irreducible matrix or it is a 1×1 0-matrix. Here, the partitioned form in (4.2) is called the *canonical form* of A. If for some k, $A_{k1}, \ldots, A_{kk-1}$ are all 0-matrices, then A_k is called an *isolated block* in the canonical form.

If A^k has a positive column for some k, then A_1 must be primitive since if A_1 is not primitive, as shown in Gantmacher (1964), its index of imprimitivity is at least 2. This assures that A_1^k, and hence A^k, never has a positive column for any k.

Corollary 4.2 *Let A be an $n \times n$ nonnegative matrix. Suppose the canonical form (4.2) of A satisfies the following:*

1. *A_1 is a primitive $m \times m$ matrix.*
2. *The canonical form for A has no isolated blocks.*

Then there is a positive integer s such that A^s has its first m columns positive.

Proof. Observe in the canonical form that since there are no isolated blocks, each vertex of a block has a path to a vertex of a block having a lower subscript. This implies that each vertex has a path to any vertex in A_1. The result now follows from the theorem. ■

A different kind of condition that can be placed on the matrices in Σ to assure r-blocks have a positive column, is that of scrambling. An $n \times n$ nonnegative matrix A is *scrambling* if $AA^t > 0$. This means, of course, that for any row indices i and j, there is a column index k such that $a_{ik} > 0$ and $a_{jk} > 0$.

A consequence of the previous corollary follows.

Corollary 4.3 *If an $n \times n$ nonnegative matrix is scrambling, then A^s has a positive column for some positive integer s.*

Proof. Suppose the canonical form for A is as in (4.2). Since A is scrambling, so is its canonical form, so this form can have no isolated blocks. And, A_1 must be primitive since, if this were not true, A_1 would have index of imprimitivity at least 2. And this would imply that A_1, and thus A, is not scrambling. ■

It is easily shown that the product of two $n \times n$ scrambling matrices is itself a scrambling matrix. Thus, we have the following.

Theorem 4.5 *The set of $n \times n$ scrambling matrices is a semigroup.*

If Σ contains only scrambling matrices, $\Sigma \cup \Sigma^2 \cup \cdots$ contains only scrambling matrices. We use this to show that for some r, every r-block from such a Σ has a positive column.

Theorem 4.6 *Suppose every matrix in Σ is scrambling. Then there is an r such that every r-block from Σ has a positive column.*

Proof. Consider any product of $r = 2^{n^2} + 1$ matrices from Σ, say $A_r \cdots A_1$. Let $\sigma(A) = A^*$, the signum matrix of the matrix A. Note that there are at most 2^{n^2} distinct $n \times n$ signum matrices. Thus,

$$\sigma(A_s \cdots A_1) = \sigma(A_t \cdots A_1)$$

for some s and t with, say, $r \geq s > t$. It follows that

$$\sigma(A_s \cdots A_{t+1}) \, \sigma(A_t \cdots A_1) = \sigma(A_t \cdots A_1)$$

when Boolean arithmetic is applied. Thus, using Boolean arithmetic, for any $k > 0$,

$$\sigma(A_s \cdots A_{t+1})^k \, \sigma(A_t \cdots A_1) = \sigma(A_t \cdots A_1).$$

We know by Corollary 4.3 that for some k, $\sigma(A_s \cdots A_{t+1})^k$ has a column of 1's. And, by the definition of Σ, $\sigma(A_t \cdots A_1)$ has no row of 0's. Thus, $\sigma(A_t \cdots A_1)$ has a positive column, and consequently so does $A_t \cdots A_1$.

From this it follows that since $r > t$, any r-block from Σ has a positive column. ■

4.2 Sarymsakov Matrices

To describe a Sarymsakov matrix, we need a few preliminary remarks.

Let A be an $n \times n$ nonnegative matrix. For all $S \subseteq \{1, \ldots, n\}$, define the consequent function F, belonging to A, as

$$F(S) = \{j : a_{ij} > 0 \text{ for some } i \in S\}.$$

Thus, $F(S)$ gives the set of all consequent indices of the indices in S. For example, if

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

then $F(\{2\}) = \{1, 2\}$ and $F(\{1, 2\}) = \{1, 2, 3\}$.

Let B be an $n \times n$ nonnegative matrix and F_1, F_2 the consequent functions belonging to A, B, respectively. Let F_{12} be consequent function belonging to AB.

Lemma 4.1 $F_2(F_1(S)) = F_{12}(S)$ *for all subsets* S.

Proof. Let $j \in F_2(F_1(S))$. Then there is a $k \in F_1(S)$ such that $b_{kj} > 0$ and an $i \in S$ such that $a_{ik} > 0$. Since the ij-th entry of AB is

$$\sum_{r=1}^{n} a_{ir} b_{rj} \tag{4.3}$$

that entry is positive. Thus, $j \in F_{12}(S)$. Since j was arbitrary, it follows that $F_2(F_1(S)) \subseteq F_{12}(S)$.

Now, let $j \in F_{12}(S)$. Then by (4.3), there is an $i \in S$ and a k such that $a_{ik} > 0$ and $b_{kj} > 0$. Thus, $k \in F_1(S)$ and $j \in F_2(\{k\}) \subseteq F_2(F_1(S))$. And, as j was arbitrary, we have that $F_{12}(S) \subseteq F_2(F_1(S))$.

Put together, this yields the result. ■

The corollary can be extended to the following.

Theorem 4.7 *Let $A_1, \ldots, A_k$ be $n \times n$ nonnegative matrices and $F_1, \ldots, F_k$ consequent functions belonging to them, respectively. Let $F_{1 \cdots k}$ be the consequent function belonging to $A_1 \cdots A_k$. Then*

$$F_k(\cdots(F_1(S))) = F_{1 \cdots k}(S)$$

for all subsets $S \subseteq \{1, \ldots, n\}$.

