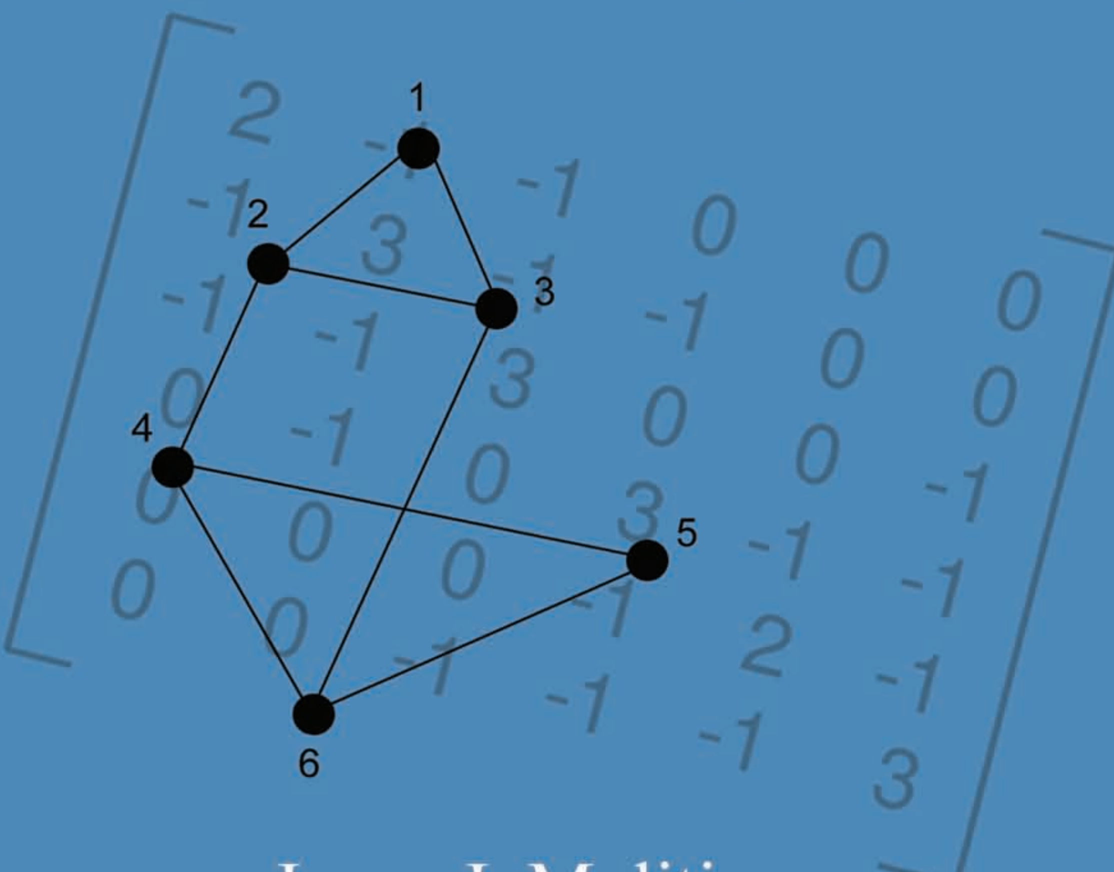
DISCRETE MATHEMATICS AND ITS APPLICATIONS
Series Editor KENNETH H. ROSEN

APPLICATIONS OF COMBINATORIAL MATRIX THEORY TO LAPLACIAN MATRICES OF GRAPHS



Jason J. Molitierno

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

APPLICATIONS OF COMBINATORIAL MATRIX THEORY TO LAPLACIAN MATRICES OF GRAPHS

DISCRETE MATHEMATICS AND ITS APPLICATIONS

Series Editor
Kenneth H. Rosen, Ph.D.

R. B. J. T. Allenby and Alan Slomson, How to Count: An Introduction to Combinatorics, Third Edition

Juergen Bierbrauer, Introduction to Coding Theory

Katalin Bimbó, Combinatory Logic: Pure, Applied and Typed

Donald Bindner and Martin Erickson, A Student's Guide to the Study, Practice, and Tools of Modern Mathematics

Francine Blanchet-Sadri, Algorithmic Combinatorics on Partial Words

Richard A. Brualdi and Dragoš Cvetković, A Combinatorial Approach to Matrix Theory and Its Applications

Kun-Mao Chao and Bang Ye Wu, Spanning Trees and Optimization Problems

Charalambos A. Charalambides, Enumerative Combinatorics

Gary Chartrand and Ping Zhang, Chromatic Graph Theory

Henri Cohen, Gerhard Frey, et al., Handbook of Elliptic and Hyperelliptic Curve Cryptography

Charles J. Colbourn and Jeffrey H. Dinitz, Handbook of Combinatorial Designs, Second Edition

Martin Erickson, Pearls of Discrete Mathematics

Martin Erickson and Anthony Vazzana, Introduction to Number Theory

Steven Furino, Ying Miao, and Jianxing Yin, Frames and Resolvable Designs: Uses, Constructions, and Existence

Mark S. Gockenbach, Finite-Dimensional Linear Algebra

Randy Goldberg and Lance Riek, A Practical Handbook of Speech Coders

Jacob E. Goodman and Joseph O'Rourke, Handbook of Discrete and Computational Geometry, Second Edition

Jonathan L. Gross, Combinatorial Methods with Computer Applications

Jonathan L. Gross and Jay Yellen, Graph Theory and Its Applications, Second Edition

Jonathan L. Gross and Jay Yellen, Handbook of Graph Theory

David S. Gunderson, Handbook of Mathematical Induction: Theory and Applications

Richard Hammack, Wilfried Imrich, and Sandi Klavžar, Handbook of Product Graphs, Second Edition

Darrel R. Hankerson, Greg A. Harris, and Peter D. Johnson, Introduction to Information Theory and Data Compression, Second Edition

Darel W. Hardy, Fred Richman, and Carol L. Walker, Applied Algebra: Codes, Ciphers, and Discrete Algorithms, Second Edition

Daryl D. Harms, Miroslav Kraetzl, Charles J. Colbourn, and John S. Devitt, Network Reliability: Experiments with a Symbolic Algebra Environment

Silvia Heubach and Toufik Mansour, Combinatorics of Compositions and Words

Leslie Hogben, Handbook of Linear Algebra

Derek F. Holt with Bettina Eick and Eamonn A. O'Brien, Handbook of Computational Group Theory

David M. Jackson and Terry I. Visentin, An Atlas of Smaller Maps in Orientable and Nonorientable Surfaces

Richard E. Klima, Neil P. Sigmon, and Ernest L. Stitzinger, Applications of Abstract Algebra with Maple™ and MATLAB®, Second Edition

Patrick Knupp and Kambiz Salari, Verification of Computer Codes in Computational Science and Engineering

William Kocay and Donald L. Kreher, Graphs, Algorithms, and Optimization

Donald L. Kreher and Douglas R. Stinson, Combinatorial Algorithms: Generation Enumeration and Search

Hang T. Lau, A Java Library of Graph Algorithms and Optimization

C. C. Lindner and C. A. Rodger, Design Theory, Second Edition

Nicholas A. Loehr, Bijective Combinatorics

Alasdair McAndrew, Introduction to Cryptography with Open-Source Software

Elliott Mendelson, Introduction to Mathematical Logic, Fifth Edition

Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone, Handbook of Applied Cryptography

Stig F. Mjølsnes, A Multidisciplinary Introduction to Information Security

Jason J. Molitierno, Applications of Combinatorial Matrix Theory to Laplacian Matrices of Graphs

Richard A. Mollin, Advanced Number Theory with Applications

Richard A. Mollin, Algebraic Number Theory, Second Edition

Richard A. Mollin, Codes: The Guide to Secrecy from Ancient to Modern Times

Richard A. Mollin, Fundamental Number Theory with Applications, Second Edition

Richard A. Mollin, An Introduction to Cryptography, Second Edition

Titles (continued)

Richard A. Mollin, Quadratics

Richard A. Mollin, RSA and Public-Key Cryptography

Carlos J. Moreno and Samuel S. Wagstaff, Jr., Sums of Squares of Integers

Goutam Paul and Subhamoy Maitra, RC4 Stream Cipher and Its Variants

Dingyi Pei, Authentication Codes and Combinatorial Designs

Kenneth H. Rosen, Handbook of Discrete and Combinatorial Mathematics

Douglas R. Shier and K.T. Wallenius, Applied Mathematical Modeling: A Multidisciplinary Approach

Alexander Stanoyevitch, Introduction to Cryptography with Mathematical Foundations and Computer Implementations

Jörn Steuding, Diophantine Analysis

Douglas R. Stinson, Cryptography: Theory and Practice, Third Edition

Roberto Togneri and Christopher J. deSilva, Fundamentals of Information Theory and Coding Design

W. D. Wallis, Introduction to Combinatorial Designs, Second Edition

W. D. Wallis and J. C. George, Introduction to Combinatorics

Lawrence C. Washington, Elliptic Curves: Number Theory and Cryptography, Second Edition

DISCRETE MATHEMATICS AND ITS APPLICATIONS

Series Editor KENNETH H. ROSEN

APPLICATIONS OF COMBINATORIAL MATRIX THEORY TO LAPLACIAN MATRICES OF GRAPHS

Jason J. Molitierno

Sacred Heart University

Fairfield, Connecticut, USA

CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

The author would like to thank Kimberly Polauf for her assistance in designing the front cover.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20111229

International Standard Book Number-13: 978-1-4398-6339-8 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Dedication

This book is dedicated to my Ph.D. advisor, Dr. Michael "Miki" Neumann, who passed away unexpectedly as this book was nearing completion. In addition to teaching me the fundamentals of combinatorial matrix theory that made writing this book possible, Miki always provided much encouragement and emotional support throughout my time in graduate school and throughout my career. Miki not only treated me as an equal colleague, but also as family. I thank Miki Neumann for the person that he was and for the profound effect he had on my career and my life. Miki was a great advisor, mentor, colleague, and friend.

This page intentionally left blank

Contents

Preface

On the surface, matrix theory and graph theory are seemingly very different branches of mathematics. However, these two branches of mathematics interact since it is often convenient to represent a graph as a matrix. Adjacency, Laplacian, and incidence matrices are commonly used to represent graphs. In 1973, Fiedler [28] published his first paper on Laplacian matrices of graphs and showed how many properties of the Laplacian matrix, especially the eigenvalues, can give us useful information about the structure of the graph. Since then, many papers have been published on Laplacian matrices. This book is a compilation of many of the exciting results concerning Laplacian matrices that have been developed since the mid 1970s. Papers written by well-known mathematicians such as (alphabetically) Fallat, Fiedler, Grone, Kirkland, Merris, Mohar, Neumann, Shader, Sunder, and several others are consolidated here. Each theorem is referenced to its appropriate paper so that the reader can easily do more in-depth research on any topic of interest. However, the style of presentation in this book is not meant to be that of a journal but rather a reference textbook. Therefore, more examples and more detailed calculations are presented in this book than would be in a journal article. Additionally, most sections are followed by exercises to aid the reader in gaining a deeper understanding of the material. Some exercises are routine calculations that involve applying the theorems presented in the section. Other exercises require a more in-depth analysis of the theorems and require the reader to prove theorems that go beyond what was presented in the section. Many of these exercises are taken from relevant papers and they are referenced accordingly.

