Universitext

Fuzhen Zhang

Matrix Theory

Basic Results and Techniques

Second Edition

Springer

Universitext

Universitext

Universitext is a series of textbooks that presents material from a wide variety of mathematical disciplines at master's level and beyond. The books, often well class-tested by their author, may have an informal, personal even experimental approach to their subject matter. Some of the most successful and established books in the series have evolved through several editions, always following the evolution of teaching curricula, to very polished texts.

Thus as research topics trickle down into graduate-level teaching, first textbooks written for new, cutting-edge courses may make their way into *Universitext*.

For further volumes:
http://www.springer.com/series/223

Fuzhen Zhang

Matrix Theory

Basic Results and Techniques

Second Edition

Linear Park, Davie, Florida, USA

Springer

Fuzhen Zhang
Division of Math, Science, and Technology
Nova Southeastern University
Fort Lauderdale, FL 33314
USA
zhang@nova.edu

ISBN 978-1-4614-1098-0 e-ISBN 978-1-4614-1099-7
DOI 10.1007/978-1-4614-1099-7
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011935372

Mathematics Subject Classification (2010): 15-xx, 47-xx

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Cheng, Sunny, Andrew, and Alan

Preface to the Second Edition

The first edition of this book appeared a decade ago. This is a revised expanded version. My goal has remained the same: to provide a text for a second course in matrix theory and linear algebra accessible to advanced undergraduate and beginning graduate students. Through the course, students learn, practice, and master basic matrix results and techniques (or *matrix kung fu*) that are useful for applications in various fields such as mathematics, statistics, physics, computer science, and engineering, etc.

Major changes for the new edition are: eliminated errors, typos, and mistakes found in the first edition; expanded with topics such as matrix functions, nonnegative matrices, and (unitarily invariant) matrix norms; included more than 1000 exercise problems; rearranged some material from the previous version to form a new chapter, Chapter 4, which now contains numerical ranges and radii, matrix norms, and special operations such as the Kronecker and Hadamard products and compound matrices; and added a new chapter, Chapter 10, "Majorization and Matrix Inequalities", which presents a variety of inequalities on the eigenvalues and singular values of matrices and unitarily invariant norms.

I am thankful to many mathematicians who have sent me their comments on the first edition of the book or reviewed the manuscript of this edition: Liangjun Bai, Jane Day, Farid O. Farid, Takayuki Furuta, Geoffrey Goodson, Roger Horn, Zejun Huang, Minghua Lin, Dennis Merino, George P. H. Styan, Götz Trenkler, Qingwen Wang, Yimin Wei, Changqing Xu, Hu Yang, Xingzhi Zhan, Xiaodong Zhang, and Xiuping Zhang. I also thank Farquhar College of Arts and Sciences at Nova Southeastern University for providing released time for me to work on this project.

Readers are welcome to communicate with me via e-mail.

Fuzhen Zhang
Fort Lauderdale
May 23, 2011
zhang@nova.edu
www.nova.edu/~zhang

Preface

It has been my goal to write a concise book that contains fundamental ideas, results, and techniques in linear algebra and (mainly) in matrix theory which are accessible to general readers with an elementary linear algebra background. I hope this book serves the purpose.

Having been studied for more than a century, linear algebra is of central importance to all fields of mathematics. Matrix theory is widely used in a variety of areas including applied math, computer science, economics, engineering, operations research, statistics, and others.

Modern work in matrix theory is not confined to either linear or algebraic techniques. The subject has a great deal of interaction with combinatorics, group theory, graph theory, operator theory, and other mathematical disciplines. Matrix theory is still one of the richest branches of mathematics; some intriguing problems in the field were long standing, such as the Van der Waerden conjecture (1926–1980), and some, such as the permanental-dominance conjecture (since 1966), are still open.