We now define the Sarymsakov matrix. Let A by an $n \times n$ nonnegative matrix and F its consequent function. Suppose that for any two disjoint nonempty subsets S, S' either

1. $F(S) \cap F(S') \neq \emptyset$ or
2. $F(S) \cap F(S') = \emptyset$ and $|F(S) \cup F(S')| > |S \cup S'|$.

Then A is a *Sarymsakov matrix*.

A diagram depicting a choice for S and S' for both (1) and (2) is given in Figure 4.2.

Note that if A is a Sarymsakov matrix, then A can have no row of 0's since, if A had a row of 0's, say the i-th row, then $S = \{i\}$ and $S' = \{1, \ldots, n\} - S$ would deny (2).

The set K of all $n \times n$ Sarymsakov matrices is called the Sarymsakov class of $n \times n$ matrices. A major property of Sarymsakov matrices follows.

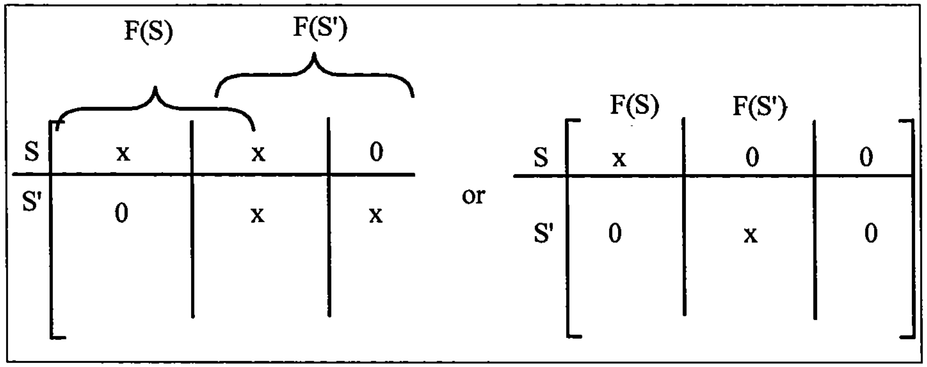

FIGURE 4.2. A diagram for Sarymsakov matrices.

Theorem 4.8 *Let $A_1, \ldots, A_{n-1}$ be $n \times n$ Sarymsakov matrices. Then $A_1 \cdots A_{n-1}$ is scrambling.*

Proof. Let $F_1, \ldots, F_{n-1}$ be the consequent functions for the matrices $A_1, \ldots, A_{n-1}$, respectively. Let $F_{1 \cdots k}$ be the consequent functions for the products $A_1 \cdots A_k$, respectively, for all k.

Now let i and j be distinct row indices. In the following, we use that if

$$F_{1\cdots k}(\{i\}) \cap F_{1\cdots k}(\{j\}) \neq \emptyset$$

for some $k < n$, then

$$F_{1\cdots n-1}(\{i\}) \cap F_{1\cdots n-1}(\{j\}) \neq \emptyset. \tag{4.4}$$

Using the definition, either $F_1(\{i\}) \cap F_1(\{j\}) \neq \emptyset$, in which case (4.4) holds or

$$|F_1(\{i\}) \cup F_1(\{j\})| > 2.$$

In the latter case, either $F_{12}(\{i\}) \cap F_{12}(\{j\}) \neq \emptyset$, so (4.4) holds or

$$|F_{12}(\{i\}) \cup F_{12}(\{j\})| > 3.$$

And continuing, we see that either (4.4) holds or

$$|F_{1\cdots n-1}(\{i\}) \cup F_{1\cdots n-1}(\{j\})| > n.$$

The latter condition cannot hold, so (4.4) holds. And since this is true for all i and j, $A_1 \cdots A_{n-1}$ is scrambling. ∎

A different description of Sarymsakov matrices follows.

Lemma 4.2 *Let A be an $n \times n$ nonnegative matrix and F the consequent function belonging to A. The two statements, which are given below, are equivalent.*

1. *A is a Sarymsakov matrix.*
2. *If C is a nonempty subset of row indices of A satisfying $|F(C)| \leq |C|$, then*

$$F(B) \cap F(C - B) \neq \emptyset$$

for any proper nonempty subset B of C.

Proof. Assuming (1), let B and C be as described in (2). Set $S = B$ and $S' = C - B$. Since S and S' are disjoint nonempty subsets, by definition, $F(S) \cap F(S') \neq \emptyset$ or $F(S) \cap F(S') = \emptyset$ and $|F(S) \cup F(S')| > |S \cup S'|$. In the latter case, we would have $|F(C)| > |C|$, which contradicts the hypothesis. Thus the first condition holds and, using that $B = S$, $C - B = S'$, we have $F(B) \cap F(C - B) \neq \emptyset$. This yields (2).

Now assume (2) and let S and S' be nonempty disjoint subsets of indices. Set $C = S \cup S'$. We need to consider two cases.

Case 1. Suppose $|F(C)| \leq |C|$. Then, setting $S = B$, we have $F(S) \cap F(S') \neq \emptyset$, thus satisfying the first part of the definition of a Sarymsakov matrix.

Case 2. Suppose $|F(C)| > |C|$. Then we have $|F(S) \cup F(S')| > |S \cup S'|$, so the second part of the definition of a Sarymsokov matrix is satisfied.

Thus, A is a Sarymsakov matrix, and so (2) implies (1), and the lemma is proved. ∎

We conclude by showing that the set of all $n \times n$ Sarymsakov matrices is a semigroup.

Theorem 4.9 *Let A_1 and A_2 be in K. Then A_1A_2 is in K.*

Proof. We show that A_1A_2 satisfies (2) of the previous lemma. We use that F_1, F_2, and F_{12} are consequent functions for A_1, A_2, and A_1A_2, respectively.

Let C be a nonempty subset of row indices, satisfying this inequality $|F_{12}(C)| \leq |C|$, and B a proper nonempty subset of C. We now argue two cases.

