E. Seneta

Non-negative Matrices and Markov Chains

Second Edition

Springer-Verlag
New York Heidelberg Berlin

E. Seneta
The University of Sydney
Department of Mathematical Statistics
Sydney, N.S.W. 2006
Australia

AMS Classification (1980): 15A48, 15A51, 15A52, 60G10, 62E99

Library of Congress Cataloging in Publication Data
Seneta, E. (Eugene), 1941–
Non-negative matrices and Markov chains.

(Springer series in statistics)
Revision of: Non-negative matrices. 1973.
Bibliography: p.
Includes indexes.
1. Non-negative matrices. 2. Markov processes. I. Title. II. Series.
QA188.S46 1981 512.9'434 81-5768
AACR2

The first edition of this book was published by George Allen & Unwin Ltd. (London, 1973)

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-90598-7 Springer-Verlag New York Heidelberg Berlin
ISBN 3-540-90598-7 Springer-Verlag Berlin Heidelberg New York

To my parents

Things are always at their best in the beginning.
Blaise Pascal, *Lettres Provinciales* [1656–1657]

Preface

Since its inception by Perron and Frobenius, the theory of non-negative matrices has developed enormously and is now being used and extended in applied fields of study as diverse as probability theory, numerical analysis, demography, mathematical economics, and dynamic programming, while its development is still proceeding rapidly as a branch of pure mathematics in its own right. While there are books which cover this or that aspect of the theory, it is nevertheless not uncommon for workers in one or another branch of its development to be unaware of what is known in other branches, even though there is often formal overlap. One of the purposes of this book is to relate several aspects of the theory, insofar as this is possible.

The author hopes that the book will be useful to mathematicians; but in particular to the workers in applied fields, so the mathematics has been kept as simple as could be managed. The mathematical requisites for reading it are: some knowledge of real-variable theory, and matrix theory; and a little knowledge of complex-variable; the emphasis is on real-variable methods. (There is only one part of the book, the second part of §5.5, which is of rather specialist interest, and requires deeper knowledge.) Appendices provide brief expositions of those areas of mathematics needed which may be less generally known to the average reader.

The first four chapters are concerned with finite non-negative matrices, while the following three develop, to a small extent, an analogous theory for infinite matrices. It has been recognized that, generally, a research worker will be interested in one particular chapter more deeply than in others; consequently there is a substantial amount of independence between them. Chapter 1 should be read by every reader, since it provides the foundation for the whole book; thereafter Chapters 2–4 have some interdependence as do Chapters 5–7. For the reader interested in the infinite matrix case, Chapter 5 should be read before Chapters 6 and 7. The exercises are intimately connected with the text, and often provide further development of the theory or deeper insight into it, so that the reader is strongly advised to (at least) look over the exercises relevant to his interests, even if not actually wishing to do them. Roughly speaking, apart from Chapter 1, Chapter 2 should be of interest to students of mathematical economics, numerical analysis, combinatorics, spectral theory of matrices, probabilists and statisticians; Chapter 3 to mathematical economists and demographers; and Chapter 4 to probabilists. Chapter 4 is believed to contain one of the first expositions in text-book form of the theory of finite inhomogeneous Markov chains, and contains due regard for Russian-language literature. Chapters 5–7 would at present appear to be of interest primarily to probabilists, although the probability emphasis in them is not great.

This book is a considerably modified version of the author's earlier book *Non-Negative Matrices* (Allen and Unwin, London/Wiley, New York, 1973, hereafter referred to as *NNM*). Since *NNM* used probabilistic techniques throughout, even though only a small part of it explicitly dealt with probabilistic topics, much of its interest appears to have been for people acquainted with the general area of probability and statistics. The title has, accordingly, been changed to reflect more accurately its emphasis and to account for the expansion of its Markov chain content. This has gone hand-in-hand with a modification in approach to this content, and to the treatment of the more general area of inhomogeneous products of non-negative matrices, via "coefficients of ergodicity," a concept not developed in *NNM*.

Specifically in regard to modification, §§2.5–§2.6 are completely new, and §2.1 has been considerably expanded, in Chapter 2. Chapter 3 is completely new, as is much of Chapter 4. Chapter 6 has been modified and expanded and there is an additional chapter (Chapter 7) dealing in the main with the problem of practical computation of stationary distributions of infinite Markov chains from finite truncations (of their transition matrix), an idea also used elsewhere in the book.

It will be seen, consequently, that apart from certain sections of Chapters 2 and 3, the present book as a whole may be regarded as one approaching the theory of Markov chains from a non-negative matrix standpoint.

Since the publication of *NNM*, another English-language book dealing exclusively with non-negative matrices has appeared (A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, New York, 1979). The points of contact with either *NNM* or its present modification (both of which it complements in that its level, approach, and subject matter are distinct) are few. The interested reader may consult the author's review in *Linear and Multilinear Algebra*, 1980, **9**; and may wish to note the extensive bibliography given by Berman and Plemmons. In the present book we have, accordingly, only added references to those of *NNM* which are cited in new sections of our text.

In addition to the acknowledgements made in the Preface to *NNM*, the

author wishes to thank the following: S. E. Fienberg for encouraging him to write §2.6 and Mr. G. T. J. Visick for acquainting him with the non-statistical evolutionary line of this work; N. Pullman, M. Rothschild and R. L. Tweedie for materials supplied on request and used in the book; and Mrs. Elsie Adler for typing the new sections.

Sydney, 1980

E. Seneta

Contents

Glossary of Notation and Symbols

T	usual notation for a non-negative matrix.
A	typical notation for a matrix.
A'	the transpose of the matrix A.
a_{ij}	the (i, j) entry of the matrix A.
$\tilde{T}$	the incidence matrix of the non-negative matrix T.
P	usual notation for a stochastic matrix.
0	zero; the zero matrix.
$\mathbf{x}$	typical notation for a column vector.
$\mathbf{0}$	the column vector with all entries 0.
$\mathbf{1}$	the column vector with all entries 1.
$P_1 \sim P_2$	the matrix P_1 has the same incidence matrix as the matrix P_2.
$\min^+$	the minimum among all strictly positive elements.
R^k, R_k	k-dimensional Euclidean space.
R	set of strictly positive integers; convergence parameter of an irreducible matrix T; a certain submatrix.
I	the identity (unit) matrix; the set of inessential indices.
$i \in R$	i is an element of the set R.
$\mathscr{C} \subset \mathscr{S}$ or $\mathscr{C} \subseteq \mathscr{S}$	$\mathscr{C}$ is a subset of the set $\mathscr{S}$.
$_{(n)}T$	$(n \times n)$ northwest corner truncation of T.
$\Delta_i, \Delta_i(s)$	the principal minor of $(sI - T)$.
$_{(n)}\Delta(\beta)$	$\det\,[_{(n)}I - \beta,\ _{(n)}T]$.
$_{(n)}\Delta$	$_{(n)}\Delta\,(R)$
M.C.	Markov chain.
$\mathscr{E}$	mathematical expectation operator.
G_1	class of $(n \times n)$ regular matrices.
M	class of $(n \times n)$ Markov matrices.
G_2	class of stochastic matrices defined on p. 143.
G_3	class of $(n \times n)$ scrambling matrices.

PART I

FINITE NON-NEGATIVE MATRICES

CHAPTER 1

Fundamental Concepts and Results in the Theory of Non-negative Matrices

We shall deal in this chapter with square non-negative matrices $T = \{t_{ij}\}$, $i, j = 1, \ldots, n$; i.e. $t_{ij} \geq 0$ for all i, j, in which case we write $T \geq 0$. If, in fact, $t_{ij} > 0$ for all i, j we shall put $T > 0$.

This definition and notation extends in an obvious way to row vectors (x') and column vectors (y), and also to expressions such as, e.g.

$$T \geq B \Leftrightarrow T - B \geq 0$$

where T, B and 0 are square non-negative matrices of compatible dimensions.

Finally, we shall use the notation $x' = \{x_i\}$, $y = \{y_i\}$ for both row vectors x' or column vectors y; and $T^k = \{t_{ij}^{(k)}\}$ for kth powers.

Definition 1.1. A square non-negative matrix T is said to be *primitive* if there exists a positive integer k such that $T^k > 0$.

It is clear that if any other matrix $\tilde{T}$ has the same dimensions as T, and has positive entries and zero entries in the same positions as T, then this will also be true of all powers T^k, $\tilde{T}^k$ of the two matrices.

As *incidence matrix* $\tilde{T}$ corresponding to a given T replaces all the positive entries of T by ones. Clearly $\tilde{T}$ is primitive if and only if T is.

1.1 The Perron—Frobenius Theorem for Primitive Matrices[1]

Theorem 1.1. *Suppose T is an $n \times n$ non-negative primitive matrix. Then there exists an eigenvalue r such that:*

(*a*) *r real,* >0;

[1] This theorem is fundamental to the entire book. The proof is necessarily long; the reader may wish to defer detailed consideration of it.