Only an undergraduate course in linear algebra and experience in proof writing are prerequisites for reading this book. To this end, Chapter 1 gives the necessities of matrix theory beyond that found in an undergraduate linear algebra course that are needed throughout this book. Topics such as matrix norms, mini-max principles, nonnegative matrices, M-matrices, doubly stochastic matrices, and generalized inverses are covered. While no prior knowledge of graph theory is required, it is helpful. Chapter 2 provides a basic overview of the necessary topics in graph theory that will be needed. Topics such as trees, special classes of graphs, connectivity, degree sequences, and the genus of graphs are covered in this chapter.

Once these basics are covered, we begin with a gentle approach to Laplacian matrices in which we motivate their study. This is done in Chapter 3. We begin with a brief study of other types of matrix representations of graphs, namely the adjacency and incidence matrices, and use these matrices to define the Laplacian matrix of a graph. Once the Laplacian matrix is defined, we present one of the most famous theorems in matrix-graph theory, the Matrix-Tree Theorem, which tells us the number of spanning trees in a given graph. Its proof is combinatoric in nature and the concepts in linear algebra that are employed are well within the grasp of a student who has a solid background in linear algebra. Chapter 3 continues to motivate the study of Laplacian matrices by deriving their construction from the continuous version of the Laplacian matrix which is used often in differential equations to study heat and energy flow through a region. We adopt these concepts to the study of energy flow on a graph. We further investigate these concepts at the end of Chapter 3 when we discuss networks which, historically, is the reason mathematicians began studying Laplacian matrices.

Once the motivation of studying Laplacian matrices is completed, we begin with a more rigorous study of their spectrum in Chapter 4. Since Laplacian matrices are symmetric, all eigenvalues are real numbers. Moreover, by the Gersgorin Disc Theorem, all of the eigenvalues are nonnegative. Since the row sums of a Laplacian matrix are all zero, it follows that zero is an eigenvalue since e, the vector of all ones, is an eigenvector corresponding to zero. We then explore the effects of the spectrum of the Laplacian matrix when taking the unions, joins, products, and complements of graphs. Once these results are established, we can then find upper bounds on the largest eigenvalue, and hence the entire spectrum, of the Laplacian matrix in terms of the structure of the graph. For example, an unweighted graph on n vertices cannot have an eigenvalue greater than n, and will have an eigenvalue of n if and only if the graph is the join of two graphs. Sharper upper bounds in terms of the number and the location of edges are also derived. Once we have upper bounds for the spectrum of the Laplacian matrix, we continue our study of its spectrum by illustrating the distribution of the eigenvalues less than, equal to, and greater than one. Additionally, the multiplicity of the eigenvalue $\lambda = 1$ gives us much insight into the number of pendant vertices of a graph. We then further our study of the spectrum by proving the recently proved Grone-Merris Conjecture which gives an upper bound on each eigenvalue of the Laplacian matrix of a graph. This is supplemented by the study of maximal or threshold graphs in which the Grone-Merris Conjecture is sharp for each eigenvalue. Such graphs have an interesting structure in that they are created by taking the successive joins and complements of complete graphs, empty graphs, and other maximal graphs. Moreover, since the upper bounds provided by the Grone-Merris Conjecture are integers, it becomes natural to study other graphs in which all eigenvalues of the Laplacian matrix are integers. In such graphs, the number of cycles comes into play.

In Chapter 5 we focus our study on the most important and most studied eigenvalue of the Laplacian matrix - the second smallest eigenvalue. This eigenvalue is known as the algebraic connectivity of a graph as it is used extensively to measure how connected a graph is. For example, the algebraic connectivity of a disconnected graph is always zero while the algebraic connectivity of a connected graph is always strictly positive. For a fixed n, the connected graph on n vertices with the largest algebraic connectivity is the complete graph as it is clearly the "most connected" graph. The path on n vertices is the connected graph on n vertices with the smallest algebraic connectivity since it is seen as the "least connected" graph. Also, the algebraic connectivity is bounded above by the vertex connectivity. Hence graphs with cut vertices such as trees will never have an algebraic connectivity greater than one. Overall, graphs containing more edges are likely to be "more connected" and hence will usually have larger algebraic connectivities. Adding an edge to a graph or increasing the weight of an existing edge will cause the algebraic connectivity to monotonically increase. Additionally, graphs with larger diameters tend to have fewer edges and thus usually have lower algebraic connectivities. The same holds true for planar graphs and graphs with low genus. In Chapter 5, we prove many theorems regarding the algebraic connectivity of a graph and how it relates to the structure of a graph.

Once we have studied the interesting ideas surrounding the algebraic connectivity of a graph, it is natural to want to study the eigenvector(s) corresponding to this eigenvalue. Such an eigenvector is known as the Fiedler vector. We dedicate Chapters 6 and 7 to the study of Fiedler vectors. Since the entries in a Fiedler vector correspond to the vertices of the graph, we begin our study of Fiedler vectors by illustrating how the entries of the Fiedler vector change as we travel along various paths in a graph. This leads us to classifying graphs into one of two types depending if there is a zero entry in the Fiedler vector corresponding to a cut vertex of the graph. We spend Chapter 6 focusing on trees since there is much literature concerning the Fiedler vectors of trees. Moreover, it is helpful to understand the ideas behind Fiedler vectors of trees before generalizing these results to graphs which is done in Chapter 7. When studying trees, we take the inverse of the submatrix of Laplacian matrix created by eliminating a row and column corresponding to a given vertex k of the tree. This matrix is known as the bottleneck matrix at vertex k. Bottleneck matrices give us much useful information about the tree. In an unweighted tree, the (i, j) entry of the bottleneck matrix is the number of edges that lie simultaneously on the path from i to k and on the path from j to k. An analogous result holds for weighted trees. Bottleneck matrices are also helpful in determining the algebraic connectivity of a tree as the spectral radius of bottleneck matrices and the algebraic connectivity are closely related. When generalizing these results to graphs, we gain much insight into the structure of a graph. We learn a great deal about its cut vertices, girth, and cycle structure.

Chapter 8 deals with the more modern aspects of Laplacian matrices. Since zero is an eigenvalue of the Laplacian matrix, it is singular, and hence we cannot take the inverse of such matrices. However, we can take the group generalized inverse of the Laplacian matrix and we discuss this in this chapter. Since the formula for the group inverse of the Laplacian matrix relies heavily on bottleneck matrices, we use many of the results of the previous two chapters to prove theorems concerning group inverses. We then apply these results to sharpen earlier results in this book. For example, we use the group inverse to create the Zenger function which is another upper bound on the algebraic connectivity. We also use the group inverse to investigate the rate of change of increase (the second derivative) in the algebraic connectivity when we increase the weight of an edge of a graph. The group inverse of the Laplacian matrix is interesting in its own right as its combinatorial properties give us much information about the stucture of a graph, especially trees. The distances between each pair of vertices in a tree is closely reflected in the entries of the group inverse. Moreover, within each row k of the group inverse, the entries in that row decrease as you travel along any path in the tree beginning at vertex k.

Matrix-graph theory is a fascinating subject that ties togtether two seemingly unrealted branches of mathematics. Because it makes use of both the combinatorial properties and the numerical properties of a matrix, this area of mathematics is fertile ground for research at the undergraduate, graduate, and experienced levels. I hope this book can serve as exploratory literature for the undergraduate student who is just learning how to do mathematical reasearch, a useful "start-up" book for the graduate student begining research in matrix-graph theory, and a convenient reference for the more experienced researcher.

Acknowledgments

The author would like to thank Dr. Stephen Kirkland and Dr. Michael Neumann for conversations that took place during the early stages of writing this book.

The author would also like to thank Sacred Heart University for providing support in the form of (i) a sabbatical leave, (ii) a University Research and Creativity Grant (URCG), and (iii) a College of Arts and Sciences release time grant.

This page intentionally left blank

Notation

$\Re$ - the set of real numbers

$\Re^n$ - the space of n-dimensional real-valued vectors

$A[X, Y]$ - the submatrix of A corresponding to the rows indexed by X and the columns indexed by Y

$A[X] = A[X, X]$

$[\overline{X}] = \{1, \ldots, n\} \setminus X$

$\|x\|$ - the Euclidean norm of the vector x

e - the column vector of all ones (the dimension is understood by the context)

$e^{(n)}$ - the n-dimensional column vector of all ones

e_i - the column vector with 1 in the i^{th} component and zeros elsewhere

y_i - the i^{th} component of the vector y

I - the identity matrix

J - the matrix of all ones

$E_{i,j}$ - the matrix with 1 in the (i, j) entry and zeros elsewhere

M_n - the set of all $n \times n$ matrices

$M_{m,n}$ - the set of all $m \times n$ matrices

$A \leq B$ - entries $a_{ij} \leq b_{ij}$ for all ordered pairs (i, j)

$A < B$ - entries $a_{ij} \leq b_{ij}$ for all ordered pairs (i, j) with strict inequality for at least one (i, j)

$A << B$ - entries $a_{ij} < b_{ij}$ for all ordered pairs (i, j).

A^T - the transpose of the matrix A

A^{-1} - the inverse of the matrix A

$A^{\#}$ - the group inverse of the matrix A

A^{+} - the Moore-Penrose inverse of the matrix A

$\mathrm{diag}(A)$ - the diagonal matrix consisting of the diagonal entries of A

$\det(A)$ - the determinant of the matrix A

$\mathrm{Tr}(A)$ - the trace of the matrix A

$m_A(\lambda)$ - the multiplicity of the eigenvalue λ of the matrix A

$L(\mathcal{G})$ - the Laplacian matrix of the graph $\mathcal{G}$

$m_{\mathcal{G}}(\lambda)$ - the multiplicity of the eigenvalue λ of $L(\mathcal{G})$

$\rho(A)$ - the spectral radius of the matrix A

$\lambda_k(A)$ - the k^{th} smallest eigenvalue of the matrix A. (Note that we will always use λ_n to denote the largest eigenvalue of the matrix A.)