Case 1. Suppose $|F_1(C)| \leq |C|$. Then since $A_1 \in K$, it follows that $F_1(B) \cap F_1(C - B) \neq \emptyset$. Thus,

$$\begin{aligned} \emptyset &\neq F_2(F_1(B)) \cap F_2(F_1(C - B)) \\ &= F_{12}(B) \cap F_{12}(C - B). \end{aligned}$$

Case 2. Suppose $|F_1(C)| > |C|$. Then, using our assumption on C, $|F_{12}(C)| \leq |C| < |F_1(C)|$. Thus,

$$|F_2(F_1(C))| < |F_1(C)|.$$

Now, using that $A_2 \in K$ and the previous lemma, if D is any proper nonempty subset of $F_1(C)$, then

$$F_2(D) \cap F_2(F_1(C) - D) \neq \emptyset. \tag{4.5}$$

Now, we look at two subcases.

Subcase a: Suppose $F_1(B) \cap F_1(C-B) = \emptyset$. Then $F_1(B)$ is a proper subset of $F_1(C)$ and $F_1(C-B) = F_1(C) - F_1(B)$. Thus, applying (4.5), with $D = F_1(B)$, we have

$$\begin{aligned}\emptyset &\neq F_2(F_1(B)) \cap F_2(F_1(C) - F_1(B)) \\ &= F_2(F_1(B)) \cap F_2(F_1(C-B)) \\ &= F_{12}(B) \cap F_{12}(C-B),\end{aligned}$$

satisfying the conclusion of (2) in the lemma.

Subcase b: Suppose $F_1(B) \cap F_1(C-B) \neq \emptyset$. Then we have $F_2(F_1(B)) \cap F_2(F_1(C-B)) \neq \emptyset$ or $F_{12}(B) \cap F_{12}(C-B) \neq \emptyset$, again the conclusion of (2) in the lemma.

Thus, $A_1A_2 \in K$. ■

The obvious corollary follows.

Corollary 4.4 *The set K is a semigroup.*

To conclude this section, we show that every scrambling matrix is a Sarymsakov matrix.

Theorem 4.10 *Every scrambling matrix is a Sarymsakov matrix.*

Proof. Let A be an $n \times n$ scrambling matrix. Using (2) of Lemma 4.2, let C and B be as the sets described there. If i and j are row indices in B and $C - B$, respectively, then since A is scrambling $F(\{i\}) \cap F(\{j\}) \neq \emptyset$. Thus $F(B) \cap F(C-B) \neq \emptyset$ and the result follows. ■

4.3 Research Notes

As shown in Brualdi and Ryser (1991), if A is primitive, then $A^{(n-1)^2-1} > 0$. This, of course, provides a test for primitivity. Other such results are also given there.

The proof of Theorem 4.2 is due to Wolfowitz (1963). The bulk of Section 2 was formed from Sarymsakov (1961) and Hartfiel and Seneta (1990). Rhodius (1989) described a class of 'almost scrambling' matrices and showed this class to be a subset of K. More work in this area can be found in Seneta (1981).

Pullman (1967) described the columns that can occur in infinite products of Boolean matrices. Also see the references there.

5
Ergodicity

This chapter begins a sequence of chapters concerned with various types of convergence of infinite products of matrices. In this chapter we consider row allowable matrices. If $A_1, A_2, \ldots$ is a sequence of $n \times n$ row allowable matrices, we let

$$P_k = A_k \ldots A_1$$

for all k. We look for conditions that assure the columns of P_k approach being proportional. In general 'ergodic' refers to this kind of behavior.

5.1 Birkhoff Coefficient Results

In this section, we use the Birkhoff contraction coefficient to obtain several results assuring ergodicity in an infinite product of row allowable matrices.

The preliminary results show what $\tau_B(P_k) \to 0$ as $k \to \infty$ means about the entries of P_k as $k \to \infty$. The first of these results uses the notion that the sequence $\langle P_k \rangle$ tends to *column proportionality* if for all r, s, $\frac{p_{ir}^{(1)}}{p_{is}^{(1)}}, \frac{p_{ir}^{(2)}}{p_{is}^{(2)}}, \ldots$ converges to some constant α_{rs}, regardless of i. (So, the r-th and the s-th columns are nearly proportional, and become even more so as $k \to \infty$.)

An example follows.

Example 5.1 *Let*

$$A_k = \begin{cases} k\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} & \text{if } k \text{ is odd,} \\ k\begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix} & \text{if } k \text{ is even.} \end{cases}$$

Then

$$P_k = \begin{cases} k!\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} & \text{if } k \text{ is odd,} \\ k!\begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix} & \text{if } k \text{ is even.} \end{cases}$$

Note that P_k tends to column proportionality, but P_k doesn't converge. Here $\alpha_{12} = 1$.

If we let $P_k = \left[p_1^{(k)} p_2^{[k]}\right]$, where $p_1^{(k)}$ and $p_2^{[k]}$ are column vectors, a picture of how column proportional might appear is given in Figure 5.1.

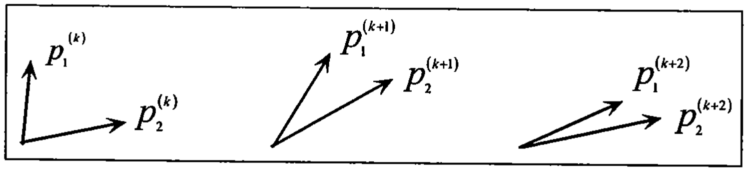

FIGURE 5.1. A view of column proportionality.