(b) *with r can be associated strictly positive left and right eigenvectors;*
(c) $r > |\lambda|$ *for any eigenvalue* $\lambda \neq r$;
(d) *the eigenvectors associated with r are unique to constant multiples.*
(e) *If* $0 \leq B \leq T$ *and* β *is an eigenvalue of* B, *then* $|\beta| \leq r$. *Moreover,* $|\beta| = r$ *implies* $B = T$.
(f) *r is a simple root of the characteristic equation of T.*

PROOF. (*a*) Consider initially a row vector $x' \geq 0'$, $\neq 0'$; and the product $x'T$. Let

$$r(x) = \min_j \frac{\sum_i x_i t_{ij}}{x_j}$$

where the ratio is to be interpreted as '∞' if $x_j = 0$. Clearly, $0 \leq r(x) < \infty$. Now since

$$x_j r(x) \leq \sum_i x_i t_{ij} \text{ for each } j,$$

$$x' r(x) \leq x'T,$$

and so

$$x'\mathbf{1} r(x) \leq x'T\mathbf{1}.$$

Since $T\mathbf{1} \leq K\mathbf{1}$ for $K = \max_i \sum_j t_{ij}$, it follows that

$$r(x) \leq x'\mathbf{1}K/x'\mathbf{1} = K = \max_i \sum_j t_{ij}$$

so $r(x)$ is uniformly bounded above for all such x. We note also that since T, being primitive, can have no column consisting entirely of zeroes, $r(\mathbf{1}) > 0$, whence it follows that

$$r = \sup_{\substack{x \geq 0 \\ x \neq 0}} \min_j \frac{\sum_i x_i t_{ij}}{x_j} \tag{1.1}$$

satisfies

$$0 < r(\mathbf{1}) \leq r \leq K < \infty.$$

Moreover, since neither numerator or denominator is altered by the norming of x,

$$r = \sup_{\substack{x \geq 0 \\ x'x = 1}} \min_j \frac{\sum_i x_i t_{ij}}{x_j}.$$

Now the region $\{x;\, x \geq 0,\, x'x = 1\}$ is compact in the Euclidean n-space R_n, and the function $r(x)$ is an upper-semicontinuous mapping of this region into R_1; hence[1] the supremum, r is actually *attained* for some x, say $\hat{x}$. Thus there exists $\hat{x} \geq 0$, $\neq 0$ such that

$$\min_j \frac{\sum_i \hat{x}_i t_{ij}}{\hat{x}_j} = r,$$

i.e.

$$\sum_i \hat{x}_i t_{ij} \geq r\hat{x}_j; \quad \text{or} \quad \hat{x}'T \geq r\hat{x}' \tag{1.2}$$

[1] see Appendix C.

for each $j = 1, \ldots, n$; with equality for some element of $\hat{x}$.

Now consider

$$z' = \hat{x}'T - r'\hat{x}' \geq 0'.$$

Either $z = 0$, or not; if not, we know that for $k \geq k_0$, $T^k > 0$ as a consequence of the primitivity of T, and so

$$z'T^k = (\hat{x}'T^k)T - r(\hat{x}'T^k) > 0',$$

i.e.

$$\frac{\{(\hat{x}'T^k)T\}_j}{\{\hat{x}'T^k\}_j} > r, \text{ each } j,$$

where the subscript j refers to the jth element. This is a contradiction to the definition of r. Hence always

$$z = 0,$$

whence

$$\hat{x}'T = r\hat{x}' \tag{1.3}$$

which proves (*a*).

(*b*) By iterating (1.3)

$$\hat{x}'T^k = r^k\hat{x}'$$

and taking k sufficiently large $T^k > 0$, and since $\hat{x} \geq 0$, $\neq 0$, in fact $\hat{x}' > 0'$.

(*c*) Let λ be any eigenvalue of T. Then for some $x \neq 0$ and possibly complex valued

$$\sum_i x_i t_{ij} = \lambda x_j \qquad \left(\text{so that } \sum_i x_i t_{ij}^{(k)} = \lambda^k x_j\right) \tag{1.4}$$

whence

$$|\lambda x_j| = \left|\sum_i x_i t_{ij}\right| \leq \sum_i |x_i| t_{ij},$$

so that

$$|\lambda| \leq \frac{\sum_i |x_i| t_{ij}}{|x_j|}$$

where the right side is to be interpreted as '∞' for any $x_j = 0$. Thus

$$|\lambda| \leq \min_j \frac{\sum_i |x_i| t_{ij}}{|x_j|},$$

and by the definition (1.1) of r

$$|\lambda| \leq r.$$

Now suppose $|\lambda| = r$; then

$$\sum_i |x_i| t_{ij} \geq |\lambda|\, |x_j| = r|x_j|$$

which is a situation identical to that in the proof of part (*a*), (1.2); so that eventually in the same way

$$\sum_i |x_i| t_{ij} = r|x_j|, >0; \qquad j = 1, 2, \ldots, n \tag{1.5}$$

and so
$$\sum_i |x_i| t_{ij}^{(k)} = r^k |x_j|, >0; \qquad j = 1, 2, \ldots, n,$$

i.e.
$$\left|\sum_i x_i t_{ij}^{(k)}\right| = |\lambda^k x_j| = \sum_i |x_i t_{ij}^{(k)}| \tag{1.6}$$

where k can be chosen so large that $T^k > 0$, by the *primitivity* assumption on T; but for two numbers $\gamma, \delta \neq 0$, $|\gamma + \delta| = |\gamma| + |\delta|$ if and only if γ, δ have the same direction in the complex plane. Thus writing $x_j = |x_j| \exp i\theta_j$, (1.6) implies $\theta_j = \theta$ is independent of j, and hence cancelling the exponential throughout (1.4) we get

$$\sum_i |x_i| t_{ij} = \lambda |x_j|$$

where, since $|x_i| > 0$ all i, λ is real and positive, and since we are assuming $|\lambda| = r$, $\lambda = r$ (or the fact follows equivalently from (1.5)).

(*d*) Suppose $\boldsymbol{x}' \neq \boldsymbol{0}'$ is a left eigenvector (possibly with complex elements) corresponding to r.

Then, by the argument in (*c*), so is $\boldsymbol{x}'_+ = \{|x_i|\} \neq \boldsymbol{0}'$, which in fact satisfies $\boldsymbol{x}_+ > \boldsymbol{0}$. Clearly

$$\eta' = \hat{x}' - cx'$$

is then also a left eigenvector corresponding to r, for any c such that $\eta \neq \boldsymbol{0}$; and hence the same things can be said about η as about $\boldsymbol{x}$; in particular $\eta_+ > \boldsymbol{0}$.

Now either $\boldsymbol{x}$ is a multiple of $\hat{x}$ or not; if not c can be chosen so that $\eta \neq \boldsymbol{0}$, but some element of η is; this is impossible as $\eta_+ > \boldsymbol{0}$.

Hence $\boldsymbol{x}'$ is a multiple of $\hat{x}'$.

Right eigenvectors. The arguments (*a*)–(*d*) can be repeated separately for right eigenvectors; (*c*) guarantees that the r produced is the same, since it is purely a statement about eigenvalues.

(*e*) Let $\boldsymbol{y} \neq \boldsymbol{0}$ be a right eigenvector of B corresponding to β. Then taking moduli as before

$$|\beta| \boldsymbol{y}_+ \leq B\boldsymbol{y}_+, \ \leq T\boldsymbol{y}_+, \tag{1.7}$$

so that using the same $\hat{x}$ as before

$$|\beta| \hat{x}'\boldsymbol{y}_+ \leq \hat{x}' T \boldsymbol{y}_+ = r\hat{x}'\boldsymbol{y}_+$$

and since $\hat{x}'\boldsymbol{y}_+ > 0$,

$$|\beta| \leq r.$$

Suppose now $|\beta| = r$; then from (1.7)

$$r\boldsymbol{y}_+ \leq T\boldsymbol{y}_+$$

whence, as in the proof of (*b*), it follows $T\boldsymbol{y}_+ = r\boldsymbol{y}_+ > \boldsymbol{0}$; whence from (1.7)

$$r\boldsymbol{y}_+ = B\boldsymbol{y}_+ = T\boldsymbol{y}_+$$

so it must follow, from $B \leq T$, that $B = T$.

(*f*) The following identities are true for all numbers, real and complex, including eigenvalues of T:

$$\left.\begin{aligned}(xI - T)\,\mathrm{Adj}\,(xI - T) &= \det\,(xI - T)I \\ \mathrm{Adj}\,(xI - T)(xI - T) &= \det\,(xI - T)I\end{aligned}\right\} \tag{1.8}$$

where I is the unit matrix and 'det' refers to the determinant. (The relation is clear for x not an eigenvalue, since then $\det\,(xI - T) \neq 0$; when x is an eigenvalue it follows by continuity.)