This book contains eight chapters covering various topics from similarity and special types of matrices to Schur complements and matrix normality. Each chapter focuses on the results, techniques, and methods that are beautiful, interesting, and representative, followed by carefully selected problems. Many theorems are given different proofs. The material is treated primarily by matrix approaches and reflects the author's tastes.

The book can be used as a text or a supplement for a linear algebra or matrix theory class or seminar. A one-semester course may consist of the first four chapters plus any other chapter(s) or section(s). The only prerequisites are a decent background in elementary linear algebra and calculus (continuity, derivative, and compactness in a few places). The book can also serve as a reference for researchers and instructors.

The author has benefited from numerous books and journals, including *The American Mathematical Monthly, Linear Algebra and Its Applications, Linear and Multilinear Algebra*, and the International Linear Algebra Society (ILAS) Bulletin *Image*. This book would not exist without the earlier works of a great number of authors (see the References).

I am grateful to the following professors for many valuable suggestions and input and for carefully reading the manuscript so that many errors have been eliminated from the earlier version of the book:

Professor R. B. Bapat (Indian Statistical Institute),

Professor L. Elsner (University of Bielefeld),

Professor R. A. Horn (University of Utah),

Professor T.-G. Lei (National Natural Science Foundation of China),
Professor J.-S. Li (University of Science and Technology of China),
Professor R.-C. Li (University of Kentucky),
Professor Z.-S. Li (Georgia State University),
Professor D. Simon (Nova Southeastern University),
Professor G. P. H. Styan (McGill University),
Professor B.-Y. Wang (Beijing Normal University), and
Professor X.-P. Zhang (Beijing Normal University).

F. Zhang
Ft. Lauderdale
March 5, 1999
zhang@nova.edu
www.nova.edu/~zhang

Contents

Preface to the Second Edition vii
Preface ix
Frequently Used Notation and Terminology xv
Frequently Used Theorems xvii

1 Elementary Linear Algebra Review **1**

1.1 Vector Spaces 1
1.2 Matrices and Determinants 8
1.3 Linear Transformations and Eigenvalues 17
1.4 Inner Product Spaces 27

2 Partitioned Matrices, Rank, and Eigenvalues **35**

2.1 Elementary Operations of Partitioned Matrices 35
2.2 The Determinant and Inverse of Partitioned Matrices 42
2.3 The Rank of Product and Sum 51
2.4 The Eigenvalues of AB and BA 57
2.5 The Continuity Argument and Matrix Functions 62
2.6 Localization of Eigenvalues: The Geršgorin Theorem 67

3 Matrix Polynomials and Canonical Forms **73**

3.1 Commuting Matrices 73
3.2 Matrix Decompositions 79
3.3 Annihilating Polynomials of Matrices 87
3.4 Jordan Canonical Forms 93
3.5 The Matrices A^T, $\overline{A}$, A^*, A^TA, A^*A, and $\overline{A}A$ 102

4 Numerical Ranges, Matrix Norms, and Special Operations 107

4.1 Numerical Range and Radius 107
4.2 Matrix Norms 113
4.3 The Kronecker and Hadamard Products 117
4.4 Compound Matrices 122

5 Special Types of Matrices **125**

5.1 Idempotence, Nilpotence, Involution, and Projections 125
5.2 Tridiagonal Matrices 133
5.3 Circulant Matrices 138
5.4 Vandermonde Matrices 143
5.5 Hadamard Matrices 150
5.6 Permutation and Doubly Stochastic Matrices 155
5.7 Nonnegative Matrices 164

6 Unitary Matrices and Contractions **171**

6.1 Properties of Unitary Matrices 171
6.2 Real Orthogonal Matrices 177
6.3 Metric Space and Contractions 182
6.4 Contractions and Unitary Matrices 188
6.5 The Unitary Similarity of Real Matrices 192
6.6 A Trace Inequality of Unitary Matrices 195