Lemma 5.1 *If $A_1, A_2, \ldots$ is a sequence of $n \times n$ positive matrices and $P_k = A_k \cdots A_1$ for all k, then $\lim_{k\to\infty} \tau_B(P_k) \to 0$ as $k \to \infty$ iff P_k tends to column proportionality.*

Proof. Suppose that P_k tends to column proportionality. Then

$$\begin{aligned} \lim_{k\to\infty} \phi(P_k) &= \lim_{k\to\infty} \min_{i,j,r,s} \frac{p_{ir}^{(k)}}{p_{jr}^{(k)}} \frac{p_{js}^{(k)}}{p_{is}^{(k)}} \\ &= 1. \end{aligned}$$

Thus,

$$\lim_{k\to\infty} \tau_B(P_k) = \lim_{k\to\infty} \frac{1-\sqrt{\phi(P_k)}}{1+\sqrt{\phi(P_k)}} = 0.$$

Conversely, suppose that $\lim_{k\to\infty} \tau_B(P_k) = 0$. Define

$$m_{rs}^{(k)} = \min_i \frac{p_{ir}^{(k)}}{p_{is}^{(k)}}, \quad M_{rs}^{(k)} = \max_i \frac{p_{ir}^{(k)}}{p_{is}^{(k)}}.$$

The idea of the proof is made clearer by letting x and y denote the r-th and s-th columns of P_{k-1}, respectively. Then

$$\begin{aligned} m_{rs}^{(k)} &= \min_i \frac{p_{ir}^{(k)}}{p_{is}^{(k)}} \\ &= \min_i \frac{\sum_{j=1}^n a_{ij}^{(k)} x_j}{\sum_{j=1}^n a_{ij}^{(k)} y_j} \\ &= \min_i \frac{\sum_{j=1}^n a_{ij}^{(k)} y_j \frac{x_j}{y_j}}{\sum_{j=1}^n a_{ij}^{(k)} y_j} \\ &= \min_i \sum_{j=1}^n \left(\frac{a_{ij}^{(k)} y_j}{\sum_{j=1}^n a_{ij}^{(k)} y_j} \right) \frac{x_j}{y_j}. \end{aligned}$$

Since $\sum_{j=1}^n \left(\frac{a_{ij}^{(k)} y_j}{\sum_{j=1}^n a_{ij}^{(k)} y_j} \right) \frac{x_j}{y_j}$ is a convex sum of $\frac{x_1}{y_1}, \ldots, \frac{x_n}{y_n}$, we have

$$m_{rs}^{(k)} \geq \min_j \frac{x_j}{y_j} = m_{rs}^{(k-1)}.$$

Similarly,

$$M_{rs}^{(k)} \leq M_{rs}^{(k-1)},$$

and it follows that $\lim_{k\to\infty} m_{rs}^{(k)} = m_{rs}$, as well as $\lim_{k\to\infty} M_{rs}^{(k)} = M_{rs}$, exist.

Finally, since $\lim_{k\to\infty} \tau_B(P_k) = 0$, $\lim_{k\to\infty} \phi(P_k) = 1$, so

$$\begin{aligned} 1 &= \lim_{k\to\infty} \min_{i,,j,r,s} \frac{p_{ir}^{(k)} p_{js}^{(k)}}{p_{is}^{(k)} p_{jr}^{(k)}} \\ &= \lim_{k\to\infty} \min_{r,s} \frac{m_{rs}^{(k)}}{M_{rs}^{(k)}} \\ &= \frac{m_{pq}}{M_{pq}} \end{aligned}$$

for some p and q. And thus $m_{pq} = M_{pq}$. Since $1 \geq \frac{m_{rs}}{M_{rs}} \geq \frac{m_{pq}}{M_{pq}}$ for all r and s, it follows that $m_{rs} = M_{rs}$. Thus,

$$\lim_{k\to\infty} \frac{p_{ir}^{(k)}}{p_{is}^{(k)}} = m_{rs}$$

for all i and $P_1, P_2, \ldots$ tend to column proportionality. ■

Formally the sequence $P_1, P_2, \ldots$ is *ergodic* if there exists a sequence of positive rank one matrices $S_1, S_2, \ldots$ such that

$$\lim_{k\to\infty} \frac{p_{ij}^{(k)}}{s_{ij}^{(k)}} = 1 \tag{5.1}$$

for all i and j. To give some meaning to this, we can think of the matrices P_k and S_k as $n^2 \times 1$ vectors. Then

$$p(P_k, S_k) = \ln \max_{i,j,r,t} \frac{p_{ij}^{(k)}}{s_{ij}^{(k)}} \frac{s_{rt}^{(k)}}{p_{rt}^{(k)}}$$

where p is the projective metric.

Now by (5.1),

$$\lim_{k\to\infty} p(P_k, S_k) = 0$$

and so

$$\lim_{k\to\infty} p\left(\frac{1}{\|P_k\|_F} P_k, \frac{1}{\|S_k\|_F} S_k\right) = 0$$

where $\|\cdot\|_F$ is the Frobenius norm (the 2-norm on the $n^2 \times 1$ vectors).

Let

$$S = \{R : R \text{ is an } n \times n \text{ rank 1 nonnegative matrix where } \|R\|_F = 1\}.$$

Recalling that

$$d\left(\frac{1}{\|P_k\|_F} P_k, S\right) = \min_{R \in S} p(P_k, R),$$

we see that

$$d\left(\frac{1}{\|P_k\|_F} P_k, S\right) \to 0 \text{ as } k \to \infty.$$

Thus, $\frac{1}{\|P_k\|_F} P_k$ tends to S. So, in a projective sense, the P_k's tend to the rank one matrices.

The result linking ergodic and τ_B follows.