Put $x = r$: then any one row of Adj $(rI - T)$ is either (i) a left eigenvector corresponding to r; or (ii) a row of zeroes; and similarly for columns. By assertions (*b*) and (*d*) (already proved) of the theorem, Adj $(rI - T)$ is either (i) a matrix with no elements zero; or (ii) a matrix with all elements zero. We shall prove that one element of Adj $(rI - T)$ is positive, which establishes that case (i) holds. The (n, n) element is

$$\det\,(r_{(n-1)}I - {}_{(n-1)}T)$$

where ${}_{(n-1)}T$ is T with last row and column deleted; and ${}_{(n-1)}I$ is the corresponding unit matrix. Since

$$0 \leq \begin{bmatrix} {}_{(n-1)}T & \boldsymbol{0} \\ \boldsymbol{0}' & 0 \end{bmatrix} \leq T, \text{ and } \neq T,$$

the last since T is primitive (and so can have no zero column), it follows from (*e*) of the theorem that no eigenvalue of ${}_{(n-1)}T$ can be as great in modulus as r. Hence

$$\det\,(r_{(n-1)}I - {}_{(n-1)}T) > 0,$$

as required; *and moreover we deduce that* Adj $(rI - T)$ *has all its elements positive.*

Write $\phi(x) = \det\,(xI - T)$; then differentiating (1.8) elementwise

$$\mathrm{Adj}\,(xI - T) + (xI - T)\frac{d}{dx}\{\mathrm{Adj}\,(xI - T)\} = \phi'(x)I.$$

Substitute $x = r$, and premultiply by $\hat{x}'$;

$$(\boldsymbol{0}' <)\hat{x}'\,\mathrm{Adj}\,(rI - T) = \phi'(r)\hat{x}'$$

since the other term vanishes. Hence $\phi'(r) > 0$ and so r is simple. □

Corollary 1.

$$\min_i \sum_{j=1}^n t_{ij} \le r \le \max_i \sum_{j=1}^n t_{ij} \tag{1.9}$$

with equality on either side implying equality throughout (i.e. r can only be equal to the maximal or minimal row sum if all row sums are equal).

A similar proposition holds for column sums.

PROOF. Recall from the proof of part (*a*) of the theorem, that

$$0 < r(\mathbf{1}) = \min_j \sum_i t_{ij} \le r \le K = \max_i \sum_j t_{ij} < \infty. \tag{1.10}$$

Since T' is also primitive and has the same r, we have also

$$\min_j \sum_i t_{ji} \le r \le \max_i \sum_j t_{ji} \tag{1.11}$$

and a combination of (1.10) and (1.11) gives (1.9).

Now assume that one of the equalities in (1.9) holds, but not all row sums are equal. Then by increasing (or, if appropriate, decreasing) the positive elements of T (but keeping them positive), produce a new primitive matrix, with all row sums equal and the same r, in view of (1.9); which is impossible by assertion (*e*) of the theorem. □

Corollary 2. *Let $\boldsymbol{v}'$ and $\boldsymbol{w}$ be positive left and right eigenvectors corresponding to r, normed so that $\boldsymbol{v}'\boldsymbol{w} = 1$. Then*

$$\text{Adj}\ (rI - T)/\phi'(r) = \boldsymbol{w}\boldsymbol{v}'.$$

To see this, first note that since the columns of Adj $(rI - T)$ are multiples of the same positive right eigenvector corresponding to r (and its rows of the same positive left eigenvector) it follows that we can write it in the form $\boldsymbol{y}\boldsymbol{x}'$ where $\boldsymbol{y}$ is a right and $\boldsymbol{x}'$ a left positive eigenvector. Moreover, again by uniqueness, there exist positive constants c_1, c_2 such that $\boldsymbol{y} = c_1 \boldsymbol{w}$, $\boldsymbol{x}' = c_2 \boldsymbol{v}'$, whence

$$\text{Adj}\ (rI - T) = c_1 c_2 \boldsymbol{w}\boldsymbol{v}'.$$

Now, as in the proof of the simplicity of r,

$$\boldsymbol{v}'\phi'(r) = \boldsymbol{v}'\ \text{Adj}\ (rI - T) = c_1 c_2 \boldsymbol{v}'\boldsymbol{w}\boldsymbol{v}' = c_1 c_2 \boldsymbol{v}'$$

so that $\boldsymbol{v}'\boldsymbol{w}\phi'(r) = c_1 c_2 \boldsymbol{v}'\boldsymbol{w}$

i.e. $\qquad c_1 c_2 = \phi'(r)$ as required. □

(Note that $c_1 c_2$ = sum of the diagonal elements of the adjoint = sum of the principal $(n-1) \times (n-1)$ minors of $(rI - T)$.)

Theorem 1.1 is the strong version of the Perron–Frobenius Theorem which holds for primitive T; we shall generalize Theorem 1.1 to a wider class of matrices, called *irreducible*, in §1.4 (and shall refer to this generalization as the Perron–Frobenius Theory).

Suppose now the *distinct* eigenvalues of a primitive T are $r, \lambda_2, \ldots, \lambda_t$, $t \le n$ where $r > |\lambda_2| \ge |\lambda_3| \ge \cdots \ge |\lambda_t|$. In the case $|\lambda_2| = |\lambda_3|$ we stipulate that the multiplicity m_2 of λ_2 is at least as great as that of λ_3, *and of any other eigenvalue having the same modulus as* λ_2.

It may happen that a primitive matrix has $\lambda_2 = 0$; an example is a matrix of form

$$T = \begin{pmatrix} a & b & c \\ a & c & b \\ a & c & b \end{pmatrix} > 0 \tag{1.12}$$

for which $r = a + b + c$. This kind of situation gives the following theorem its dual form, the example (1.12) illustrating that in part (*b*) the bound $(n-1)$ cannot be reduced.

Theorem 1.2. *For a primitive matrix T:*

(*a*) *if $\lambda_2 \neq 0$, then as $k \to \infty$*

$$T^k = r^k \boldsymbol{w}\boldsymbol{v}' + 0(k^s |\lambda_2|^k)$$

elementwise, where $s = m_2 - 1$;

(*b*) *if $\lambda_2 = 0$, then for $k \ge n - 1$*

$$T^k = r^k \boldsymbol{w}\boldsymbol{v}'.$$

In both cases $\boldsymbol{w}$, $\boldsymbol{v}'$ are any positive right and left eigenvectors corresponding to r guaranteed by Theorem 1.1, providing only they are normed so that $\boldsymbol{v}'\boldsymbol{w} = 1$.

PROOF. Let $R(z) = (I - zT)^{-1} = \{r_{ij}(z)\}$, $z \neq \lambda_i^{-1}$, $i = 1, 2, \ldots$ (where $\lambda_1 = r$). Consider a general element of this matrix

$$r_{ij}(z) = \frac{c_{ij}(z)}{(1 - zr)(1 - z\lambda_2)^{m_2} \cdots (1 - z\lambda_t)^{m_t}}$$

where m_i is the multiplicity of λ_i and $c_{ij}(z)$ is a polynomial in z, of degree at most $n - 1$ (see Appendix B).

Here using partial fractions, *in case* (a)

$$r_{ij}(z) = p_{ij}(z) + \frac{a_{ij}}{(1 - zr)} + \sum_{s=0}^{m_2-1} \frac{b_{ij}^{(m_2-s)}}{(1 - z\lambda_2)^{m_2-s}}$$

$$+ \text{ similar terms for any other non-zero eigenvalues,}$$

where the a_{ij}, $b_{ij}^{(m_2-s)}$ are constants, and $p_{ij}(z)$ is a polynomial of degree at most $(n-2)$. Hence for $|z| < 1/r$,

$$r_{ij}(z) = p_{ij}(z) + a_{ij} \sum_{k=0}^{\infty} (zr)^k + \sum_{s=0}^{m_2-1} b_{ij}^{(m_2-s)} \left\{ \sum_{k=0}^{\infty} \binom{-m_2+s}{k} (-z\lambda_2)^k \right\}$$

$$+ \text{ similar terms for other non-zero eigenvalues.}$$

in matrix form, with obvious notation

$$R(z) = P(z) + A \sum_{k=0}^{\infty} (zr)^k + \sum_{s=0}^{m_2-1} B^{(m_2-s)} \left\{ \sum_{k=0}^{\infty} \binom{-m_2+s}{k} (-z\lambda_2)^k \right\}$$

$$+ \text{ possible like terms.}$$

From Stirling's formula, as $k \to \infty$

$$\binom{-m_2+s}{k} \sim \text{const. } k^{m_2-s-1},$$

so that, identifying coefficients of z^k on both sides (see Appendix B) for large k

$$T^k = Ar^k + 0(k^{m_2-1} |\lambda_2|^k).$$

In case (b), we have merely, with the same symbolism as in case (a)

$$r_{ij}(z) = p_{ij}(z) + \frac{a_{ij}}{(1-zr)}$$

so that for $k \geq n-1$,

$$T^k = Ar^k.$$

It remains to determine the nature of A. We first note that

$$T^k/r^k \to A \geq 0 \text{ elementwise, as } k \to \infty,$$

and that the series

$$\sum_{k=0}^{\infty} (r^{-1}T)^k z^k$$

has non-negative coefficients, and is convergent for $|z| < 1$, so that by a wellknown result (see e.g. Heathcote, 1971, p. 65).

$$\lim_{x \to 1-} (1-x) \sum_{k=0}^{\infty} (r^{-1}T)^k x^k = A \text{ elementwise.}$$

Now, for $0 < x < 1$,

$$\sum_{k=0}^{\infty} (r^{-1}T)^k x^k = (I - r^{-1}xT)^{-1} = \frac{\text{Adj } (I - r^{-1}xT)}{\det (I - r^{-1}xT)}$$

$$= \frac{r \text{ Adj } (rx^{-1}I - T)}{x \det (rx^{-1}I - T)}$$

so that
$$A = -r \text{ Adj } (rI - T)/c$$

where
$$c = \lim_{x \to 1-} \{-\det (rx^{-1}I - T)/(1-x)\}$$

$$= \frac{d}{dx} [\phi(rx^{-1})]_{x=1}$$

$$= -r\phi'(r)$$

which completes the proof, taking into account Corollary 2 of Theorem 1.1. □

In conclusion to this section we point out that assertion (d) of Theorem 1.1 states that the *geometric multiplicity* of the eigenvalue r is one, whereas (f) states that its *algebraic multiplicity* is one. It is well known in matrix theory that geometric multiplicity one for the eigenvalue of a square arbitrary matrix does not in general imply algebraic multiplicity one. A simple example to this end is the matrix (which is non-negative, but of course not primitive):

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

which has repeated eigenvalue unity (algebraic multiplicity two), but a corresponding left eigenvector can only be a multiple of $\{0, 1\}$ (geometric multiplicity one).