$\sigma(A)$ - the spectrum of A, i.e., the set of eigenvalues of the matrix A counting multiplicity

$\sigma(\mathcal{G})$ - the set of eigenvalues, counting multiplicity, of $L(\mathcal{G})$

$\mathcal{Z}(A)$ - the Zenger of the matrix A

$|X|$ - the cardinality of a set X

$w(e)$ - the weight of the edge e

$|\mathcal{G}|$ - the number of vertices in the graph $\mathcal{G}$

d_v or $\deg(v)$ - the degree of vertex v

m_v - the average of the degrees of the vertices adjacent to v

$v \sim w$ - vertices v and w are adjacent

$N(v)$ - the set of vertices in $\mathcal{G}$ adjacent to the vertex v

$d(u, v)$ - the distance between vertices u and v

$\tilde{d}(u, v)$ - the inverse weighted distance between vertices u and v

$\tilde{d}_v$ - the inverse status of the vertex v

$\operatorname{diam}(\mathcal{G})$ - the diameter of the graph $\mathcal{G}$

$\overline{\rho}(\mathcal{G})$ - the mean distance of the graph $\mathcal{G}$

$V(\mathcal{G})$ - the vertex set of the graph $\mathcal{G}$

$E(\mathcal{G})$ - the edge set of the graph $\mathcal{G}$

$v(\mathcal{G})$ - the vertex connectivity of the graph $\mathcal{G}$

$e(\mathcal{G})$ - the edge connectivity of the graph $\mathcal{G}$

$a(\mathcal{G})$ - the algebraic connectivity of the graph $\mathcal{G}$

$\delta(\mathcal{G})$ - the minimum vertex degree of the graph $\mathcal{G}$

$\Delta(\mathcal{G})$ - the maximum vertex degree of the graph $\mathcal{G}$

$\gamma(\mathcal{G})$ - the genus of the graph $\mathcal{G}$

$p(\mathcal{G})$ - the number of pendant vertices of the graph $\mathcal{G}$

$q(\mathcal{G})$ - the number of quasipendant vertices of the graph $\mathcal{G}$

K_n - the complete graph on n vertices

$K_{m,m}$ - the complete bipartite graph whose partite sets contain m and n vertices, respectively

P_n - the path on n vertices

C_n - the cycle on n vertices

W_n - the wheel on $n+1$ vertices

$\mathcal{G}^c$ - the complement of the graph $\mathcal{G}$

$\mathcal{G}_1 + \mathcal{G}_2$ - the sum (union) of the graphs $\mathcal{G}_1$ and $\mathcal{G}_2$

$\mathcal{G}_1 \vee \mathcal{G}_2$ - the join of the graphs $\mathcal{G}_1$ and $\mathcal{G}_2$

$\mathcal{G}_1 \times \mathcal{G}_2$ - the product of the graphs $\mathcal{G}_1$ and $\mathcal{G}_2$

$\mathcal{L}(\mathcal{G})$ - the line graph of the graph $\mathcal{G}$

Chapter 1

Matrix Theory Preliminaries

As stated in the Preface, this book assumes an undergraduate knowledge of linear algebra. In this chapter, we study topics that are typically beyond that of an undergraduate linear algebra course, but are useful in later chapters of this book. Much of the material is taken from [6] and [41] which are two standard resources in linear algebra. We begin with a study of vector and matrix norms. Vector and matrix norms are useful in finding bounds on the spectral radius of a square matrix. We study the spectral radius of matrices more extensively in the next section which covers Perron-Frobenius theory. Perron-Frobenius theory is the study of nonnegative matrices. We will study nonnegative matrices in general, but also study interesting subsets of this class of matrices, namely positive matrices and irreducible matrices. We will see that positive matrices and irreducible matrices have many of the same properties. Nonnegative matrices will play an important role throughout this book and will be useful in understanding the theory behind M-matrices which also play an important role in later chapters. Hence we dedicate a section to M-matrices and apply the theory of nonnegative matrices to proofs of theorems involving M-matrices. Nonnegative matrices are also useful in the study of doubly stochastic matrices. Doubly stochastic matrices, which we study in the section following the section on M-matrices, are nonnegative matrices whose row sums and column sums are each one. Doubly stochastic matrices will play an important role in the study of the algebraic connectivity of graphs. Finally, we close this chapter with a section on generalized inverses of matrices. Since many of the matrices we will utilize in this book are singular, we need to familiarize ourselves with more general inverses, namely the group inverse of matrices.

1.1 Vector Norms, Matrix Norms, and the Spectral Radius of a Matrix

Vector and matrix norms have many uses in mathematics. In this section, we investigate vector and matrix norms and show how they give us insight into the spectral radius of a square matrix. To do this, we begin by understanding vector norms. In $\Re^n$, vectors are used to quantify length and distance. The length of a vector, or equivalently, the distance between two points in $\Re^n$, can be defined in many ways. However, for the sake of convenience, there are conditions that are often placed on the way such distances can be defined. This leads us to the formal definition of a vector norm:

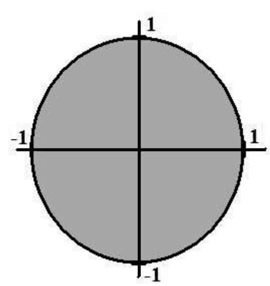
DEFINITION 1.1.1 In $\Re^n$, the function $\| \bullet \| : \Re^n \to \Re$ is a *vector norm* if for all vectors $x, y \in \Re^n$, it satisfies the following properties:

i) $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$
ii) $\|cx\| = |c|\|x\|$ for all scalars c
iii) $\|x + y\| \leq \|x\| + \|y\|$

EXAMPLE 1.1.2 The most commonly used norm is the *Euclidean norm*:

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$$

Given two vectors x and y whose initial point is at the origin, we often use the Euclidean norm to find the distance between the end points of these vectors. We do this by finding $\|x - y\|_2$. In other words, the Euclidean norm is often used to find the distance between two points in $\Re^n$. For example, the set of all points in $\Re^2$ whose Euclidean distance from the origin is at most 1 is the following:



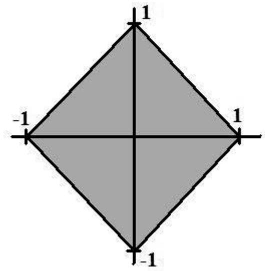
EXAMPLE 1.1.3 We can generalize the Euclidean norm to the ℓ_p norm for $p \geq 1$:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

OBSERVATION 1.1.4 The ℓ_1 norm is often referred to as the *sum norm* since:

$$\|x\|_1 = |x|_1 + |x|_2 + \ldots + |x|_n.$$

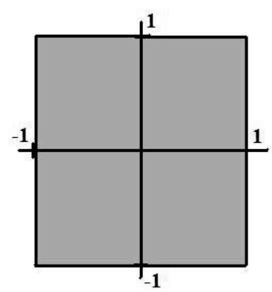
Since norms are often used to measure distance, we can compare the manners in which distance is defined between the norms ℓ_1 and ℓ_2. We saw above that the set of all points whose distance from the origin in $\Re^2$ is at most 1 with respect to ℓ_2 is the unit disc. However, the set of all points whose distance from the origin in $\Re^2$ is at most 1 with respect to ℓ_1 is the following:



OBSERVATION 1.1.5 The ℓ_∞ norm is often referred to as the *max norm* since:

$$\|x\|_\infty = \max\{|x|_1, |x|_2, \ldots, |x|_n\}.$$

Keeping with the concept of distance, the set of all points whose distance from the origin in $\Re^2$ is at most 1 with respect to ℓ_∞ is the following



Since norms are used to quantify distance in $\Re^n$, this leads us to the concept of a sequence of vectors converging. To this end, we have the following definition:

DEFINITION 1.1.6 Let $\{x^{(k)}\}$ be a sequence of vectors in $\Re^n$. We say that $\{x^{(k)}\}$ converges to the vector x with respect to the norm $\| \bullet \|$ if $\|x^{(k)} - x\| \to 0$ as $k \to \infty$.

With the idea of convergence, we are now able to compare various vector norms in $\Re^n$. We do this in the following theorem from [41]:

THEOREM 1.1.7 *Let $\| \bullet \|_\alpha$ and $\| \bullet \|_\beta$ be any two vector norms in $\Re^n$. Then there exist finite positive constants c_m and c_M such that $c_m\|x\|_\alpha \leq \|x\|_\beta \leq c_M\|x\|_\alpha$ for all $x \in \Re^n$.*

Proof: Define the function $h(x) = \|x\|_\beta / \|x\|_\alpha$ on the Euclidean unit ball $S = \{x \in \Re^n \mid \|x\|_2 = 1\}$ which is a compact set in $\Re^n$. Observe that the denominator of $h(x)$ is never zero on S by (i) of Definition 1.1.1. Since vector norms are continuous functions and since the denominaror of $h(x)$ is never zero on S, it follows that $h(x)$ is continuous on the compact set S. Hence by the Weierstrass theorem, h achieves a finite positive maximum c_M and a positive minimum c_m on S. Hence $c_m\|x\|_\alpha \leq \|x\|_\beta \leq c_M\|x\|_\alpha$ for all $x \in S$. Because $x/\|x\|_2 \in S$ for every nonzero vector $x \in \Re^n$, it follows that these inequalities hold for all nonzero $x \in \Re^n$.

These inequalities trivially hold for $x = 0$. This completes the proof. $\square$

Theorem 1.1.7 suggests that given a vector $x \in \Re^n$, the values of x with respect to various norms will not vary too much. This leads to the idea of equivalent norms.

DEFINITION 1.1.8 Two norms are *equivalent* if whenever a sequence of vectors $\{x^{(k)}\}$ converges to a vector x with respect to the first norm, then it converges to the same vector with respect to the second norm.

With this definition, we can now prove a corollary for Theorem 1.1.7 which is also from [41].

COROLLARY 1.1.9 *All vector norms in $\Re^n$ are equivalent.*

Proof: Let $\| \bullet \|_\alpha$ and $\| \bullet \|_\beta$ be vector norms in $\Re^n$. Let $\{x^{(k)}\}$ be a sequence of vectors that converges to a vector x with respect to $\| \bullet \|_\alpha$. By Theorem 1.1.7, there exist constants $c_M \geq c_m > 0$ such that

$$c_m \|x^{(k)} - x\|_\alpha \;\leq\; \|x^{(k)} - x\|_\beta \;\leq\; c_M \|x^{(k)} - x\|_\alpha$$

for all k. Therefore, it follows that $\|x^{(k)} - x\|_\alpha \to 0$ if and only if $\|x^{(k)} - x\|_\beta \to 0$ as $k \to \infty$. $\square$

The idea of equivalent norms will be useful as we turn our attention to matrix norms. We begin with a definition of a matrix norm. Observe that this definition is of similar flavor to that of a vector norm.