Theorem 5.1 *The sequence $P_1, P_2, \ldots$ is ergodic iff $\lim_{k\to\infty} \tau_B(P_k) = 0$.*

Proof. If $P_1, P_2, \ldots$ is ergodic, there are rank 1 positive $n \times n$ matrices $S_1, S_2, \ldots$ satisfying (5.1). Thus, using that S_k is rank one,

$$\begin{aligned} \phi(P_k) &= \min_{i,j,r,s} \frac{p_{ir}^{(k)}}{p_{is}^{(k)}} \frac{p_{js}^{(k)}}{p_{jr}^{(k)}} \\ &= \min_{i,j,r,s} \frac{p_{ir}^{(k)}}{p_{is}^{(k)}} \frac{p_{js}^{(k)}}{p_{jr}^{(k)}} \frac{s_{is}^{(k)}}{s_{ir}^{(k)}} \frac{s_{jr}^{(k)}}{s_{js}^{(k)}}, \end{aligned}$$

so

$$\lim_{k\to\infty} \phi(P_k) = 1.$$

Hence,

$$\lim_{k\to\infty} \tau_B(P_k) = \lim_{k\to\infty} \frac{1 - \sqrt{\phi(P_k)}}{1 + \sqrt{\phi(P_k)}} = 0.$$

Conversely, suppose $\lim_{k\to\infty} \tau_B(P_k) = 0$. For $e = (1, 1, \ldots, 1)^t$, define

$$S_k = \frac{P_k e e^t P_k}{e^t P_k e} = \left[\frac{\sum_{r=1}^{n} p_{ir}^{(k)} \cdot \sum_{s=1}^{n} p_{sj}^{(k)}}{\sum_{r=1}^{n}\sum_{s=1}^{n} p_{rs}^{(k)}} \right]$$

an $n \times n$ rank one positive matrix. Then

$$\frac{p_{ij}^{(k)}}{s_{ij}^{(k)}} = \frac{p_{ij}^{(k)} \sum_{r=1}^{n}\sum_{s=1}^{n} p_{rs}^{(k)}}{\sum_{r=1}^{n} p_{ir}^{(k)} \sum_{s=1}^{n} p_{sj}^{(k)}} = \frac{\sum_{r=1}^{n}\sum_{s=1}^{n} p_{ij}^{(k)} p_{rs}^{(k)}}{\sum_{r=1}^{n}\sum_{s=1}^{n} p_{ir}^{(k)} p_{sj}^{(k)}}.$$

Using the quotient bound result (2.3), we have that

$$\phi(P_k) \le \frac{p_{ij}^{(k)}}{s_{ij}^{(k)}} \le \frac{1}{\phi(P_k)}.$$

And since $\lim_{k\to\infty} \phi(P_k) = 1$, we have

$$\lim_{k\to\infty} \frac{p_{ij}^{(k)}}{s_{ij}^{(k)}} = 1.$$

Thus, the sequence $P_1, P_2, \ldots$ is ergodic. ■

We now give some conditions on matrices $A_1, A_2, \ldots$ that assure the sequence $P_1, P_2, \ldots$ is ergodic. Basically, these conditions assure that $\langle \phi(A_k) \rangle$ doesn't converge to 0 too fast.

Theorem 5.2 *Let $A_1, A_2, \ldots$ be a sequence of $n \times n$ row allowable matrices. If $\sum_{k=1}^{\infty} \sqrt{\varphi(A_k)} = \infty$, then $P_1, P_2, \ldots$ is ergodic.*

Proof. Since $\sum_{k=1}^{\infty} \sqrt{\varphi(A_k)} = \infty$, it follows by Theorem 51 (See Hyslop (1959) or the Appendix.) that $\prod_{k=1}^{\infty} \left(1 + \sqrt{\varphi(A_k)}\right) = \infty$. Thus, since

$$\tau_B(P_k) \le \tau_B(A_k) \cdots \tau_B(A_1) = \frac{\left(1 - \sqrt{\varphi(A_k)}\right)}{\left(1 + \sqrt{\varphi(A_k)}\right)} \cdots \frac{\left(1 - \sqrt{\varphi(A_1)}\right)}{\left(1 + \sqrt{\varphi(A_1)}\right)} \le \frac{1}{\left(1 + \sqrt{\varphi(A_k)}\right) \cdots \left(1 + \sqrt{\varphi(A_1)}\right)},$$

$\lim_{k\to\infty} \tau_B(P_k) = 0$. ■

A corollary, more easily applied than the theorem, follows.

Corollary 5.1 *Let m_k and M_k be the smallest and largest entries in A_k, respectively. If $\sum_{k=1}^{\infty} \left(\frac{m_k}{M_k}\right) = \infty$, then $P_1, P_2, \ldots$ is ergodic.*

Proof. Since

$$\frac{m_k}{M_k} \le \sqrt{\varphi(A_k)},$$

the corollary follows. ■

A final such result follows.

Theorem 5.3 *Let $A_1, A_2, \ldots$ be a sequence of row allowable matrices. Suppose that for some positive integer r and for some $\gamma > 0$, we have that $\phi\left(A_{(k+1)r} \cdots A_{kr+1}\right) \ge \gamma^2$ for all k. Then*

$$\tau_B(A_k \cdots A_1) \le \left(\frac{1-\gamma}{1+\gamma}\right)^{\left[\frac{k}{r}\right]}.$$

Proof. Write

$$k = rq + s \text{ where } 0 \le s < r.$$

Block $A_k \cdots A_1$ into lengths r forming

$$A_k \cdots A_{rq+1} B_q B_{q-1} \cdots B_1.$$

Then

$$\begin{aligned}\tau_B(A_k\cdots A_1) &\le \tau_B(A_k\cdots A_{rq+1})\,\tau_B(B_q)\cdots\tau_B(B_1)\\ &\le \left(\frac{1-\gamma}{1+\gamma}\right)^q\\ &= \left(\frac{1-\gamma}{1+\gamma}\right)^{\left[\frac{k}{r}\right]},\end{aligned}$$

which yields the result. ■

A lower bound for γ, using $m = \inf_{a_{ij}>0} a_{ij}$ and $M = \sup_{i,j} a_{ij}$, where the inf and sup are over all matrices $A_1, A_2, \ldots$, can be found by noting that $\inf(B)_{ij} \ge m^r$, $\sup(B)_{ij} \le n^{r-1}M^r$ for all r-blocks $B = A_{rk}\cdots A_{r(k-1)+1}$. Thus

$$\phi(B) \ge \left(\frac{m^r}{n^{r-1}M^r}\right)^2$$

and so γ can be taken as

$$\gamma = \frac{m^r}{n^{r-1}M^r}.$$

Furthermore, types of matrices which produce positive r-blocks, for some r, were given in Chapter 4.