The distinction between geometric and algebraic multiplicity in connection with r in a primitive matrix is slurred over in some treatments of nonnegative matrix theory.

1.2 Structure of a General Non-negative Matrix

In this section we are concerned with a general square matrix $T = \{t_{ij}\}$, $i, j = 1, \ldots, n$, satisfying $t_{ij} \geq 0$, with the aim of showing that the behaviour of its powers T^k reduces, to a substantial extent, to the behaviour of powers of a fundamental type of non-negative square matrix, called *irreducible*. The class of irreducible matrices further subdivides into matrices which are *primitive* (studied in §1.1), and *cyclic* (imprimitive), whose study is taken up in §1.3.

We introduce here a definition, which while frequently used in other expositions of the theory, and so possibly useful to the reader, will be used by us only to a limited extent.

Definition 1.2. A sequence $(i, i_1, i_2, \ldots, i_{t-1}, j)$, for $t \geq 1$ (where $i_0 = i$), from the index set $\{1, 2, \ldots, n\}$ of a non-negative matrix T is said to form a *chain* of length t between the ordered pair (i, j) if

$$t_{ii_1} t_{i_1 i_2} \cdots t_{i_{t-2} i_{t-1}} t_{i_{t-1} j} > 0.$$

Such a chain for which $i = j$ is called a *cycle* of length t between i and itself.

Clearly in this definition, we may without loss of generality impose the restriction that, for fixed (i, j), $i, j \neq i_1 \neq i_2 \neq \cdots \neq i_{t-1}$, to obtain a 'minimal' length chain or cycle, from a given one.

Classification of indices

Let i, j, k be arbitrary indices from the index set $1, 2, \ldots, n$ of the matrix T.

We say that i *leads to* j, and write $i \to j$, if there exists an integer $m \geq 1$ such that $t_{ij}^{(m)} > 0$.[1] If i does not lead to j we write $i \not\to j$. Clearly, if $i \to j$ and $j \to k$ then, from the rule of matrix multiplication, $i \to k$.

We say that i and j *communicate* if $i \to j$ and $j \to i$, and write in this case $i \leftrightarrow j$.

The indices of the matrix T may then be classified and grouped as follows.

(*a*) If $i \to j$ but $j \not\to i$ for some j, then the index i is called *inessential*. An index which leads to no index at all (this arises when T has a row of zeros) is also called inessential.

(*b*) Otherwise the index i is called *essential*. Thus if i is essential, $i \to j$ implies $i \leftrightarrow j$; and there is at least one j such that $i \to j$.

(*c*) It is therefore clear that all essential indices (if any) can be subdivided into *essential classes* in such a way that all the indices belonging to one class communicate, but cannot lead to an index outside the class.

(*d*) Moreover, all inessential indices (if any) which communicate with some index, may be divided into *inessential classes* such that all indices in a class communicate.

Classes of the type described in (*c*) and (*d*) are called *self-communicating* classes.

(*e*) In addition there may be inessential indices which communicate with no index; these are defined as forming an *inessential class* by themselves (which, of course, if not self-communicating). These are of nuisance value only as regards applications, but are included in the description for completeness.

This description appears complex, but should be much clarified by the example which follows, and similar exercises.

Before proceeding, we need to note that the classification of indices (and hence grouping into classes) *depends only on the location of the positive elements, and not on their size*, so any two non-negative matrices with the same incidence matrix will have the same index classification and grouping (and, indeed, *canonical form*, to be discussed shortly).

Further, given a non-negative matrix (or its incidence matrix), classification and grouping of indices is made easy by a *path diagram* which may be described as follows. Start with index 1—this is the zeroth stage;

[1] Or, equivalently, if there is a chain between i and j.

determine all j such that $1 \to j$ and draw arrows to them—these j form the 2nd stage; for each of these j now repeat the procedure to form the 3rd stage; and so on; *but* as soon as an index occurs which has occurred at an earlier stage, ignore further consequents of it. Thus the diagram terminates when every index in it has repeated. (Since there are a finite total number of indices, the process must terminate.) This diagram will represent all possible consequent behaviour for the set of indices which entered into it, which may not, however, be the entire index set. If any are left over, choose one such and draw a similar diagram for it, regarding the indices of the previous diagram also as having occurred 'at an earlier stage'. And so on, till all indices of the index set are accounted for.

EXAMPLE. A non-negative matrix T has incidence matrix

$$\begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7\\8\\9 \end{array}\begin{array}{c} \begin{array}{ccccccccc} 1&2&3&4&5&6&7&8&9 \end{array} \\ \begin{bmatrix} 1&1&0&0&0&0&0&0&0\\ 1&1&1&0&0&0&1&0&0\\ 0&0&0&0&0&0&1&0&0\\ 0&0&0&1&0&0&0&0&1\\ 0&0&0&0&1&0&0&0&0\\ 0&0&1&0&0&1&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&1&0&0&0&1&0&1&0\\ 0&0&0&1&0&0&0&0&1 \end{bmatrix} \end{array}$$

Thus Diagram 1 tells us $\{3, 7\}$ is an essential class, while $\{1, 2\}$ is an inessential (communicating) class.
Diagram 2 tells us $\{4, 9\}$ is an essential class.

Diagram 1

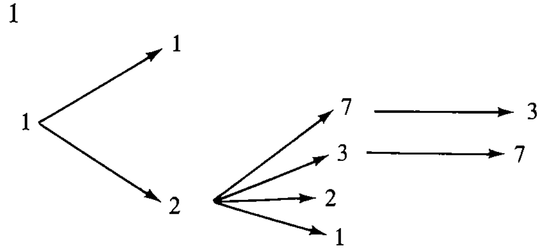


Diagram 2

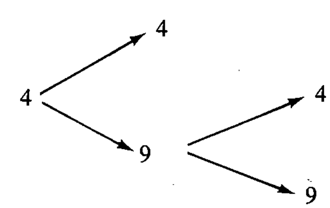


Diagram 3 tells us $\{5\}$ is an essential class.
Diagram 4 tells us $\{6\}$ is an inessential (self-communicating) class.
Diagram 5 tells us $\{8\}$ is an inessential (self-communicating) class.

Diagram 3

$$5 \longrightarrow 5$$

Diagram 4

$$\begin{tikzcd}[row sep=small] & 3 \text{ (Diagram 1)} \\ 6 \arrow[ur] \arrow[dr] & \\ & 6 \end{tikzcd}$$

Diagram 5

$$\begin{tikzcd}[row sep=small] & 2 \text{ (Diagram 1)} \\ 8 \arrow[ur] \arrow[r] \arrow[dr] & 6 \text{ (Diagram 4)} \\ & 8 \end{tikzcd}$$

Canonical Form

Once the classification and grouping has been carried out, the definition 'leads' may be extended to classes in the obvious sense e.g. the statement $\mathscr{C}_1 \to \mathscr{C}_2 (\mathscr{C}_1 \neq \mathscr{C}_2)$ means that there is an index of $\mathscr{C}_1$ which leads to an index of $\mathscr{C}_2$. Hence all indices of $\mathscr{C}_1$ lead to all indices of $\mathscr{C}_2$, and the statement can only apply to an inessential class $\mathscr{C}_1$.

Moreover, the matrix T may be put into *canonical form* by first relabelling the indices in a specific manner. Before describing the manner, we mention that relabelling the indices using the same index set $\{1, \ldots, n\}$ and rewriting T accordingly merely amounts to performing a *simultaneous permutation* of rows and columns of the matrix. Now such a simultaneous permutation only amounts to a *similarity transformation* of the original matrix, T, so that (*a*) its powers are similarly transformed; (*b*) its spectrum (i.e. set of eigenvalues) is unchanged. Generally any given ordering is as good as any other in a physical context; but the canonical form of T, arrived at by one such ordering, is particularly convenient.

The canonical form is attained by first taking the indices of one essential class (if any) and renumbering them consecutively using the lowest integers,

and following by the indices of another essential class, if any, until the essential classes are exhausted. The numbering is then extended to the indices of the inessential classes (if any) which are themselves arranged in an order such that an inessential class occurring earlier (and thus higher in the arrangement) does not *lead* to any inessential class occurring later.