DEFINITION 1.1.10 Let M_n denote the set of all $n \times n$ matrices. The function $\| \bullet \| : M_n \to \Re$ is a *matrix norm* if for all $A, B \in M_n$, it satisfies the following properties:

i) $\|A\| \geq 0$ and $\|A\| = 0$ if and only if $A = 0$
ii) $\|cA\| = |c|\|A\|$ for all complex scalars c
iii) $\|A + B\| \leq \|A\| + \|B\|$
iv) $\|AB\| \leq \|A\|\|B\|$

Matrix norms are often defined in terms of vector norms. For example, a commonly used matrix norm is $\|A\|_p$ which is defined as

$$\|A\|_p \;=\; \max_{\|x\|_p \neq 0} \frac{\|Ax\|_p}{\|x\|_p} \;=\; \max_{\|x\|_p = 1} \|Ax\|_p.$$

As with vector norms, letting $p = 1$ and letting $p \to \infty$ are of interest. We now present the following observations from [41] concerning p-norms for matrices for important values of p:

OBSERVATION 1.1.11 For any $n \times n$ matrix A,

$$\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{n} |a_{i,j}|.$$

In other words, the 1-norm of a matrix is the maximum of the 1-norm of the column vectors of the matrix.

OBSERVATION 1.1.12 For any $n \times n$ matrix A,

$$\|A\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{i,j}|.$$

In other words, the ∞-norm of a matrix is the maximum of the 1-norm of the row vectors of the matrix.

Matrix norms are very useful in finding bounds on the eigenvalues of a square matrix. The following theorem from [41] shows that the spectral radius of a matrix is always bounded above by any norm of a matrix:

THEOREM 1.1.13 *If $\|\bullet\|$ is any matrix norm and if $A \in M_n$, then $\rho(A) \le \|A\|$.*

Proof: Let λ be an eigenvalue of A such that $|\lambda| = \rho(A)$. Let x be a corresponding eigenvector. Using the properties of matrix norms, we have

$$|\lambda|\|x\| = \|\lambda x\| = \|Ax\| \le \|A\|\|x\|.$$

Since $\|x\| > 0$, dividing through by $\|x\|$ gives us $\rho(A) = |\lambda| \le \|A\|$. □

We can use Observations 1.1.11 and 1.1.12 to obtain the following corollary from [41] which gives conditions as to when $\rho(A)$ and $\|A\|$ can be equal.

COROLLARY 1.1.14 *Let $A \in M_n$ and suppose that A is nonnegative. If the row sums of A are constant, then $\rho(A) = \|A\|_\infty$. If the column sums are constant, then $\rho(A) = \|A\|_1$.*

Proof: We know from Theorem 1.1.13 that $\rho(A) \le \|A\|$ for any matrix norm $\|\bullet\|$. However, if the row sums are constant, then e is an eigenvector of A with eigenvalue $\|A\|_\infty$, and so $\rho(A) = \|A\|_\infty$. The statement for column sums follows from applying the same argument to A^T. □

The goal for the remainder of this section is to prove a theorem which gives us a formula for the spectral radius in terms of matrix norms. To this end, we begin with an important lemma from [41].

LEMMA 1.1.15 *Let $A \in M_n$ and $\epsilon > 0$ be given. Then there is a matrix norm $\|\bullet\|$ such that $\rho(A) \le \|A\| \le \rho(A) + \epsilon$.*

Proof: By the Schur triangularization theorem (see [41]), there is a unitary matrix U and an upper triangular matrix V such that $A = U^TVU$. Let $D_t = \text{diag}(t, t^2, \ldots, t^n)$ and observe

$$D_tVD_t^{-1} = \begin{bmatrix} \lambda_1 & t^{-1}d_{12} & t^{-2}d_{13} & \ldots & t^{-n+1}d_{1n} \\ 0 & \lambda_2 & t^{-1}d_{23} & \ldots & t^{-n+2}d_{2n} \\ 0 & 0 & \lambda_3 & \ldots & t^{-n+3}d_{3n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & t^{-1}d_{n-1,n} \\ 0 & 0 & 0 & 0 & \lambda_n \end{bmatrix},$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. For $t > 0$ large enough, the sum of the off-diagonal entries of $D_tVD_t^{-1}$ are less that ϵ. In particular, by Observation 1.1.11 we have $\|D_tVD_t^{-1}\|_1 \leq \rho(A) + \epsilon$ for large enough t. Hence if we define the matrix norm $\| \bullet \|$ by

$$\|B\| = \|D_tUTBUD_t^{-1}\|_1 = \|(UD_t^{-1})^{-1}B(UD_t^{-1})\|_1$$

for any $B \in M_n$, and if we choose t large enough, we will have constructed a matrix norm such that $\|A\| \leq \rho(A) + \epsilon$. Since by Theorem 1.1.13, we have $\rho(A) \leq \|A\|$, this lemma is proven. $\square$

We now consider matrices whose norm is less than one for some norm. We do this with a lemma from [41].

LEMMA 1.1.16 *Let $A \in M_n$ be a given matrix. If there is a matrix norm $\| \bullet \|$ such that $\|A\| < 1$, then $\lim_{k\to\infty} A^k = 0$; that is, all the entries of A^k tend to zero as $k \to \infty$.*

Proof: If $\|A\| < 1$, then $\|A^k\| \leq \|A\|^k \to 0$ as $k \to \infty$. Thus $\|A^k\| \to 0$ as $k \to \infty$. But since all vector norms on the n^2-dimensional space M_n are equivalent by Corollary 1.1.9, it must also be the case that $\|A^k\|_\infty \to 0$. The result follows. $\square$

Intuitively, if $\lim_{k\to\infty} A^k = 0$, then the entries of A must be relatively small. Hence the spectral radius should be small. In the following lemma from [41], we make this idea more precise.

LEMMA 1.1.17 *Let $A \in M_n$. Then $\lim_{k\to\infty} A^k = 0$ if and only if $\rho(A) < 1$.*

Proof: If $A^k \to 0$ and if $x \neq 0$ is an eigenvector corresponding to the eigenvalue λ, then $A^kx = \lambda^kx \to 0$ if and only if $|\lambda| < 1$. Since this inequality must hold for every eigenvalue of A, we conclude that $\rho(A) < 1$. Conversely, if $\rho(A) < 1$, then by Lemma 1.1.15, there is some matrix norm $\| \bullet \|$ such that $\|A\| < 1$. Thus by Lemma 1.1.16, it follows that $A^k \to 0$ as $k \to \infty$. $\square$

We now prove the main result of this section which gives us a formula for the spectral radius of a matrix. This result is from [41].

THEOREM 1.1.18 *Let $A \in M_n$. For any matrix norm $\| \bullet \|$*

$$\rho(A) \;=\; \lim_{k\to\infty} \|A^k\|^{1/k}$$

Proof: Observe $\rho(A)^k = \rho(A^k) \leq \|A^k\|$, the last inequality follows from Theorem 1.1.13. Hence $\rho(A) \leq \|A^k\|^{1/k}$ for all natural numbers k. Given $\epsilon > 0$, the matrix $\hat{A} := [1/(\rho(A)+\epsilon)]A$ has a spectral radius strictly less than one and hence it follows from Lemma 1.1.17 that $\|\hat{A}^k\| \to 0$ as $k \to \infty$. Thus for a fixed A and ϵ, there exists N (depending on A and ϵ) such that $\|\hat{A}^k\| < 1$ for all $k \geq N$. But this is equivalent to saying $\|A^k\| \leq (\rho(A)+\epsilon)^k$ for all $k \geq N$, or that $\|A^k\|^{1/k} \leq \rho(A)+\epsilon$ for all $k \geq N$. Since ϵ was arbitrary, it follows that $\|A^k\|^{1/k} \leq \rho(A)$ for $k \geq N$. But we saw earlier in the proof that $\rho(A) \leq \|A^k\|^{1/k}$ for all k. Hence $\rho(A) \;=\; \lim_{k\to\infty} \|A^k\|^{1/k}$. □

Theorem 1.1.18 will be useful to us in later sections and chapters when we need to compare the spectral radii of matrices, especially nonnegative matrices. To this end, we close this section with three corollaries from [41] which allow us to compare the spectral radii of matrices. We prove the first corollary and leave the proofs of the remaining corollaries as exercises.

COROLLARY 1.1.19 *Let A and B be $n \times n$ matrices. If $|A| \leq B$, then $\rho(A) \leq \rho(|A|) \leq \rho(B)$.*

Proof: First note that for every natural number m we have $|A^m| \leq |A|^m \leq B^m$. Hence

$$\|A^m\|_2 \leq \||A|^m\|_2 \leq \|B^m\|_2$$

and

$$\|A^m\|_2^{1/m} \leq \||A|^m\|_2^{1/m} \leq \|B^m\|_2^{1/m}$$

for all natural numbers m. Letting m tend to infinity and applying Theorem 1.1.18 results in $\rho(A) \leq \rho(|A|) \leq \rho(B)$. □

COROLLARY 1.1.20 *Let A and B be $n \times n$ matrices. If $0 \leq A \leq B$, then $\rho(A) \leq \rho(B)$.*

COROLLARY 1.1.21 *Let A be an $n \times n$ matrix where $A \geq 0$. If $\tilde{A}$ is any principle submatrix of A, then $\rho(\tilde{A}) \leq \rho(A)$. In particular, $max_{1\leq i\leq n} a_{i,i} \leq \rho(A)$.*

Exercises:

1. (See [41]) Prove that for each $p \geq 1$ that ℓ_p is a vector norm by verifying the properties in Definition 1.1.1.

2. (See [41]) Prove that ℓ_∞ is a vector norm by verifying the properties in Definition 1.1.1, and show that

$$\|x\|_\infty \;=\; \lim_{p\to\infty} \|x\|_p.$$

3. Define the Frobenius norm for a matrix A as

$$\|A\|_F \;=\; \left(\sum_{i,j=1}^{n} |a_{ij}|^2\right)^{1/2}$$

Use Definition 1.1.10 to verify that this is a matrix norm.

4. Prove Corollary 1.1.20.

5. Prove Corollary 1.1.21.

1.2 Location of Eigenvalues

In this section, we develop theory that shows where the eigenvalues of a matrix lie and how the eigenvalues of a matrix change when the matrix is perturbed. Most of this section will focus on symmetric matrices since mainly symmetric matrices will be used throughout this book. We begin with a well-known theorem known as the Gersgorin Disc Theorem which states that all of the eigenvalues of a square matrix lie in certain discs on the complex plane.