5.2 Direct Results

In this section, we look at matrix sets Σ such that if $A_1, A_2, \ldots$ is a sequence taken from Σ and x, y positive vectors, then

$$p(P_k x, P_k y) \to 0 \text{ as } k \to \infty.$$

Note that this implies that $\frac{P_k x}{\|P_k x\|}$ and $\frac{P_k y}{\|P_k y\|}$, as vectors, get closer. So we will call such sets *ergodic sets*. As we will see, the results in this section apply to very special matrices. However, as we will point out later, these matrices arise in applications.

A preliminary such result follows.

Theorem 5.4 *Let Σ be a set of $n \times n$ row-allowable matrices M of the form*

$$M = \begin{bmatrix} A & 0 \\ B & 0 \end{bmatrix}$$

where A is $n_1 \times n_1$. If τ is a positive constant, $\tau < 1$, and $\tau_B(A) < \tau$ for all matrices M in Σ, then Σ is an ergodic set. And the rate of convergence is geometric.

Proof. Let $M_1, M_2, \ldots$ be matrices in Σ where

$$M_k = \begin{bmatrix} A_k & 0 \\ B_k & 0 \end{bmatrix},$$

and A_k is $n_1 \times n_1$ for all k. Then

$$M_k\cdots M_1 = \begin{bmatrix} A_k\cdots A_1 & 0 \\ B_kA_{k-1}\cdots A_1 & 0 \end{bmatrix}.$$

Now, let x and y be positive vectors where

$$x = \begin{bmatrix} x_A \\ x_C \end{bmatrix}, y = \begin{bmatrix} y_A \\ y_C \end{bmatrix}$$

partitioned compatibly to M. Then

$$\begin{aligned}M_k\cdots M_1 x &= \begin{bmatrix} A_k & 0 \\ B_k & 0 \end{bmatrix}\begin{bmatrix} A_{k-1}\cdots A_1 & 0 \\ B_{k-1}A_{k-2}\cdots A_1 & 0 \end{bmatrix} x\\ &= \begin{bmatrix} A_k \\ B_k \end{bmatrix} A_{k-1}\cdots A_1 x_A\end{aligned}$$

and

$$M_k\cdots M_1 y = \begin{bmatrix} A_k \\ B_k \end{bmatrix} A_{k-1}\cdots A_1 y_A.$$

Thus,

$$\begin{aligned}&p(M_k\cdots M_1 x, M_k\cdots M_1 y)\\ &= p\left(\begin{bmatrix} A_k \\ B_k \end{bmatrix} A_{k-1}\cdots A_1 x_A, \begin{bmatrix} A_k \\ B_k \end{bmatrix} A_{k-1}\cdots A_1 y_A\right)\end{aligned}$$

and by Lemma 2.1, we continue to get

$$\begin{aligned}&\le p(A_{k-1}\cdots A_1 x_A, A_{k-1}\cdots A_1 y_A)\\ &\le \tau_B(A_{k-1})\cdots\tau_B(A_1)\,p(x_A, y_A)\\ &\le \tau^{k-1} p(x_A, y_A).\end{aligned}$$

Thus, as $k \to \infty$, $p(M_k\cdots M_1 x, M_k\cdots M_1 y) \to 0$. ■

A special set $\hat{\Sigma}$ of row allowable matrices $M = \begin{bmatrix} A & 0 \\ B & C \end{bmatrix}$, where A is $n_1 \times n_1$, occurring in practice has $m_{n_1+1,n_1} > 0$ and C lower triangular with 0 main diagonal. For example, Leslie matrices have this property. A corollary now follows.

Corollary 5.2 *Suppose for some integer r, $\hat{\Sigma}^r$ satisfies the hypothesis of the theorem. Then $\hat{\Sigma}$ is an ergodic set and the convergence rate of products is geometric.*

The next result uses that Σ is a set of $n \times n$ row allowable matrices with form

$$M = \begin{bmatrix} A & 0 \\ B & C \end{bmatrix} \tag{5.2}$$

where A is $n_1 \times n_1$, B is row allowable, and C is $n_2 \times n_2$. Let Σ_A be the set of all matrices A which occur as an upper left $n_1 \times n_1$ submatrix of some $M \in \Sigma$. Concerning the submatrices A, B, C of any $M \in \Sigma$, we assume $a_h, b_h, c_h, a_l, b_l, c_l$, are positive constants such that

$$\max a_{ij} \le a_h, \ \max b_{ij} \le b_h, \ \max c_{ij} \le c_h,$$
$$\min_{a_{ij}>0} a_{ij} \ge a_\ell, \ \min_{b_{ij}>0} b_{ij} \ge b_\ell, \ \min_{c_{ij}>0} c_{ij} \ge c_\ell.$$

The major theorem of this section follows.

Theorem 5.5 *Let Σ be described as above and suppose that $\tau_B(\Sigma_A^r) \le \tau < 1$ for some positive integer r. Further, suppose that there exists a constant K_1 where*

$$n_2 K_1 < 1 \ \text{and} \ \frac{c_h}{a_\ell} \le K_1,$$

and that a constant K_2 satisfies

$$\frac{c_h}{b_l} \le K_2, \frac{b_h}{a_l} \le K_2, \frac{a_h}{b_\ell} \le K_2, \frac{a_h}{a_l} \le K_2.$$

Let x and y be positive vectors. Suppose

$$\max_{i,j} \frac{x_i}{y_j} \le K_2 \ \text{and} \ \max_{i,j} \frac{y_i}{x_j} \le K_2.$$

Then there is a positive constant $T, T < 1$, such that

$$p(P_k x, P_k y) \le K T^k$$

where K is a constant. Thus Σ is an ergodic set. (This K depends on x and y.)