EXAMPLE (continued). For the given matrix T the essential classes are $\{5\}$, $\{4, 9\}$, $\{3, 7\}$; and the inessential classes $\{1, 2\}$, $\{6\}$, $\{8\}$ which from Diagrams 4 and 5 should be ranked in this order. Thus a possible canonical form for T is

$$\begin{array}{c} \\ 5 \\ 4 \\ 9 \\ 3 \\ 7 \\ 1 \\ 2 \\ 6 \\ 8 \end{array} \begin{array}{c} \begin{array}{ccccccccc} 5 & 4 & 9 & 3 & 7 & 1 & 2 & 6 & 8 \end{array} \\ \left[\begin{array}{c|cc|cc|cc|c|c} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{array}\right] \end{array}.$$

It is clear that the canonical form consists of square diagonal blocks corresponding to 'transition within' the classes in one 'stage', zeros to the right of these diagonal blocks, but at least one non-zero element to the left of each inessential block unless it corresponds to an index which leads to no other. Thus the general version of the canonical form of T is

$$T = \left[\begin{array}{cccc|c} T_1\, 0 & \ldots & 0 \ldots & & 0 \\ 0\, T_2 & & & & \vdots \\ 0 & & & & \\ \vdots & & & & \\ 0\, 0 & \ldots & & T_z & 0 \\ \hline & R & & & Q \end{array}\right]$$

where the T_i correspond to the z essential classes, and Q to the inessential indices, with $R \neq 0$ in general, with Q itself having a structure analogous to T, except that there may be non-zero elements to the left of any of its diagonal blocks:

$$Q = \begin{bmatrix} Q_1 & & & \\ & Q_2 & & \\ & & \ddots & 0 \\ S & & & Q_w \end{bmatrix}.$$

Now, in most applications we are interested in the behaviour of the *powers* of T. Let us assume it is in canonical form. Since

$$T^k = \left[\begin{array}{cccc|c} T_1^k & & & & \\ & T_2^k & & & \\ 0 & & \ddots & 0 & 0 \\ & & & T_z^k & \\ \hline & & R_k & & Q^k \end{array}\right], \quad Q^k = \begin{bmatrix} Q_1^k & & & \\ & Q_2^k & & \\ & & \ddots & 0 \\ S_k & & & Q_w^k \end{bmatrix}$$

it follows that a substantial advance in this direction will be made in studying the powers of the diagonal *block submatrices corresponding to self—communicating classes* (the other diagonal block submatrices, if any, are 1×1 zero matrices; the evolution of R_k and S_k is complex, with k). In fact if one is interested in only the essential indices, as is often the case, this is sufficient.

A (sub)matrix corresponding to a single self-communicating class is called *irreducible*.

It remains to show that, *normally*, there is at least one self-communicating (indeed essential) class of indices present for any matrix T; although it is nevertheless *possible* that all indices of a non-negative matrix fall into non self-communicating classes (and are therefore inessential): for example

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

Lemma 1.1. *An $n \times n$ non-negative matrix with at least one positive entry in each row possesses at least one essential class of indices.*

PROOF. Suppose all indices are inessential. The assumption of non-zero rows then implies that for any index i, $i = 1, \ldots, n$, there is at least one j such that $i \to j$, but $j \not\to i$.

Now suppose i_1 is any index. Then we can find a sequence of indices i_2, $i_3, \ldots$ etc. such that

$$i_1 \to i_2 \to i_3 \to \cdots \to i_n \to i_{n+1} \cdots$$

but such that $i_{k+1} \not\to i_k$, and hence $i_{k+1} \not\to i_1, i_2, \ldots,$ or i_{k-1}. However, since the sequence $i_1, i_2, \ldots, i_{n+1}$ is a set of $n+1$ indices, each chosen from the same n possibilities, $1, 2, \ldots, n$, at least *one index repeats* in the sequence. This is a contradiction to the deduction that no index can lead to an index with a lower subscript. □

We come now to the important concept of the period of an index.

Definition 1.3. If $i \to i$, $d(i)$ is the *period* of the index i if it is the greatest common divisor of those k for which

$$t_{ii}^{(k)} > 0$$

(see Definition A.2 in Appendix A). N.B. If $t_{ii} > 0$, $d(i) = 1$.

We shall now prove that in a communicating class all indices have the same period.

Lemma 1.2. *If $i \leftrightarrow j$, $d(i) = d(j)$.*

PROOF. Let $t_{ij}^{(M)} > 0$, $t_{ji}^{(N)} > 0$. Then for any positive integer s such that $t_{jj}^{(s)} > 0$

$$t_{ii}^{(M+s+N)} \geq t_{ij}^{(M)} t_{jj}^{(s)} t_{ji}^{(N)} > 0,$$

the first inequality following from the rule of matrix multiplication and the non-negativity of the elements of T. Now, for such an s it is also true that $t_{jj}^{(2s)} > 0$ necessarily, so that

$$t_{ii}^{(M+2s+N)} > 0.$$

Therefore $\quad d(i)$ divides $M + 2s + N - (M + s + N) = s$.

Hence: for every s such that $t_{jj}^{(s)} > 0$, $d(i)$ divides s.

Hence $$d(i) \leq d(j).$$

But since the argument can be repeated with i and j interchanged,

$$d(j) \leq d(i).$$

Hence $d(i) = d(j)$ as required. □

Note that, again, consideration of an incidence matrix is adequate to determine the period.

Definition 1.4. The period of a communicating class is the period of any one of its indices.

EXAMPLE (continued): Determine the periods of all communicating classes for the matrix T with incidence matrix considered earlier.

Essential classes:

$\{5\}$ has period 1, since $t_{55} > 0$.

$\{4, 9\}$ has period 1, since $t_{44} > 0$.

$\{3, 7\}$ has period 2, since $t_{33}^{(k)} > 0$

for every even k, and is zero for every odd k.

Inessential self-communicating classes:

$\{1, 2\}$ has period 1 since $t_{11} > 0$.

$\{6\}$ has period 1 since $t_{66} > 0$.

$\{8\}$ has period 1 since $t_{88} > 0$.

Definition 1.5. An index i such that $i \to i$ is *aperiodic* (acyclic) if $d(i) = 1$. [It is thus contained in an *aperiodic* (*self-communicating*) *class*.]

1.3 Irreducible Matrices

Towards the end of the last section we called a non-negative square matrix, corresponding to a single self-communicating class of indices, irreducible. We now give a general definition, independent of the previous context, which is, nevertheless, easily seen to be equivalent to the one just given. The part of the definition referring to periodicity is justified by Lemma 1.2.

Definition 1.6. An $n \times n$ non-negative matrix T is *irreducible* if for every pair i, j of its index set, there exists a positive integer $m \equiv m(i, j)$ such that $t_{ij}^{(m)} > 0$. An irreducible matrix is said to be cyclic (periodic) with period d, if the period of any one (and so of each one) of its indices satisfies $d > 1$, and is said to be acyclic (aperiodic) if $d = 1$.

The following results all refer to an irreducible matrix with period d. Note that an irreducible matrix T cannot have a zero row or column.

Lemma 1.3. *If $i \to i$, $t_{ii}^{(kd)} > 0$ for all integers $k \geq N_0(=N_0(i))$.*

Proof.

Suppose $$t_{ii}^{(kd)} > 0, \; t_{ii}^{(sd)} > 0.$$

Then $$t_{ii}^{([k+s]d)} \geq t_{ii}^{(kd)} t_{ii}^{(sd)} > 0.$$

Hence the positive integers $\{kd\}$ such that

$$t_{ii}^{(kd)} > 0,$$

form a closed set under addition, and their greatest common divisor is d. An appeal to Lemma A.3 of Appendix A completes the proof. □

Theorem 1.3. *Let i be any fixed index of the index set $\{1, 2, \ldots, n\}$ of T. Then, for every index j there exists a unique integer r_j in the range $0 \leq r_j < d$ (r_j is called a residue class modulo d) such that*

(a) $t_{ij}^{(s)} > 0$ implies $s \equiv r_j \pmod d$;[1] *and*
(b) $t_{ij}^{(kd+r_j)} > 0$ for $k \geq N(j)$, where $N(j)$ is some positive integer.

Proof. Let $t_{ij}^{(m)} > 0$ and $t_{ij}^{(m')} > 0$.
There exists a p such that $t_{ji}^{(p)} > 0$, whence as before

$$t_{ii}^{(m+p)} > 0 \text{ and } t_{ii}^{(m'+p)} > 0.$$

Hence d divides each of the superscripts, and hence their difference $m - m'$. Thus $m - m' \equiv 0 \pmod d$, so that

$$m \equiv r_j \pmod d.$$

[1] Recall from Appendix A, that this means that if qd is the multiple of d nearest to s from below, then $s = qd + r_j$; it reads 's is equivalent to r_j, modulo d'.

This proves (a).

To prove (b), since $i \to j$ and in view of (a), there exists a positive m such that

$$t_{ij}^{(md+r_j)} > 0.$$

Now, let $N(j) = N_0 + m$, where N_0 is the number guaranteed by Lemma 1.3 for which $t_{ii}^{(sd)} > 0$ for $s \geq N_0$. Hence if $k \geq N(j)$, then

$$kd + r_j = sd + md + r_j, \text{ where } s \geq N_0.$$

Therefore $t_{ij}^{(kd+r_j)} \geq t_{ii}^{(sd)} t_{ij}^{(md+r_j)} > 0$, for all $k \geq N(j)$. □

Definition 1.7. The set of indices j in $\{1, 2, \ldots, n\}$ corresponding to the same residue class (mod d) is called a subclass of the class $\{1, 2, \ldots, n\}$, and is denoted by C_r, $0 \leq r < d$.