THEOREM 1.2.1 The Gersgorin Disc Theorem. *Let A be an $n \times n$ matrix and let σ be the set of all eigenvalues of A. Then*

$$\sigma \subset \bigcup_{i=1}^{n} \left\{ r \in C \,:\, |a_{i.i} - r| \;\leq\; \sum_{\substack{k=1\\k\neq i}}^{n} |a_{i,k}| \right\} \tag{1.2.1}$$

Proof: Suppose λ is an eigenvalue of A with x as a corresponding eigenvector, i.e., $Ax = \lambda x$. Let x_i be the entry of x such that $x_i = \max_{1\leq k\leq n} |x_k|$. Observe

$$\sum_{k=1}^{n} a_{i,k}x_k \;=\; \lambda x_i$$

and therefore

$$(\lambda - a_{i,i})x_i \;=\; \sum_{\substack{k=1\\k\neq i}}^{n} a_{i,k}x_k.$$

By the triangle inequality we have

$$|\lambda - a_{i,i}|\,|x_i| \;\leq\; \sum_{\substack{k=1\\k\neq i}}^{n} |a_{i,k}|\,|x_k|\,.$$

Dividing through by $|x_i|$ and recalling that $x_i = \max_{1\leq k\leq n} |x_k|$, we obtain

$$|\lambda - a_{i,i}| \;\leq\; \sum_{\substack{k=1\\k\neq i}}^{n} |a_{i,k}|\,\frac{|x_k|}{|x_i|} \;\leq\; \sum_{\substack{k=1\\k\neq i}}^{n} |a_{i,k}|\,.$$

Therefore, the distance from $a_{i,i}$ to λ is at most $\sum_{\substack{k=1 \\ k\neq i}}^{n} |a_{i,k}|$ on the complex plane, i.e.,

$$\lambda \in \{r \in C : |a_{i,i} - r| \leq \sum_{\substack{k=1 \\ k\neq i}}^{n} |a_{i,k}|\}.$$

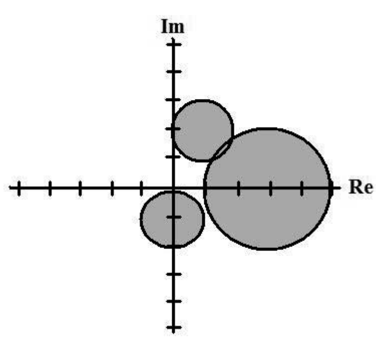
Taking all eigenvalues of A into account gives us (1.2.1). □

In summary, the Gersgorin Disc Theorem states that all of the eigenvalues of a square matrix lie in the union of discs whose centers are the diagonal entries of the matrix and whose radii are the sum of the absolute values of the off-diagonal entries in the corresponding row.

EXAMPLE 1.2.2 Consider the matrix

$$A = \begin{bmatrix} 1+2i & 0 & 1 \\ -1 & 3 & 1 \\ 0 & i & -i \end{bmatrix}$$

We create three discs in accordance with the Gersgorin Disc Theorem. The first disc has center $1+2i$ and radius 1; the second disc has center 3 and radius 2; the third disc has center $-i$ and radius 1. All eigenvalues of A will lie in the union of these discs.



Note that the eigenvalues of A are $3.1+0.2i$, $1.1+2.1i$, and $-0.2-1.3i$.

Since we will primarily deal with symmetric matrices in this book, we present a well-known theorem which shows that all eigenvalues of a symmetric matrix are real numbers.

THEOREM 1.2.3 *Let A be a real symmetric matrix. Then all eigenvalues of A are real.*

Proof: Let x^H and A^H denote the conjugate transpose of the vector x and matrix A, respectively. If λ is a complex number such that $\lambda = a + bi$ for real numbers a and b, note that $\lambda^H = a - bi$. We will prove this statement for the set of complex matrices A such that $A = A^H$ noting that the set of real symmetric matrices is a subset of this set. Let λ be an eigenvalue of A with corresponding eigenvector x normalized so that $x^H x = 1$. Then

$$\lambda \;=\; x^H A x \;=\; x^H A^H x \;=\; (x^H A x)^H \;=\; \lambda^H.$$

Since $\lambda = \lambda^H$, it follows that λ is real. □

Since all of eigenvalues of a symmetric matrix are real, we can order the eigenvalues as follows:

$$\lambda_{min} \;=\; \lambda_1 \;\leq\; \lambda_2 \;\leq\; \ldots \;\leq\; \lambda_{n-1} \;\leq\; \lambda_n \;=\; \lambda_{max}$$

Now that we know that all of the eigenvalues of a symmetric matrix are real and the approximate location of such eigenvalues via the Gersgorin Disc Theorem, we now proceed with the goal of this section which is to gain insight into the eigenvalues of symmetric matrices with respect to unit vectors. We begin by investigating the well-known Rayleigh-Ritz equations with a theorem found in [41] which give us useful formulas for the largest and smallest eigenvalues of a symmetric matrix in terms of unit vectors.

THEOREM 1.2.4 *Let $A \in M_n$ be symmetric. Then*

$$(i) \quad \lambda_1 x^T x \;\leq\; x^T A x \;\leq\; \lambda_n x^T x$$

for all $x \in \Re^n$. In addition

$$(ii) \quad \lambda_n \;=\; \max_{x \neq 0} \frac{x^T A x}{x^T x} \;=\; \max_{x^T x = 1} x^T A x$$

and

$$(iii) \quad \lambda_1 \;=\; \min_{x \neq 0} \frac{x^T A x}{x^T x} \;=\; \min_{x^T x = 1} x^T A x.$$

Proof: Since A is symmetric, there exists a unitary matrix $U \in M_n$ such that $A = UDU^T$ where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. For any vector $x \in \Re^n$, we have

$$x^T A x \;=\; x^T U D U^T x \;=\; (U^T x)^T D (U^T x) \;=\; \sum_{i=1}^{n} \lambda_i |(U^T x)_i|^2.$$

Since each term $|(U^T x)_i|^2$ is nonnegative, it follows that

$$\lambda_1 \sum_{i=1}^{n} |(U^T x)_i|^2 \;\leq\; x^T A x \;=\; \sum_{i=1}^{n} \lambda_i |(U^T x)_i|^2 \;\leq\; \lambda_n \sum_{i=1}^{n} |(U^T x)_i|^2.$$

Since U is unitary, it follows that

$$\sum_{i=1}^{n} |(U^T x)_i|^2 \;=\; \sum_{i=1}^{n} |x^T U U^T x| \;=\; \sum_{i=1}^{n} |x_i|^2 \;=\; x^T x.$$

Therefore

$$\lambda_1 x^T x \;\le\; x^T A x \;\le\; \lambda_n x^T x, \tag{1.2.2}$$

which proves (i).

To prove (ii), we see that dividing (1.2.2) through by $x^T x$ we obtain

$$\lambda_1 \;\le\; \frac{x^T A x}{x^T x} \;\le\; \lambda_n$$

However, if x is an eigenvector of A corresponding the eigenvalue λ_n, then

$$\frac{x^T A x}{x^T x} \;=\; \frac{\lambda_n x^T x}{x^T x} \;=\; \lambda_n$$

which implies

$$\max_{x \neq 0} \frac{x^T A x}{x^T x} \;=\; \lambda_n. \tag{1.2.3}$$

Finally, if $x \neq 0$ then

$$\frac{x^T A x}{x^T x} \;=\; \left(\frac{x}{\sqrt{x^T x}}\right)^T A \left(\frac{x}{\sqrt{x^T x}}\right) \quad \text{and} \quad \left(\frac{x}{\sqrt{x^T x}}\right)^T \left(\frac{x}{\sqrt{x^T x}}\right) \;=\; 1$$

which shows (1.2.3) is equivalent to

$$\max_{x^T x = 1} x^T A x \;=\; \lambda_n$$

This finishes the proof of (ii). The proof of (iii) is similar. □

Our goal will be to generalize the Rayleigh-Ritz equations to obtain formulas for the other eigenvalues of a symmetric matrix. This is known as the Courant-Fischer Minimax Principle. Before making such generalizations, we need a lemma from [41]:

LEMMA 1.2.5 *Let $A \in M_n$ and let $U = [u_1, \ldots, u_n]$ be a unitary matrix such that $A = U^T D U$ where $D = diag(\lambda_1, \ldots, \lambda_n)$. Then*

$$\max_{\substack{x \neq 0 \\ x \perp u_n, u_{n-1}, \ldots, u_{n-k+1}}} \frac{x^T A x}{x^T x} \;=\; \max_{\substack{x^T x = 1 \\ x \perp u_n, u_{n-1}, \ldots, u_{n-k+1}}} x^T A x \;=\; \lambda_{n-k}$$

where $u_1, \ldots, u_n$ are the columns of U.

__**Proof**__: Suppose we consider only those vectors $x \in \Re^n$ that are orthogonal to $u_n, u_{n-1}, \ldots, u_{n-k+1}$. Then

$$x^T A x \;=\; \sum_{i=1}^{n} \lambda_i |(U^T x)_i|^2 \;=\; \sum_{i=1}^{n} \lambda_i |u_i^T x|^2 \;=\; \sum_{i=1}^{n-k} \lambda_i |u_i^T x|^2.$$

This is a nonnegative linear combination of $\lambda_1, \ldots, \lambda_{n-k}$. Therefore

$$x^T A x \;=\; \sum_{i=1}^{n-k} \lambda_i |u_i^T x|^2 \le \lambda_{n-k} \sum_{i=1}^{n-k} |u_i^T x|^2 \;=\; \lambda_{n-k} \sum_{i=1}^{n} |(U^T x)_i|^2 \;=\; \lambda_{n-k} x^T x.$$

The inequality is sharp if $x = u_{n-k}$. The result now follows. □

REMARK 1.2.6 For each $k = 1, \ldots, n$, the column vector u_k of U is a unit eigenvector corresponding to the eigenvalue λ_k of A.

We are now ready to prove the main theorem of this section which generalizes the Rayleigh-Ritz equations. In this theorem from [41], we present the well-known Courant-Fischer Minimax Theorem.