Proof. Let

$$P(1,k) = M_k \cdots M_1, \ P(t,k) = M_k \cdots M_t$$

where each $M_i \in \Sigma$ and $1 \le t < k$. Partition these matrices as in (5.2),

$$P(1,k) = \begin{bmatrix} P_{11}(k) & 0 \\ P_{21}(k) & P_{22}(k) \end{bmatrix},$$
$$P(t,k) = \begin{bmatrix} P_{11}(t,k) & 0 \\ P_{21}(t,k) & P_{22}(t,k) \end{bmatrix}.$$

Note that

$$P_{11}(k) = A_k \cdots A_1$$
$$P_{21}(k) = \sum_{j=1}^{k} C_k \cdots C_{j+1} B_j A_{j-1} \cdots A_1$$

where $C_k \cdots C_{j+1} = I$ if $j = k$ and $A_{j-1} \cdots A_1 = I$ if $j = 1$,

$$P_{22}(k) = C_k \cdots C_1.$$

Thus, for $k \ge r$, $P_{11}(k) > 0$ and since B_k is row allowable, for $k > r$, $P_{21}(k) > 0$. Further, by rearrangement, for any t, $1 \le t < k$,

$$P_{21}(1,k) = P_{22}(t+1,k) P_{21}(1,t) + P_{21}(t+1,k) P_{11}(1,t).$$

Now, using that $P_{*1}(t+1,k) = \begin{bmatrix} P_{11}(t+1,k) \\ P_{21}(t+1,k) \end{bmatrix}$, partitioning $x = \begin{bmatrix} x_A \\ x_C \end{bmatrix}$, $y = \begin{bmatrix} y_A \\ y_C \end{bmatrix}$ as is M, and applying the triangle inequality, we get

$$\begin{aligned} & p(P(1,k)x, P(1,k)y) \le \\ & p(P(1,k)x, P_{*1}(t+1,k) P_{11}(1,t) x_A) \\ & + p(P_{*1}(t+1,k) P_{11}(1,t) x_A, P_{*1}(t+1,k) P_{11}(1,t) y_A) \\ & + p(P_{*1}(t+1,k) P_{11}(1,t) y_A, P(1,k) y). \end{aligned}$$

We now find bounds on each of these terms. To keep our notation compact when necessary, we use $\hat{P} = P(t+1,k)$ and $\bar{P} = P(1,t)$.

1. $p(P(1,k)x, P_{*1}(t+1,k) P_{11}(1,t) x_A)$. To bound this term, we need to consider the expressions

$$r_i = \frac{[P(1,k)x]_i}{[P_{*1}(t+1,k) P_{11}(1,t) x_A]_i}.$$

By definition, it is clear that $r_i \geq 1$ for all i with equality assured when $i \leq n_1$. So,

$$\frac{\max r_i}{\min r_i} = \max r_i.$$

Since

$$P(1,k) = \begin{bmatrix} \hat{P}_{11} & 0 \\ \hat{P}_{21} & \hat{P}_{22} \end{bmatrix} \begin{bmatrix} \bar{P}_{11} & 0 \\ \bar{P}_{21} & \bar{P}_{22} \end{bmatrix} = \begin{bmatrix} \hat{P}_{11}\bar{P}_{11} & 0 \\ \hat{P}_{21}\bar{P}_{11} + \hat{P}_{22}\bar{P}_{21} & \hat{P}_{22}\bar{P}_{22} \end{bmatrix},$$

$$r_i = \frac{\left[\hat{P}_{21}\bar{P}_{11}x_A + \hat{P}_{22}\bar{P}_{21}x_A + \hat{P}_{22}\bar{P}_{22}x_C\right]_i}{\left[\hat{P}_{21}\bar{P}_{11}x_A\right]_i} = 1 + \frac{\left[\hat{P}_{22}\bar{P}_{21}x_A + \hat{P}_{22}\bar{P}_{22}x_C\right]_i}{\left[\hat{P}_{21}\bar{P}_{11}x_A\right]_i}.$$

We will define several numbers, the importance of which will be seen later. Set $Q = n_1K_2\left(n_2K_1\right)^f$ and let f be sufficiently large that $Q < 1$ and $t = \left[\frac{k}{f+1}\right]$. (We assume $k > f+1$.) Then using that $P_{21}(t+1,k)\,P_{11}(1,t)\,x_A \geq B_kA_{k-1}\cdots A_1x_A$, we get

$$r_i \leq 1 + \frac{\left[C_k\cdots C_{t+1}\left(\sum_{j=1}^{t} C_t\cdots C_{j+1}B_jA_{j-1}\cdots A_1\right)x_A\right]_i}{\left[B_kA_{k-1}\cdots A_1x_A\right]_i} + \frac{\left[C_k\cdots C_1x_C\right]_i}{\left[B_kA_{k-1}\cdots A_1x_A\right]_i}.$$

We now bound this expression by one involving K_1 and K_2. Let $x_l = \min x_i$ and $x_h = \max x_i$. Then

$$r_i \leq 1 + \frac{c_h^{k-t}\sum_{j=1}^{t} n_2^{t-j-1}n_1^j\,c_h^{t-j}b_ha_h^{j-1}x_h}{b_la_l^{k-1}x_l} + \frac{n_2^kc_h^kx_h}{b_la_l^{k-1}x_l}.$$

Thus,

$$\begin{aligned} r_i - 1 &\leq K_1^{k-t}n_2^{k-t}\sum_{j=1}^{t} n_2^{t-j-1}n_1^j\,K_1^{t-j}K_2^{j+1} + n_2^kK_1^{k-1}K_2^2 \\ &\leq \left(n_2K_1\right)^{k-t}K_2\sum_{j=1}^{t}\left(n_2K_1\right)^{t-j}n_1^jK_2^j + n_2K_2^2\left(n_2K_1\right)^{k-1} \\ &= \left(n_2K_1\right)^{k-t}K_2\left(n_2K_1\right)^t\sum_{j=1}^{t}\left(\frac{n_1K_2}{n_2K_1}\right)^j + n_2K_2^2\left(n_2K_1\right)^{k-1}. \end{aligned}$$