It is clear that the d subclasses C_r are disjoint, and their union is $\{1, 2, \ldots, n\}$. It is not yet clear that the composition of the classes does not depend on the choice of initial fixed index i, which we prove in a moment; nor that each subclass contains at least one index.

Lemma 1.4. *The composition of the residue classes does not depend on the initial choice of fixed index i; an initial choice of another index merely subjects the subclasses to a cyclic permutation.*

Proof. Suppose we take a new fixed index i'. Then

$$t_{ij}^{(md+r_{i'}+kd+r'_j)} \geq t_{ii'}^{(kd+r_{i'})} t_{i'j}^{(md+r'_j)}$$

where r'_j denotes the residue class corresponding to j according to classification with respect to fixed index i'. Now, by Theorem 1.3 for large k, m, the right hand side is positive, so that the left hand side is also, whence, in the old classification,

$$md + r_{i'} + kd + r'_j \equiv r_j \pmod d$$

i.e. $$r_{i'} + r'_j \equiv r_j \pmod d.$$

Hence the composition of the subclasses $\{C_i\}$ is unchanged, and their order is merely subjected to a cyclic permutation σ:

$$\begin{pmatrix} 0 & 1 & \cdots & d-1 \\ \sigma(0) & \sigma(1) & \cdots & \sigma(d-1) \end{pmatrix}.$$ □

For example, suppose we have a situation with $d = 3$, and $r_{i'} = 2$. Then the classes which were C_0, C_1, C_2 in the old classification according to i (according to which $i' \in C_2$) now become, respectively, C'_1, C'_2, C'_0 since we must have $2 + r'_j \equiv r_j \pmod d$ for $r_j = 0, 1, 2$.

Let us now define C_r for all non-negative integers r by putting $C_r = C_{r_j}$ if $\equiv r_j \pmod d$, using the initial classification with respect to i. Let m be a ositive integer, and consider any j for which $t_{ij}^{(m)} > 0$. (There is at least one ppropriate index j, otherwise T^m (and hence higher powers) would have ith ow consisting entirely of zeros, contrary to irreducibility of T.) Then $m \equiv r_j$ mod d), i.e. $m = sd + r_j$ and $j \in C_{r_j}$, Now, similarly, let k be any index such hat

$$t_{ik}^{(m+1)} > 0.$$

hen, since $m + 1 = sd + r_j + 1$, it follows $k \in C_{r_j+1}$.

Hence it follows that, looking at the ith row, the positive entries occur, for uccessive powers, in successive subclasses. In particular each of the d cyclic lasses is non-empty. If subclassification has initially been made according to he index i', since we have seen the subclasses are merely subjected to a cyclic ermutation, the classes still 'follow each other' in order, looking at succesive powers, and ith (hence any) row.

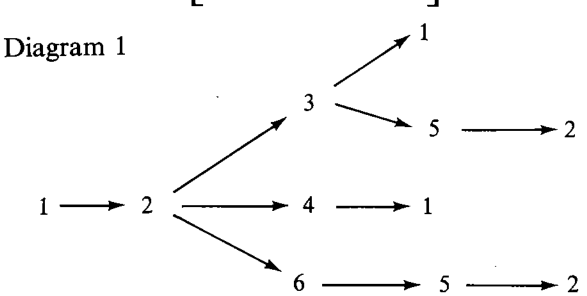
It follows that if $d > 1$ (so there is more than one subclass) a canonical orm of T is possible, by relabelling the indices so that the indices of C_0 come irst, of C_1 next, and so on. This produces a version of T of the sort

$$T_c = \begin{bmatrix} 0 & Q_{01} & 0 & \cdots & & 0 \\ 0 & 0 & Q_{12} & \cdots & & 0 \\ \cdot & & 0 & & & \\ \cdot & \cdot & & & & \\ 0 & \cdot & \cdot & & 0 & Q_{d-2,\,d-1} \\ Q_{d-1,\,0} & 0 & 0 & \cdots & & 0 \end{bmatrix}$$

EXAMPLE: Check that the matrix, whose incidence matrix is given below is rreducible, find its period, and put into a canonical form if periodic.

$$\begin{array}{c c} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

Diagram 1



Clearly $i \to j$ for any i and j in the index set, so the matrix is certainly irreducible. Let us now carry out the determination of subclasses on the basis of index 1. Therefore index 1 must be in the subset C_0; 2 must be in C_1; 3, 4, 6 in C_2; 1, 5 in C_3; 2 in C_4. Hence C_0 and C_3 are identical; C_1 and C_4; etc., and so $d = 3$. Moreover

$$C_0 = \{1, 5\},\ C_1 = \{2\},\ C_2 = \{3, 4, 6\},$$

so canonical form is

$$\begin{array}{c c} & \begin{matrix} 1 & 5 & 2 & 3 & 4 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 5 \\ 2 \\ 3 \\ 4 \\ 6 \end{matrix} & \left[\begin{array}{cc|c|ccc} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 1 & 1 \\ \hline 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}.$$

Theorem 1.4. *An irreducible acyclic matrix T is primitive and conversely. The powers of an irreducible cyclic matrix may be studied in terms of powers of primitive matrices.*

PROOF. If T is irreducible, with $d = 1$, there is only one subclass of the index set, consisting of the index set itself, and Theorem 1.3 implies

$$t_{ij}^{(k)} > 0 \quad \text{for} \quad k \geq N(i, j).$$

Hence for $N^* = \max_{i,j} N(i, j)$

$$t_{ij}^{(k)} > 0,\ k \geq N^*, \quad \text{for all } i, j.$$

i.e.

$$T^k > 0 \quad \text{for} \quad k \geq N^*.$$

Conversely, a primitive matrix is trivially irreducible, and has $d = 1$, since for any fixed i, and k great enough $t_{ii}^{(k)} > 0$, $t_{ii}^{(k+1)} > 0$, and the greatest common divisor of k and $k + 1$ is 1.

The truth of the second part of the assertion may be conveniently demonstrated in the case $d = 3$, where the canonical form of T is

$$T_c = \begin{bmatrix} 0 & Q_{01} & 0 \\ 0 & 0 & Q_{12} \\ Q_{20} & 0 & 0 \end{bmatrix},$$

$$\text{and } T_c^2 = \begin{bmatrix} 0 & 0 & Q_{01}Q_{12} \\ Q_{12}Q_{20} & 0 & 0 \\ 0 & Q_{20}Q_{01} & 0 \end{bmatrix},$$

$$T_c^3 = \begin{bmatrix} Q_{01}Q_{12}Q_{20} & 0 & 0 \\ 0 & Q_{12}Q_{20}Q_{01} & 0 \\ 0 & 0 & Q_{20}Q_{01}Q_{12} \end{bmatrix}.$$

Now, the diagonal matrices of T_c^3 (of T_c^d in general) are square and *primitive*, for Lemma 1.3 states that $t_{ii}^{(3k)} > 0$ for all k sufficiently large. Hence

$$T_c^{3k} = (T_c^3)^k,$$

so that powers which are integral multiples of the period may be studied with the aid of the primitive matrix theory of §1.1. One needs to consider also

$$T_c^{3k+1} \quad \text{and} \quad T_c^{3k+2}$$

but these present no additional difficulty since we may write $T_c^{3k+1} = (T_c^{3k})T_c$, $T_c^{3k+2} = (T_c^{3k})T_c^2$ and proceed as before. □

These remarks substantiate the reason for considering primitive matrices as of prime importance, and for treating them first. It is, nevertheless, convenient to consider a theorem of the type of the fundamental Theorem 1.1 for the broader class of irreducible matrices, which we now expect to be closely related.

1.4 Perron–Frobenius Theory for Irreducible Matrices

Theorem 1.5. *Suppose T is an $n \times n$ irreducible non-negative matrix. Then all of the assertions* (*a*)–(*f*) *of Theorem* 1.1 *hold, except that* (*c*) *is replaced by the weaker statement*: $r \geq |\lambda|$ *for any eigenvalue* λ *of* T. *Corollaries* 1 *and* 2 *of Theorem* 1.1 *hold also.*

PROOF. The proof of (*a*) of Theorem 1.1 holds to the stage where we need to assume

$$z' = \hat{x}'T - r\hat{x}' \geq \mathbf{0}' \quad \text{but} \quad \neq \mathbf{0}'.$$

The matrix $I + T$ is primitive, hence for some k, $(I + T)^k > 0$; hence

$$z'(I + T)^k = \{\hat{x}'(I + T)^k\}T - r\{\hat{x}'(I + T)^k\} > \mathbf{0}$$

which contradicts the definition of r; (*b*) is then proved as in Theorem 1.6 following; and the rest follows as before, except for the last part in (*c*). □

We shall henceforth call r the Perron–Frobenius eigenvalue of an irreducible T, and its corresponding positive eigenvectors, the Perron–Frobenius eigenvectors.

The above theorem does not answer in detail questions about eigenvalues λ such that $\lambda \neq r$ but $|\lambda| = r$ in the cyclic case.

The following auxiliary result is more general than we shall require immediately, but is important in future contexts.