THEOREM 1.2.7 *Let $A \in M_n$ be symmetric and let k be an integer $1 \le k \le n$. Then*

$$\lambda_k = \min_{w_1, w_2, \ldots, w_{n-k} \in \Re^n} \max_{\substack{x \neq 0 \\ x \in \Re^n \\ x \perp w_1, w_2, \ldots, w_{n-k}}} \frac{x^T A x}{x^T x} \tag{1.2.4}$$

and

$$\lambda_k = \max_{w_1, w_2, \ldots, w_{k-1} \in \Re^n} \min_{\substack{x \neq 0 \\ x \in \Re^n \\ x \perp w_1, w_2, \ldots, w_{k-1}}} \frac{x^T A x}{x^T x} \tag{1.2.5}$$

<u>Proof</u>: We will only prove (1.2.4) as the proof of (1.2.5) is similar. Writing $A = UDU^T$ as in the proof of Lemma 1.2.5 and fixing k where $2 \le k \le n$, then if $x \neq 0$, we have

$$\frac{x^T A x}{x^T x} = \frac{(U^T x)^T D (U^T x)}{x^T x} = \frac{(U^T x)^T D (U^T x)}{(U^T x)^T (U^T x).}$$

Since U is unitary, we have

$$\{U^T x : x \in \Re^n, x \neq 0\} = \{y \in \Re^n : y \neq 0\}.$$

Therefore, if $w_1, \ldots, w_{n-k} \in \Re^n$ are given, we have

$$\begin{aligned}
\sup_{\substack{x \neq 0 \\ x \perp w_1, \ldots, w_{n-k}}} \frac{x^T A x}{x^T x} &= \sup_{\substack{y \neq 0 \\ y \perp U^T w_1, \ldots, U^T w_{n-k}}} \frac{y^T D y}{y^T y} \\
&= \sup_{\substack{y^T y = 1 \\ y \perp U^T w_1, \ldots, U^T w_{n-k}}} \textstyle\sum_{i=1}^n \lambda_i |y_i|^2 \\
&\ge \sup_{\substack{y^T y = 1 \\ y \perp U^T w_1, \ldots, U^T w_{n-k} \\ y_1 = y_2 = \ldots = y_{k-1} = 0}} \textstyle\sum_{i=1}^n \lambda_i |y_i|^2 \\
&= \sup_{\substack{|y_k|^2 + |y_{k+1}|^2 + \ldots + |y_n|^2 = 1 \\ y \perp U^T w_1, \ldots, U^T w_{n-k}}} \textstyle\sum_{i=k}^n \lambda_i |y_i|^2 \\
&\ge \lambda_k.
\end{aligned}$$

Therefore

$$\sup_{\substack{x \neq 0 \\ x \perp w_1, \ldots, w_{n-k}}} \frac{x^T A x}{x^T x} \ge \lambda_k$$

for any $n-k$ vectors $w_1, \dots, w_{n-k}$. However, Lemma 1.2.5 and Remark 1.2.6 show that equality holds for one choice of the vectors w_i, namely $w_i = u_{n-i+1}$. Therefore

$$\inf_{w_1,\dots,w_{n-k}} \sup_{\substack{x \neq 0 \\ x \perp w_1,\dots,w_{n-k}}} \frac{x^T A x}{x^T x} = \lambda_k$$

Since the extrema is achieved in all of these cases, we replace "inf" and "sup" with "min" and "max," respectively. This completes the proof. □

One of the most important consequences of the Courant-Fisher Minimax Theorem are the interlacing theorems of eigenvalues. In the following theorem and corollaries from [41], we show that if we perturb a given symmetric matrix A to obtain a symmetric matrx B, then the eigenvalues of A and B interlace in some fashion. In the following theorem, we investigate the eigenvalues of the matrix $A + zz^T$ where A is symmetric and z is any real vector.

THEOREM 1.2.8 *Let $A \in M_n$ be symmetric and let $z \in \Re^n$ be a given vector. If the eigenvalues of A and $A + zz^T$ are arranged in increasing order, then*

$$(i)\ \lambda_k(A + zz^T) \;\leq\; \lambda_{k+1}(A) \;\leq\; \lambda_{k+2}(A + zz^T),\ \text{for}\, k = 1, 2, \dots, n-2$$

$$(ii)\ \lambda_k(A) \;\leq\; \lambda_{k+1}(A + zz^T) \;\leq\; \lambda_{k+2}(A),\ \text{for}\, k = 1, 2, \dots, n-2.$$

Proof: Let $1 \leq k \leq n-2$. Then by Theorem 1.2.7 we have

$$\begin{aligned}
\lambda_{k+2}(A \pm zz^T) &= \min_{w_1,\dots,w_{n-k-2}} \max_{\substack{x \neq 0 \\ x \perp w_1,\dots,w_{n-k-2}}} \frac{x^T(A+zz^T)x}{x^T x} \\
&\geq \min_{w_1,\dots,w_{n-k-2}} \max_{\substack{x \neq 0,\, x \perp z \\ x \perp w_1,\dots,w_{n-k-2}}} \frac{x^T(A+zz^T)x}{x^T x} \\
&= \min_{\substack{w_1,\dots,w_{n-k-2} \\ w_{n-k-1} = z}} \max_{\substack{x \neq 0 \\ x \perp w_1,\dots,w_{n-k-1}}} \frac{x^T(A+zz^T)x}{x^T x} \\
&\geq \min_{w_1,\dots,w_{n-k-1}} \max_{\substack{x \neq 0 \\ x \perp w_1,\dots,w_{n-k-1}}} \frac{x^T(A+zz^T)x}{x^T x} \\
&= \lambda_{k+1}(A).
\end{aligned}$$

Similarly, for $2 \leq k \leq n-1$ we have

$$\begin{aligned}
\lambda_k(A \pm zz^T) &= \max_{w_1,\ldots,w_{k-1}} \min_{\substack{x\neq 0\\ x\perp w_1,\ldots,w_{k-1}}} \frac{x^T(A+zz^T)x}{x^Tx} \\
&\leq \max_{w_1,\ldots,w_{k-1}} \min_{\substack{x\neq 0,\, x\perp z\\ x\perp w_1,\ldots,w_{k-1}}} \frac{x^T(A+zz^T)x}{x^Tx} \\
&= \max_{\substack{w_1,\ldots,w_{k-1}\\ w_k=z}} \min_{\substack{x\neq 0\\ x\perp w_1,\ldots,w_k}} \frac{x^T(A+zz^T)x}{x^Tx} \\
&\leq \max_{w_1,\ldots,w_k} \min_{\substack{x\neq 0\\ x\perp w_1,\ldots,w_k}} \frac{x^T(A+zz^T)x}{x^Tx} \\
&= \lambda_{k+1}(A).
\end{aligned}$$

Combining these inequalities proves the theorem. □

We close this section with three useful corollaries (see [41]) of Theorem 1.2.8 whose proofs we leave as exercises.

COROLLARY 1.2.9 *Let $A, B \in M_n$ be symmetric and suppose that B has rank at most r. Then*

$$(i)\ \lambda_k(A+B) \ \leq\ \lambda_{k+r}(A) \ \leq\ \lambda_{k+2r}(A+B), \text{ for } k = 1, 2, \ldots, n-2r$$

$$(ii)\ \lambda_k(A) \ \leq\ \lambda_{k+r}(A+B) \ \leq\ \lambda_{k+2r}(A), \text{ for } k = 1, 2, \ldots, n-2r.$$

COROLLARY 1.2.10 *Let $A \in M_n$ be symmetric, $z \in \Re^n$ be a vector, and $c \in \Re$. Let $\hat{A} \in M_{n+1}$ be the symmetric matrix obtained from A by bordering A with z and c as follows*

$$\hat{A} \ = \ \left[\begin{array}{c|c} A & z \\ \hline z^T & c \end{array}\right].$$

Then

$$\lambda_1(\hat{A}) \leq \lambda_1(A) \leq \lambda_2(\hat{A}) \leq \lambda_2(A) \leq \ldots \leq \lambda_{n-1}(A) \leq \lambda_n(\hat{A}) \leq \lambda_n(A) \leq \lambda_{n+1}(\hat{A}).$$

COROLLARY 1.2.11 *Let $A, B \in M_n$ be symmetric where B is positive semidefinite. Then*

$$\lambda_k(A) \ \leq\ \lambda_k(A+B)$$

for all $k = 1, \ldots, n$.

Exercises:

1. Prove Corollary 1.2.9.

2. Prove Corollary 1.2.10.

3. Prove Corollary 1.2.11.

1.3 Perron-Frobenius Theory

Perron-Frobenius theory deals with the eigenvalues and eigenvectors corresponding to the spectral radius of a nonnegative matrix. Nonnegative matrices are of great importance in matrix theory and will be of special importance later in this book as we apply them extensively in graph theory. Therefore, we dedicate a section to these results. We begin with a definition:

DEFINITION 1.3.1 A matrix A is *nonnegative* if all entries of A are nonnegative. In this case, we write $A \geq 0$. If all entires of A are strictly positive, then we say A is *positive* and write $A >> 0$.

Note that the set of positive matrices is a subset of the set of nonnegative matrices. Further if we want to denote that a nonnegative matrix A has at least one positive entry, we write $A > 0$.

In this section, we will first develop Perron-Frobenius theory for positive matrices. We then relax the condition of the matrices being positive and investigate how Perron-Frobenius theory changes when dealing with nonnnegative matrices. Finally, we study a special class of nonnegative matrices known as irreducible matrices and show that they behave similarly to positive matrices. We begin with the study of positive matrices. Since the set of positive matrices is a subset of nonnegative matrices, we begin with an important preliminary lemma and three useful corollaries from [41] concerning the larger class of nonnegative matrices:

LEMMA 1.3.2 *Let $A \in M_n$ be nonnegative. Then*

$$\min_{1\leq i\leq n} \sum_{j=1}^{n} a_{i,j} \ \leq \ \rho(A) \ \leq \ \max_{1\leq i\leq n} \sum_{j=1}^{n} a_{i,j} \tag{1.3.1}$$

and

$$\min_{1\leq j\leq n} \sum_{i=1}^{n} a_{i,j} \ \leq \ \rho(A) \ \leq \ \max_{1\leq j\leq n} \sum_{i=1}^{n} a_{i,j} \tag{1.3.2}$$

Proof: Let $\alpha = \min_{1\leq i\leq n} \sum_{j=1}^{n} a_{i,j}$ and let $B \in M_n$ be such that $b_{i,j} = \alpha a_{i,j} / \sum_{j=1}^{n} a_{i,j}$. Observe $A \geq B \geq 0$. By Corollary 1.1.14 we see that $\rho(B) = \alpha$; by Corollary 1.1.20 we have $\rho(B) \leq \rho(A)$. Hence $\alpha \leq \rho(A)$ which establishes the first inequality in (1.3.1). The second inequality in (1.3.1) is established in a similar fashion. Finally, (1.3.2) is established by applying the above argument to A^T. □

Now that we have some preliminary bounds on the spectral radius of nonnegative matrices, we can apply this lemma to get more precise results. In our first corollary, we recall that if S is an invertible matrix, then $\rho(S^{-1}AS) = \rho(A)$.