For simplicity, set $\beta = \frac{n_1K_2}{n_2K_1}$. Since $n_1K_2 \geq 1$, $n_2K_1 < 1$,

$$r_i - 1 \leq \left(n_2K_1\right)^kK_2\beta\left(\frac{\beta^t-1}{\beta-1}\right) + n_2K_2^2\left(n_2K_1\right)^{k-1},$$

$$r_i - 1 \leq \left(n_2K_1\right)^kK_2\beta\frac{\beta^t}{\beta-1} + n_2K_2^2\left(n_2K_1\right)^{k-1}.$$

For the first term of this expression,

$$\begin{aligned} \left(n_2K_1\right)^k\beta^t &= \left(n_2K_1\right)^k\left(\frac{n_1K_2}{n_2K_1}\right)^{\left[\frac{k}{f+1}\right]} \\ &\leq \left(n_2K_1\right)^k\left(\frac{n_1K_2}{n_2K_1}\right)^{\frac{k}{f+1}} \\ &= \left[\left(n_2K_1\right)^{f+1}\left(\frac{n_1K_2}{n_2K_1}\right)\right]^{\frac{k}{f+1}} \\ &= \left[\left(n_2K_1\right)^f\left(n_1K_2\right)\right]^{\frac{k}{f+1}} \\ &\leq Q^{\frac{k}{f+1}} \\ &= \left(Q^{\frac{1}{f+1}}\right)^k. \end{aligned}$$

Continuing,

$$r_i - 1 \leq K_2\frac{\beta}{\beta-1}\left(Q^{\frac{1}{f+1}}\right)^k + n_2K_2^2\left(n_2K_1\right)^{k-1}$$

and thus,

$$\begin{aligned} &p\left(P\left(1,k\right)x, P_{*1}\left(t+1,k\right)P_{11}\left(1,t\right)x_A\right) \\ &\le \max \ln r_i \\ &\le K_2 \frac{\beta}{\beta-1}\left(Q^{\frac{1}{f+1}}\right)^k + n_2 K_2^2 \left(n_2 K_1\right)^{k-1}. \end{aligned}$$

Let $T_1 = \max\left\{Q^{\frac{1}{f+1}}, n_2 K_1\right\}$, so $T < 1$. Then by setting $K_3 = \frac{K_2\beta}{\beta-1} + \frac{K_2^2}{K_1}$ and continuing

$$p\left(P\left(1,k\right)x, P_{*1}\left(t+1,k\right)P_{11}\left(1,t\right)x_A\right) \le K_3 T_1^k.$$

Similarly we can show

$$p\left(P_{*1}\left(t+1,k\right)P_{11}\left(1,t\right)y_A, P\left(1,k\right)y\right) \le K_4 T_1^k$$

for some constant K_4.

2. $p\left(P_{*1}\left(t+1,k\right)P_{11}\left(1,t\right)x_A, P_{*1}\left(t+1,k\right)P_{11}\left(1,t\right)y_A\right)$

$$\begin{aligned} &\le p\left(P_{11}\left(1,t\right)x_A, P_{11}\left(1,t\right)y_A\right) \\ &\le \tau_B^{\left[\frac{t}{r}\right]} p\left(x_A, y_A\right) \\ &\le \tau^{\frac{t}{r}} p\left(x_A, y_A\right) \\ &\le \left(\tau^{\frac{1}{r}}\right)^{\frac{k}{f+1}} p\left(x_A, y_A\right) \\ &\le \left(\tau^{\frac{1}{r(f+1)}}\right)^k p\left(x_A, y_A\right) \\ &= K_5 T_2^k \end{aligned}$$

where $T_2 = T^{\frac{1}{r(f+1)}}$ and $K_5 = p\left(x_A, y_A\right)$.

Putting (1) and (2) together,

$$\begin{aligned} p\left(P\left(1,k\right)x, P\left(1,k\right)y\right) &\le K_3 T_1^k + K_4 T_1^k + K_5 T_2^k \\ &\le KT^k \end{aligned}$$

where $T = \max\{T_1, T_2\}$ and $K = K_3 + K_4 + K_5$, the desired result. ■

The condition $\frac{c_h}{a_h} \le K_1$, $K_1 < \frac{1}{n_2}$, which we need to assure $T < 1$, may seem a bit restrictive; however, in applications we would expect that for large k and $P_k = \begin{bmatrix} A_k \cdots A_1 & 0 \\ P_{21}\left(k\right) & C_k \cdots C_1 \end{bmatrix}$,

$$A_k \cdots A_1 > C_k \cdots C_1$$

and so the theorem can be applied to blocks from Σ. For blocks we would use the previous theorem together with the following one.

Theorem 5.6 *Let Σ be a row proper matrix set and x, y positive vectors. Suppose K, T, with $T < 1$, are constants such that*

$$p\left(B_q \cdots B_1 x,\ B_q \cdots B_1 y\right) \le K T^q$$

for all r-blocks $B_1, \ldots, B_q$ from Σ. Then

$$p\left(M_k \cdots M_1 x,\ M_k \cdots M_1 y\right) \le K T^{\left[\frac{k}{r}\right]}$$

for any matrices $M_1, \ldots, M_k$ in Σ.

Proof. Write $M_k \cdots M_1 = M_k \cdots M_{rq+1} B_q \cdots B_1$ where the subscripts satisfy $k = rq + t, 0 \le t < r$. Then

$$\begin{aligned} p\left(M_k \cdots M_1 x,\ M_k \cdots M_1 y\right) &\le p\left(B_q \cdots B_1 x,\ B_q \cdots B_1 y\right) \\ &\le KT^q \\ &= KT^{\left[\frac{k}{r}\right]}, \end{aligned}$$

the desired result. ■

5.3 Research Notes

The results in Section 1 were based on Hajnal (1976), while those in Section 2 were formed from Cohen (1979).

In a related paper, Cohn and Nerman (1990) showed results, such as those in this chapter, by linking nonnegative matrix products and nonhomogeneous Markov chains. Cohen (1979) discussed how ergodic theorems apply to demographics. And Geramita and Pullman (1984) provided numerous examples of demographic problems in the study of biology.