Theorem 1.6. (*The Subinvariance Theorem*). *Let T be a non-negative irreducible matrix, s a positive number, and $y \geq \mathbf{0}$, $\neq \mathbf{0}$, a vector satisfying*

$$Ty \leq sy.$$

Then (*a*) $y > \mathbf{0}$; (*b*) $s \geq r$, *where r is the Perron–Frobenius eigenvalue of T. Moreover, $s = r$ if and only if $Ty = sy$.*

PROOF. Suppose at least one element, say the ith, of y is zero. Then since $T^k y \leq s^k y$ it follows that

$$\sum_{j=1}^{n} t_{ij}^{(k)} y_j \leq s^k y_i.$$

Now, since T is irreducible, for this i and any j, there exists a k such that $t_{ij}^{(k)} > 0$; and since $y_i > 0$ for some j, it follows that

$$y_i > 0$$

which is a contradiction. Thus $y > \mathbf{0}$. Now, premultiplying the relation $Ty \leq sy$ by $\hat{x}'$, a positive left eigenvector of T corresponding to r,

$$s\hat{x}'y \geq \hat{x}'Ty = r\hat{x}'y$$

i.e.

$$s \geq r.$$

Now suppose $Ty \leq ry$ with strict inequality in at least one place; then the preceding argument, on account of the strict positivity of Ty and ry, yields $r < r$, which is impossible. The implication $s = r$ follows from $Ty = sy$ similarly. □

In the sequel, any subscripts which occur should be understood as reduced modulo d, to bring them into the range $[0, d - 1]$, if they do not already fall in the range.

Theorem 1.7. *For a cyclic matrix T with period $d > 1$, there are present precisely d distinct eigenvalues λ with $|\lambda| = r$, where r is the Perron–Frobenius eigenvalue of T. These eigenvalues are*: $r \exp i2\pi k/d$, $k = 0, 1, \ldots, d - 1$ (*i.e. the d roots of the equation* $\lambda^d - r^d = 0$).

PROOF. Consider an arbitrary one, say the ith, of the primitive matrices:

$$Q_{i,\, i+1} Q_{i+1,\, i+2} \cdots Q_{i+d-1,\, i+d}$$

occurring as diagonal blocks in the dth power, T^d, of the canonical form T_c of T (recall that T_c has the same eigenvalues as T), and denote by $r(i)$ its Perron–Frobenius eigenvalue, and by $y(i)$ a corresponding positive right eigenvector, so that

$$Q_{i,\, i+1} Q_{i+1,\, i+2} \cdots Q_{i+d-1,\, i+d}\, y(i) = r(i) y(i).$$

Now premultiply this by $Q_{i-1,i}$:

$$Q_{i-1,i}Q_{i,i+1}Q_{i+1,i+2}\cdots Q_{i+d-2,i+d-1}Q_{i+d-1,i+d}y(i) = r(i)Q_{i-1,i}y(i),$$

and since $Q_{i+d-1,i+d} \equiv Q_{i-1,i}$, we have

$$Q_{i-1,i}Q_{i,i+1}Q_{i+1,i+2}\cdots Q_{i+d-2,i+d-1}(Q_{i-1,i}y(i)) = r(i)(Q_{i-1,i}y(i))$$

whence it follows from Theorem 1.6 that $r(i) \geq r(i-1)$. Thus

$$r(0) \geq r(d-1) \geq r(d-2) \cdots \geq r(0),$$

so that, for all i, $r(i)$ is constant, say $\tilde{r}$, and so there are precisely d dominant eigenvalues of T^d, all the other eigenvalues being strictly smaller in modulus. Hence, since the eigenvalues of T^d are dth powers of the eigenvalues of T, there must be precisely d dominant roots of T, and all must be dth roots of $\tilde{r}$. Now, from Theorem 1.5, the positive dth root is an eigenvalue of T and is r. Thus every root λ of T such that $|\lambda| = r$ must be of the form

$$\lambda = r \exp \mathrm{i}(2\pi k/d),$$

where k is one of $0, 1, \ldots, d-1$, and there are d of them.

It remains to prove that there are no coincident eigenvalues, so that in fact all possibilities $r \exp \mathrm{i}(2\pi k/d)$, $k = 0, 1, \ldots, d-1$ occur.

Suppose that y is a positive $(n \times 1)$ right eigenvector corresponding to the Perron–Frobenius eigenvalue r of T_c (i.e. T written out in canonical form), and let y_j, $j = 0, \ldots, d-1$ be the subvector of components corresponding to subclass C_j.

Thus
$$y' = [y'_0, y'_1, \ldots, y'_{d-1}]$$
and
$$Q_{j,j+1}y_{j+1} = ry_j.$$

Now, let $\bar{y}_k$, $k = 0, 1, \ldots, d-1$ be the $(n \times 1)$ vector obtained from y by making the transformation

$$y_j \to \exp \mathrm{i}\left(\frac{2\pi jk}{d}\right)y_j$$

of its components as defined above. It is easy to check that $\bar{y}_0 = y$, and indeed that $\bar{y}_k$, $k = 0, 1, \ldots, d-1$ is an eigenvector corresponding to an eigenvector $r \exp \mathrm{i}(2\pi k/d)$, as required. This completes the proof of the theorem. □

We note in conclusion the following corollary on the structure of the eigenvalues, whose validity is now clear from the immediately preceding.

Corollary. *If $\lambda \neq 0$ is any eigenvalue of T, then the numbers $\lambda \exp \mathrm{i}(2\pi k/d)$, $k = 0, 1, \ldots, d-1$ are eigenvalues also. (Thus, rotation of the complex plane about the origin through angles of $2\pi/d$ carries the set of eigenvalues into itself.)*

Bibliography and Discussion

§1.1. and §1.4

The exposition of the chapter centres on the notion of a primitive non-negative matrix as the fundamental notion of non-negative matrix theory. The approach seems to have the advantage of proving the fundamental theorem of nonnegative matrix theory at the outset, and of avoiding the slight awkwardness entailed in the usual definition of irreducibility merely from the permutable structure of T.

The fundamental results (Theorems 1.1, 1.5 and 1.7) are basically due to Perron (1907) and Frobenius (1908, 1909, 1912), Perron's contribution being associated with strictly positive T. Many modern expositions tend to follow the simple and elegant paper of Wielandt (1950) (whose approach was anticipated in part by Lappo-Danilevskii (1934)); see e.g. Cherubino (1957), Gantmacher (1959) and Varga (1962). This is essentially true also of our proof of Theorem 1.1 (=Theorem 1.5) with some slight simplifications, especially in the proof of part (*e*), under the influence of the well-known paper of Debreu & Herstein (1953), which deviates otherwise from Wielandt's treatment also in the proof of (*a*). (The proof of Corollary 1 of Theorem 1.1 also follows Debreu & Herstein.)

The proof of Theorem 1.7 is not, however, associated with Wielandt's approach, due to an attempt to bring out, again, the primacy of the primitivity property. The last part of the proof (that all dth roots of r are involved), as well as the corollary, follows Romanovsky (1936). The possibility of evolving §1.4 in the present manner depends heavily on §1.3.

For other approaches to the Perron–Frobenius theory see Bellman (1960, Chapter 16), Brauer (1957*b*), Fan (1958), Householder (1958), Karlin (1959, §8.2; 1966, Appendix), Pullman (1971), Samelson (1957) and Sevastyanov (1971, Chapter 2). Some of these references do not deal with the most general case of an irreducible matrix, containing restrictions of one sort or another. In their recent treatise on non-negative matrices, Berman and Plemmons (1979) begin with a chapter studying the spectral properties of the set of $n \times n$ matrices which leave a proper cone in R^n invariant, combining the use of the Jordan normal form of a matrix, matrix norms and some assumed knowledge of cones. These results are specialized to non-negative matrices in their Chapter 2; and together with additional direct proofs give the full structure of the Perron–Frobenius theory. We have sought to present this theory in a simpler fashion, at a lower level of mathematical conception and technique.

Finally we mention that Schneider's (1977) survey gives, interalia, a historical survey of the concept of irreducibility.

§1.2 and §1.3

The development of these sections is motivated by probabilistic considerations from the theory of Markov chains, where it occurs in connection with *stochastic* non-negative matrices $P = \{p_{ij}\}, i, j = 1, 2, \ldots$, with $p_{ij} \geq 0$ and

$$\sum_j p_{ij} = 1, \qquad i = 1, 2, \ldots.$$

In this setting the classification theory is essentially due to Kolmogorov (1936); an account may be found in the somewhat more general exposition of Chung (1967, Chapter 1, §3), which our exposition tends to follow.

A weak analogue of the Perron–Frobenius Theorem for any square $T \geq 0$ is given as Exercise 1.12. Another approach to Perron–Frobenius-type theory in this general case is given by Rothblum (1975), and taken up in Berman and Plemmons (1979, §2.3).

Just as in the case of stochastic matrices, the corresponding exposition *is not restricted to finite matrices* (this in fact being the reason for the development of this kind of classification in the probabilistic setting), and virtually all of the present exposition goes through for infinite non-negative matrices T, so long as all powers T^k, $k = 1, 2, \ldots$ exist (with an obvious extension of the rule of matrix multiplication of finite matrices). This point is taken up again to a limited extent in Chapters 5 and 6, where infinite T are studied.