COROLLARY 1.3.3 *Let $A \in M_n$ be nonnegative. Then for any positive vector $x \in \Re^n$ we have*

$$\min_{1 \le i \le n} \frac{1}{x_i} \sum_{j=1}^{n} a_{i,j} x_j \;\le\; \rho(A) \;\le\; \max_{1 \le i \le n} \frac{1}{x_i} \sum_{j=1}^{n} a_{i,j} x_j$$

and

$$\min_{1 \le j \le n} x_j \sum_{i=1}^{n} \frac{a_{i,j}}{x_i} \;\le\; \rho(A) \;\le\; \max_{1 \le j \le n} x_j \sum_{i=1}^{n} \frac{a_{i,j}}{x_i}$$

Proof: Let $S = diag(x_1, \ldots, x_n)$. Since S is invertible, it follows that $\rho(S^{-1}AS) = \rho(A)$. Moreover, $S^{-1}AS$ is nonnegative. Thus we can apply Lemma 1.3.2 to $S^{-1}AS = [a_{i,j}x_j/x_i]$ to obtain the result. □

We now continue to sharpen our bounds on the spectral radius of nonnegative matrices found in Lemma 1.3.2 in the next corollary which helps us determine bounds on the spectral radius in terms of vectors. Observe that this corollary is somewhat reminiscent of Theorem 1.2.4(i).

COROLLARY 1.3.4 *Let $A \in M_n$ be nonnegative and suppose $x \in \Re^n$ is a positive vector. If $\alpha, \beta \ge 0$ are such that $\alpha x \le Ax \le \beta x$, then $\alpha \le \rho(A) \le \beta$. Moreover, if $\alpha x < Ax$ then $\alpha < \rho(A)$; if $Ax < \beta x$, then $\rho(A) < \beta$.*

Proof: If $\alpha x \le Ax$, then $\alpha \le \min_{1 \le i \le n}(1/x_i) \sum_{j=1}^{n} a_{i,j}x_j$. Thus by Corollary 1.3.3, it follows that $\alpha \le \rho(A)$. If $\alpha x < Ax$, then there exists some $\alpha' > \alpha$ such that $\alpha' \le Ax$. In this case, $\rho(A) \ge \alpha' > \alpha$, thus $\rho(A) > \alpha$. The upper bounds are verified in a similar fashion. □

The previous two corollaries have led up to the next corollary which will be useful when proving the first main result of this section.

COROLLARY 1.3.5 *Let $A \in M_n$ be nonnegative. If A has a positive eigenvector, then the corresponding eigenvalue is $\rho(A)$.*

Proof: Suppose $Ax = \lambda x$ where $x >> 0$. Then $\lambda x \le Ax \le \lambda x$ by Corollary 1.3.4. Applying Corollary 1.3.4 again, we obtain $\lambda \le \rho(A) \le \lambda$. □

Now that we have some preliminary results concerning nonnegative matrices, we return our focus to positive matrices. The first goal of this section is to prove Perron's theorem which is a well-known theorem concerning the eigenvalues and eigenvectors of positive matrices. First we need a lemma from [41].

LEMMA 1.3.6 *Let $A \in M_n$. Suppose that λ is an eigenvalue of A such that $|\lambda| = \rho(A)$ and that λ is the only eigenvalue of A with modulus $\rho(A)$. Suppose x and y are vectors such that $Ax = \lambda x$ and $A^T y = \lambda y$ where x and y are normalized so that $x^T y = 1$. Let $L = xy^T$. Then $\lim_{m\to\infty}[(1/\lambda(A))A]^m = L$.*

Proof: First, observe that (a) $L^m = L$ and (b) $A^m L = LA^m = \lambda^m L$ for all integers m. Then (a) and (b) imply $(A - \lambda L)^m = A^m - \lambda^m L$ for all integers m. Hence

$$\left(\frac{1}{\lambda}A - L\right)^m = \left[\frac{1}{\lambda}(A - \lambda L)\right]^m = \frac{1}{\lambda^m}A^m - L.$$

Therefore

$$\left(\frac{1}{\lambda}A\right)^m = L + \left(\frac{1}{\lambda}A - L\right)^m. \tag{1.3.3}$$

Since

$$\rho\left(\frac{1}{\lambda}A - L\right) = \frac{\rho(A - \lambda L)}{\rho(A)} \leq \frac{|\lambda_{n-1}(A)|}{\rho(A)} < 1,$$

the result follows from (1.3.3). □

OBSERVATION 1.3.7 Since L is the product of two vectors, it follows that the rank of L is 1.

We are now ready to prove Perron's Theorem for positive matrices which is the first main result of this section. The proof is adapted from [41].

THEOREM 1.3.8 *Let $A \in M_n$ be positive. Then*
(i) $\rho(A)$ is an eigenvalue of A,
(ii) There is a positive eigenvector corresponding to $\rho(A)$,
(iii) $|\lambda| < \rho(A)$ for every eigenvalue such that $\lambda \neq \rho(A)$,
(iv) $\rho(A)$ is a simple eigenvalue of A.

Proof: Let $x \neq 0$ be such that $Ax = \lambda x$ where $|\lambda| = \rho(A)$. Then

$$\rho(A)|x| = |\lambda||x| = |\lambda x| = |Ax| \leq |A||x| = A|x|.$$

Thus $y := A|x| - \rho(A)|x| \geq 0$. Since $|x| > 0$ and $A >> 0$, it follows that $z := A|x| >> 0$. If $y \neq 0$ then

$$0 < Ay = Az - \rho(A)z$$

which simplifies to $Az > \rho(A)z$. This implies that $\rho(A) > \rho(A)$ which is clearly false. Thus $y = 0$, and therefore $A|x| = \rho(A)|x|$. Hence $\rho(A)$ is a positive eigenvalue of A corresponding to the positive eigenvector $|x|$, thus (i) and (ii) are proved.

To prove (iii), we will show that if λ is an eigenvalue of A where $|\lambda| = \rho(A)$, then $\lambda = \rho(A)$. Let x be an eigenvector corresponding to λ. We first show that there exits an argument $0 \leq \theta < 2\pi$ such that $e^{-i\theta}x = |x| >> 0$. To see this, observe from (i) and (ii) that

$$\rho(A)|x_k| = |\lambda||x_k| = |\lambda x_k| = \left|\sum_{p=1}^{n} a_{kp}x_p\right| \leq \sum_{p=1}^{n} |a_{kp}||x_p| = \sum_{p=1}^{n} a_{kp}|x_p| = \rho(A)|x_k|.$$

Thus equality must hold in the triangle inequality and hence the nonzero complex numbers $a_{kp}x_p$, $p = 1, \ldots, n$ must all have the same argument, say θ. Since $a_{kp} > 0$

for all p, it follows that $e^{-i\theta}x >> 0$. Letting $w = e^{-i\theta}x >> 0$, we have $Aw = \lambda w$. But by Corollary 1.3.5 it follows that $\lambda = \rho(A)$.

To prove (iv), write $A = U\Delta U^T$ where U is unitary and Δ is an upper triangular matrix with main diagonal entries $\rho, \ldots, \rho, \lambda_{k+1}, \ldots, \lambda_n$, where $\rho = \rho(A)$ is an eigenvalue of A with algebraic multiplicity $k \geq 1$; the eigenvalues λ_i are all such that $|\lambda_i| < \rho(A)$ for all $k+1 \leq i \leq n$ (by part (iii)). Using Lemma 1.3.6 we have

$$L = \lim_{m\to\infty}\left(\frac{1}{\rho(A)}A\right)^m = U \lim_{m\to\infty}\begin{bmatrix} 1 & & & & & \\ & \ldots & & & * & \\ & & 1 & & & \\ & & & \frac{\lambda_{k+1}}{\rho} & & \\ & 0 & & & \ldots & \\ & & & & & \frac{\lambda_n}{\rho} \end{bmatrix}^m U^T$$

$$= U\begin{bmatrix} 1 & & & & & \\ & \ldots & & & * & \\ & & 1 & & & \\ & & & 0 & & \\ & 0 & & & \ldots & \\ & & & & & 0 \end{bmatrix} U^T$$

where the diagonal entry 1 is repeated k times in the last two expressions, and the diagonal entry 0 is repeated $n-k$ times. Since the upper triangular matrix in the last expression has rank at least k, and since L has rank 1 (Observation 1.3.7), we conclude that $k > 1$ is impossible, thus proving (iv). □

EXAMPLE 1.3.9 Consider the positive matrix

$$A = \begin{bmatrix} 8 & 8 & 8 \\ 4 & 2 & 1 \\ 4 & 12 & 4 \end{bmatrix}$$

The eigenvalues of A are 16, $-1 + 3.32i$, and $-1 - 3.32i$. Note that the eigenvalue of largest modulus is 16 and that $\rho(A) = 16$. Moreover, the eigenvector corresponding to 16 is positive, namely $[3, 1, 2]^T$. Finally, 16 is the only eigenvalue with a positive eigenvector as the eigenvectors corresponding to $-1+3.32i$ and $-1-3.32i$ are $[-0.52+1.23i, 1, -0.93-1.6i]^T$ and $[-0.52-1.23i, 1, -0.93+1.6i]^T$, respectively.

Since the eigenvector corresponding to the spectral radius of a positive matrix is of special importance, we have the following definition. Note in this definition, we relax the conditions of the matrix and eigenvector corresponding to the spectral radius to be nonnegative rather than positive. We will see in the theorem that follows that relaxing such conditions is desirable.

DEFINITION 1.3.10 A nonnegative eigenvector of $A \geq 0$ corresponding to $\rho(A)$ is called a *Perron vector* of A.

We now turn our attention to nonnegative matrices. Since we are relaxing the conditions of Theorem 1.3.8 by allowing our matrices to have entries of zero, we expect the conclusions of the theorem to be more relaxed in that the eigenvector corresponding to the spectral radius be allowed entries of zero. This is indeed the case as we see in the following theorem from [41].

THEOREM 1.3.11 *Let A be a nonnegative $n \times n$ matrix. Then*
(i) $\rho(A)$ is an eigenvalue of A, and
(ii) A has a nonnegative eigenvector corresponding to $\rho(A)$.

Proof: For any $\epsilon > 0$, define the matrix $A(\epsilon) := [a_{i,j} + \epsilon] >> 0$. Let $x(\epsilon)$ be the positive eigenvector of $A(\epsilon)$ corresponding to $\rho(A(\epsilon))$ as per Theorem 1.3.8(i). Normalize each vector $x(\epsilon)$ so that $\sum_{i=1}^{n} x(\epsilon)_i = 1$. Since the set of vectors $\{x(\epsilon) : \epsilon > 0\}$ is contained in the compact set $\{x : x \in C^n, \|x\|_1 \leq 1\}$, there is a monotone decreasing sequence $\epsilon_1, \epsilon_2, \ldots$, with $\lim_{k\to\infty} \epsilon_k = 0$ such that $x := \lim_{k\to\infty} x(\epsilon_k)$ exists. Since $x(\epsilon_k) >> 0$ for all k, it follows that $x \geq 0$. However, since

$$\sum_{i=1}^{n} x_i \;=\; \lim_{k\to\infty} \sum_{i=1}^{n} x(\epsilon_k)_i \;=\; 1$$

it follows that $x \neq 0$, hence $x > 0$. By Corollary 1.1.20 it follows that $\rho(A(\epsilon_k)) \geq \rho(A(\epsilon_{k+1})) \geq \ldots \geq \rho(A)$, for any k. Thus the sequence of real numbers $\{\rho(A(\epsilon_k))\}_{k=1,2,\ldots}$ is a bounded monotone decreasing sequence and hence $\rho := \lim_{k\to\infty} \rho(A(\epsilon_k))$ exists and $\rho \geq \rho(A)$. However,

$$\begin{aligned} Ax &= \lim_{k\to\infty} A(\epsilon_k)x(\epsilon_k) \\ &= \lim_{k\to\infty} \rho(A(\epsilon_k))x(\epsilon_k) \\ &= \lim_{k\to\infty} \rho(A(\epsilon_k)) \lim_{k\to\infty} x(\epsilon_k) \;=\; \rho x. \end{aligned}$$

Since $x \neq 0$, it follows that ρ is an eigenvalue of A with x as the corresponding eigenvector. Therefore $\rho = \rho(A)$ and $x > 0$ is a corresponding eigenvector. □

In Theorem 1.3.8 which concerns positive matrices, i.e., nonnegative matrices which do not contain a zero entry, we see that the eigenvector corresponding to the largest eigenvalue in modulus is also positive, hence it does not contain a zero entry. Moreover, the spectral radius of such a matrix is a simple eigenvalue. However, when we relax the conditions of allowing zero entries in a nonnegative matrix as we do in Theorem 1.3.11, we see that while the spectral radius is still an eigenvalue, it need not be a simple eigenvalue. Moreover, the eigenvector corresponding to such an eigenvalue is nonnegative, hence it may have a zero entry. We now turn our attention to a specific class of nonnegative matrices known as irreducible matrices. We will see that while these matrices may have a zero entry, they will behave like positive matrices. To this end, we have a definition:

DEFINITION 1.3.12 A matrix $A \in M_n$ is *reducible* if A is permutationally similar to a matrix of the form

$$\begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

where B and D are both square matrices. If A is not permutationally similar to a matrix of this form, we say that A is *irreducible*.

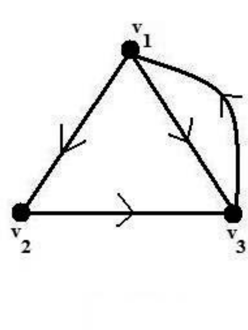
In order to be able to determine if a nonnegative matrix is irreducible, it is helpful for us to have a pictorial representation of the matrix:

DEFINITION 1.3.13 The associated *directed graph*, $G(A)$, of a matrix $A \in M_n$ is a graph on n vertices $v_1, \ldots, v_n$ where there is a directed edge from v_i to v_j if and only if $a_{i,j} \neq 0$.

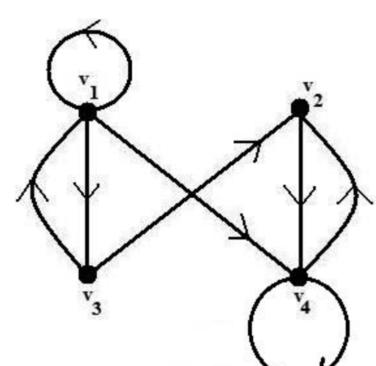
EXAMPLE 1.3.14 Consider the following matrices:

$$A = \begin{bmatrix} 0 & 5 & 3 \\ 0 & 0 & 1 \\ 4 & 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 0 & 4 & 1 \\ 0 & 0 & 0 & 3 \\ 6 & 5 & 0 & 0 \\ 0 & 7 & 0 & 2 \end{bmatrix}$$

Their associated directed graphs are



G(A)



G(B)

DEFINITION 1.3.15 A directed graph is *strongly connected* if for any pair of vertices v_i and v_j, it is possible to travel from v_i to v_j along a sequence of directed edges. We refer to such a sequence of directed edges as a *directed path*.

Observe that $G(A)$ is strongly connected. However, $G(B)$ is not strongly connected as there does not exist a directed path from v_4 to v_1 (or v_4 to v_3). The existence of directed paths between any pairs of vertices leads us to the following theorem from [6] concerning irreducible matrices:

THEOREM 1.3.16 *A matrix A is irreducible if and only if $G(A)$ is strongly connected.*

Proof: Suppose A is reducible, then there exists a permutation matrix P such that

$$A \;=\; P\begin{bmatrix} B & C \\ 0 & D \end{bmatrix}P^T \;:=\; P\hat{A}P^T$$

where $B \in M_r$, $D \in M_{n-r}$, $C \in M_{r,n-r}$, and $0 \in M_{n-r,r}$ for some $1 \le r \le n$. Let $v_1, \ldots, v_n \in V(G(A))$ and $\hat{v}_1, \ldots, \hat{v}_n \in V(G(\hat{A}))$. In $G(\hat{A})$, observe that there does not exist a directed path from $\hat{v}_i$ to $\hat{v}_j$ if $r+1 \le i \le n$ and $1 \le j \le r$. Hence $G(\hat{A})$ is not strongly connected. Since $G(A)$ and $G(\hat{A})$ are isomorphic, it follows that $G(A)$ is not strongly connected.

Now suppose $G(A)$ is not strongly connected. Then there exists nonempty sets of vertices S_1 and S_2 of $G(A)$ such that no directed path from v_i to v_j exists if $v_i \in S_2$ and $v_j \in S_1$. Let $|S_1| = r$ and $|S_2| = n - r$. Relabel the vertices of $G(A)$ as $\hat{v}_1, \ldots, \hat{v}_n$ where $\hat{v}_1, \ldots, \hat{v}_r \in S_1$ and $\hat{v}_{r+1}, \ldots, \hat{v}_n \in S_2$; permute the matrix A in the same fashion to create $\hat{A}$. Thus graph created from relabeling the vertices of $G(A)$ is precisely $G(\hat{A})$. Since there are no directed paths in $G(\hat{A})$ from the vertices in S_2 to the vertices in S_1, it follows that $\hat{A}$ must have an $(n-r)\times r$ block of zeros in the lower left corner. Thus $\hat{A}$ is reducible. Since A and $\hat{A}$ are permutationally similar, it follows that A is reducible. $\square$

EXAMPLE 1.3.17 Revisiting the nonnegative matrices in Example 1.3.14, we see that A is irreducible since the associated directed graph is strongly connected. However B is reducible since its associated directed graph is not strongly connected. Partioning the matrix below to highlight reducibility, observe that B is permutationally similar to

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

Our goal will be to show that nonnegative irreducible matrices behave in a similar way to positive matrices. To this end, we present two lemmas from [41] and [6] which shed light on the relationship between nonnegative and positive matrices.

LEMMA 1.3.18 *Let $A \in M_n$ and suppose A is nonnegative. Then A is irreducible if and only if $(I + A)^{n-1}$ is positive.*

Proof: Suppose first that A is reducible. Then for some permutation matrix P we have $A = P\hat{A}P^T$ where $\hat{A}$ is as in Theorem 1.3.16. Observe

$$(I+A)^{n-1} \;=\; (I+P\hat{A}P^T)^{n-1} \;=\; (P[I+\hat{A}]P^T)^{n-1}$$

$$= P\left[I + (n-1)\hat{A} + \binom{n-1}{2}\hat{A}^2 + \ldots + \binom{n-1}{n-1}\hat{A}^{n-1}\right]P^T.$$

By matrix multiplication, note that $\hat{A}^2, \hat{A}^3, \ldots, \hat{A}^{n-1}$ all have the same $(n-r)\times r$ block of 0's in the lower left corner as $\hat{A}$. Therefore, all of the terms in the square brackets have an $(n-r)\times r$ block of 0's in the lower left corner, and hence $(I+A)^{n-1}$ does also. Therefore $(I+A)^{n-1}$ is not positive.

Suppose now that A is irreducible, then so is $I+A$. Let Y be the set of all nonnegative nonzero vectors in $\Re^n$ with at least one entry of zero. Since $I+A$ is irreducible, it follows by matrix-vector multiplication that $(I+A)y$ will have fewer zero entries than y for each vector $y \in Y$. Hence $(I+A)^{n-1}y$ is positive for all vectors $y \in Y$. The only way that this can hold for *every* vector $y \in Y$ is for $(I+A)^{n-1}$ to be positive. $\square$

LEMMA 1.3.19 *If $A \in M_n$, A is nonnegative, and A^k is positive for some $k \geq 1$, then $\rho(A)$ is a simple eigenvalue of A.*

Proof: If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, then $\lambda_1^k, \ldots, \lambda_n^k$ are the eigenvalues of A^k. By Theorem 1.3.11, $\rho(A)$ is an eigenvalue of A. Hence if $\rho(A)$ were a multiple eigenvalue of A, then $\rho(A)^k = \rho(A^k)$ would be a multiple eigenvalue of A^k. But this is impossible since $\rho(A^k)$ is a simple eigenvalue of A^k by Theorem 1.3.8. $\square$

We are now able to present the culminating theorem (from [41]) of this section which illustrates that the majority of Theorem 1.3.8 still holds for nonnegative matrices so long as the matrix is irreducible.

THEOREM 1.3.20 *Let $A \in M_n$ be an irreducible nonnegative matrix. Then*
(i) $\rho(A) > 0$
(ii) $\rho(A)$ is an eigenvalue of A
(iii) There is a positive eigenvector x corresponding to $\rho(A)$
(iv) $\rho(A)$ is a simple eigenvalue.

Proof: Since A is nonnegative and irreducible, all row sums are positive. Thus (i) follows from Lemma 1.3.2. Statement (ii) follows from Theorem 1.3.11 and the fact that A is nonnegative (irreducibility is not even required here). Theorem 1.3.11 also guarantees that there exists a nonnegative eigenvector x corresponding to $\rho(A)$. To prove (iii), we need only show that x does not have a zero coordinate. Since $Ax = \rho(A)x$, it follows that $(I+A)^{n-1}x = (1+\rho(A))^{n-1}x$. By Lemma 1.3.18 the matrix $(I+A)^{n-1}$ is positive, so by Theorem 1.3.8, x is positive. To prove (iv), suppose $\rho(A)$ is a multiple eigenvalue of A. Then $1+\rho(A) = \rho(I+A)$ is a multiple eigenvalue of $I+A$. Hence $(1+\rho(A))^{n-1} = \rho((I+A)^{n-1})$ is a multiple eigenvalue of $(I+A)^{n-1}$. But $I+A$ is nonnegative and $(I+A)^{n-1}$ is positive by Lemma 1.3.18. Therefore by Lemma 1.3.19, $\rho((I+A)^{n-1})$ cannot be a multiple eigenvalue of $(I+A)^{n-1}$, producing a contradiction. Thus $\rho(A)$ is a simple eigenvalue of A. $\square$