The reader acquainted with graph theory will recognize its relationship with the notion of path diagrams used in our exposition. For development along the lines of graph theory see Rosenblatt (1957), the brief account in Varga (1962, Chapters 1 and 2), Paz (1963) and Gordon (1965, Chapter 1). The relevant notions and usage in the setting of non-negative matrices implicitly go back at least to Romanovsky (1936).

Another development, not explicitly graph theoretical, is given in the papers of Pták (1958) and Pták & Sedlaček (1958); and it is utilized to some extent in §2.4 of the next chapter.

Finite stochastic matrices and finite Markov chains will be treated in Chapter 4. The general infinite case will be taken up in Chapter 5.

Exercises

1.1 Find all essential and inessential classes of a non-negative matrix with incidence matrix:

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

Find the periods of all self-communicating classes, and write the matrix T in full canonical form, so that the matrices corresponding to all self-communicating classes are also in canonical form.

1.2. Keeping in mind Lemma 1.1, construct a non-negative matrix T whose index set contains no essential class, but has, nevertheless, a self-communicating class.

1.3. Let $T = \{t_{i,j}\}, i, j = 1, 2, \ldots, n$ be a non-negative matrix. If, for some fixed i and j, $t_{i,j}^{(k)} > 0$ for some $k \equiv k(i, j)$, show that there exists a sequence $k_1, k_2, \ldots, k_r$ such that

$$t_{i,k_1} t_{k_1,k_2} \cdots t_{k_{r-1},k_r} t_{k_r,j} > 0$$

where $r \leq n - 2$ if $i \neq j$, $r \leq n - 1$ if $i = j$. Hence show that:

(*a*) if T is irreducible and $t_{j,j} > 0$ for some j, then $t_{i,j}^{(k)} > 0$ for $k \geq n - 1$ and every i; and, hence, if $t_{j,j} > 0$ for every j, then $T^{n-1} > 0$;

(*b*) T is irreducible if and only if $(I + T)^{n-1} > 0$.

(Wielandt, 1960; Herstein, 1954.)

Further results along the lines of (*a*) are given as Exercises 2.17–2.19, and again in Lemma 3.9.

1.4. Given $T = \{t_{i,j}\}, i, j = 1, 2, \ldots, n$ is a non-negative matrix, suppose that for some power $m \geq 1$, $T^m = \{t_{i,j}^{(m)}\}$ is such that

$$t_{i,i+1}^{(m)} > 0, \; i = 1, 2, \ldots, n-1, \quad \text{and} \quad t_{n,1}^{(m)} > 0.$$

Show that: T is irreducible; and (by example) that it may be periodic.

1.5. By considering the vector $x' = (\alpha, \alpha, 1 - 2\alpha)$, suitably normed, when: (i) $\alpha = 0$, (ii) $0 < \alpha < \frac{1}{2}$, and the matrix

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 1 & 0 & 2 \end{bmatrix}.$$

show that $r(x)$, as defined in the proof of Theorem 1.1 is not continuous in $x \geq \mathbf{0}$, $x'x = 1$.

(Schneider, 1958)

1.6.[1] If r is the Perron–Frobenius eigenvalue of an irreducible matrix $T = \{t_{ij}\}$, show that for any vector $x \in \mathscr{P}$, where $\mathscr{P} = \{x; x > \mathbf{0}\}$

$$\min_i \frac{\sum_j t_{ij} x_j}{x_i} \leq r \leq \max_i \frac{\sum_j t_{ij} x_j}{x_i}.$$

(Collatz, 1942)

1.7. Show, in the situation of Exercise 1.6, that equality on either side implies equality on both; and by considering when this can happen show that r is the supremum of the left hand side, and the infimum of the right hand side, over $x \in \mathscr{P}$, and is actually attained as both supremum and infimum for vectors in $\mathscr{P}$.

[1] Exercises 1.6 to 1.8 have a common theme.

1.8. In the framework of Exercise 1.6, show that

$$\max_{x \in \mathscr{P}} \left\{ \min_{y \in \mathscr{P}} \frac{y'Tx}{y'x} \right\} = r = \min_{y \in \mathscr{P}} \left\{ \max_{x \in \mathscr{P}} \frac{y'Tx}{y'x} \right\}.$$

(Birkhoff and Varga, 1958)

1.9.[1] Let B be a matrix with possibly complex elements and denote by B_+ the matrix of moduli of elements of B and β an eigenvalue of B. Let T be irreducible and such that $0 \le B_+ \le T$. Show that $|\beta| \le r$; and moreover that $|\beta| = r$ implies $B_+ = T$, where r is the Perron–Frobenius eigenvalues of T.

(Frobenius, 1909)

1.10. If, in Exercise 1.9, $|\beta| = r$, so that $\beta = re^{i\theta}$, say, it can be shown (Wielandt, 1950) that B has the representation

$$B = e^{i\theta} DTD^{-1}$$

where D is a diagonal matrix whose diagonal elements have modulus one. Show as consequences:
(i) that if $|\beta| = r$, $B_+ = T$;
(ii) that given there are d dominant eigenvalues of modulus r for a given periodic irreducible matrix of period d, they must in fact all be simple, and take on the values $r \exp i(2\pi j/d)$, $j = 0, 1, \ldots, d-1$. (Put $B = T$ in the representation.)

1.11. Let T be an irreducible non-negative matrix and E a non-zero non-negative matrix of the same size. If x is a positive number, show that $A = xE + T$ is irreducible, and that its Perron–Frobenius eigenvalue may be made to equal any positive number exceeding the Perron–Frobenius eigenvalue r of T by suitable choice of x.
(Consider first, for orientation, the situation where at least one diagonal element of E is positive. Make eventual use of the continuity of the eigenvalues of A with x.)

(Birkhoff & Varga, 1958)

1.12. If $T \ge 0$ is any square non-negative matrix, use the canonical form of T to show that the following weak analogue of the Perron–Frobenius Theorem holds: there exists an eigenvalue ρ such that
(a') ρ real, ≥ 0;
(b') with ρ can be associated non-negative left and right eigenvectors;
(c') $\rho \ge |\lambda|$ for any eigenvalue λ of T;
(e') if $0 \le B \le T$ and β is an eigenvalue of B, then $|\beta| \le \rho$.
(In such problems it is often useful to consider a sequence of matrices each $\ge T$ and converging to T elementwise {particularly in relation to (b') here}—Debreu and Herstein (1953).)

1.13. Show in relation to Exercise 1.12, that $\rho > 0$ if and only if T contains a cycle of elements.

(Ullman, 1952)

[1] Exercises 1.9 to 1.11 have a common theme.

1.14. Use the relevant part of Theorem 1.4, in conjunction with Theorem 1.2, to show that for an irreducible T with Perron–Frobenius eigenvalue r, as $k \to \infty$

$$s^{-k}T^k \to 0$$

if and only if $s > r$; and if $0 < s < r$, for each pair (i, j)

$$\limsup_{k \to \infty} s^{-k} t_{ij}^{(k)} = \infty.$$

Hence deduce that the power series

$$T_{ij}(z) = \sum_{k=0}^{\infty} t_{ij}^{(k)} z^k$$

have common convergence radius $R = r^{-1}$ for each pair (i, j). (This result is relevant to the development of the theory of countable irreducible T in Chapter 6.)

1.15. Let T be a non-negative matrix. Show that:
(a) $Ty \le sy$, where $s \ne 0$; $y \ge \mathbf{0}, \ne \mathbf{0} \Rightarrow y > \mathbf{0}$ if and only if T is irreducible;
(b) T has a single non-negative (left or right) eigenvector (to constant multiples) and this eigenvector is positive if and only if T is irreducible.

1.16. If A and B are non-negative matrices such that $0 \le B \le A$, $A - B \ne 0$, and $A + B$ is irreducible, show that $\rho(B) < \rho(A)$ where $\rho(\cdot)$ is the eigenvalue alluded to in Exercise 1.12.

1.17. Let T be a non-negative irreducible matrix, s a positive number, and $y \ge \mathbf{0}, \ne \mathbf{0}$ a vector satisfying

$$Ty \ge sy.$$

Show that $r \ge s$, where r is the Perron-Frobenius eigenvector of T, and $s = r$ if and only if $Ty = sy$. [This is a dual result to the (Subinvariance) Theorem 1.6.]

1.18. Suppose T is a non-negative matrix which, by simultaneous permutation of rows and columns may be put in the form

$$\begin{bmatrix} 0 & T_1 & 0 & \cdots & 0 \\ 0 & 0 & T_2 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & T_{d-1} \\ T_d & 0 & 0 & \cdots & 0 \end{bmatrix}$$

where the zero blocks on the diagonal are square. If T has no zero rows or columns, and $T_1 T_2 \cdots T_d$ is irreducible, show using Exercise 1.15(a), that T is irreducible. [*Hint:* Consider $y \ge \mathbf{0}, \ne \mathbf{0}$ partitioned according to $y' = [y_1', y_2', \ldots, y_d']$ where y_i has as many entries as the columns of T_i. Assuming $Ty \le sy$ for some $s > 0$, show first that $y_1 > \mathbf{0}$, and then that $y_{i+1} > \mathbf{0} \Rightarrow y_i > \mathbf{0}$, $i = 1, \ldots, d$, $y_{d+1} \equiv y_1$.]

(Minc, 1974, Pullman, 1975)