7 Positive Semidefinite Matrices **199**

7.1 Positive Semidefinite Matrices 199
7.2 A Pair of Positive Semidefinite Matrices 207
7.3 Partitioned Positive Semidefinite Matrices 217
7.4 Schur Complements and Determinant Inequalities 227
7.5 The Kronecker and Hadamard Products of Positive Semidefinite Matrices 234
7.6 Schur Complements and the Hadamard Product 240
7.7 The Wielandt and Kantorovich Inequalities 245

8 Hermitian Matrices **253**

8.1 Hermitian Matrices and Their Inertias 253
8.2 The Product of Hermitian Matrices 260
8.3 The Min-Max Theorem and Interlacing Theorem 266
8.4 Eigenvalue and Singular Value Inequalities 274
8.5 Eigenvalues of Hermitian matrices A, B, and $A+B$ 281
8.6 A Triangle Inequality for the Matrix $(A^*A)^{1/2}$ 287

9 Normal Matrices **293**

9.1 Equivalent Conditions 293
9.2 Normal Matrices with Zero and One Entries 306
9.3 Normality and Cauchy–Schwarz–Type Inequality 312
9.4 Normal Matrix Perturbation 319

10 Majorization and Matrix Inequalities **325**

10.1 Basic Properties of Majorization 325
10.2 Majorization and Stochastic Matrices 334
10.3 Majorization and Convex Functions 340
10.4 Majorization of Diagonal Entries, Eigenvalues, and Singular Values 349
10.5 Majorization for Matrix Sum 356
10.6 Majorization for Matrix Product 363
10.7 Majorization and Unitarily Invariant Norms 372

References 379
Notation 391
Index 395

Frequently Used Notation and Terminology

$\dim V$, 3	dimension of vector space V
$\mathbb{M}_n$, 8	$n \times n$ (i.e., n-square) matrices with complex entries
$A = (a_{ij})$, 8	matrix A with (i,j)-entry a_{ij}
I, 9	identity matrix
A^T, 9	transpose of matrix A
$\overline{A}$, 9	conjugate of matrix A
A^*, 9	conjugate transpose of matrix A, i.e., $A^* = \overline{A}^T$
A^{-1}, 13	inverse of matrix A
$\operatorname{rank}(A)$, 11	rank of matrix A
$\operatorname{tr} A$, 21	trace of matrix A
$\det A$, 12	determinant of matrix A
$\lvert A \rvert$, 12, 83, 164	determinant for a block matrix A or $(A^*A)^{1/2}$ or $(\lvert a_{ij} \rvert)$
(u, v), 27	inner product of vectors u and v
$\lVert \cdot \rVert$, 28, 113	norm of a vector or a matrix
$\operatorname{Ker}(A)$, 17	kernel or null space of A, i.e., $\operatorname{Ker}(A) = \{x : Ax = 0\}$
$\operatorname{Im}(A)$, 17	image space of A, i.e., $\operatorname{Im}(A) = \{Ax\}$
$\rho(A)$, 109	spectral radius of matrix A
$\sigma_{\max}(A)$, 109	largest singular value (spectral norm) of matrix A
$\lambda_{\max}(A)$, 124	largest eigenvalue of matrix A
$A \geq 0$, 81	A is positive semidefinite (or all $a_{ij} \geq 0$ in Section 5.7)
$A \geq B$, 81	$A - B$ is positive semidefinite (or $a_{ij} \geq b_{ij}$ in Section 5.7)
$A \circ B$, 117	Hadamard (entrywise) product of matrices A and B
$A \otimes B$, 117	Kronecker (tensor) product of matrices A and B
$x \prec_w y$, 326	weak majorization, i.e., all $\sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k y_i^\downarrow$ hold
$x \prec_{\mathrm{wlog}} y$, 344	weak log-majorization, i.e., all $\prod_{i=1}^k x_i^\downarrow \leq \prod_{i=1}^k y_i^\downarrow$ hold

An $n \times n$ matrix A is said to be

upper-triangular	if all entries below the main diagonal are zero
diagonalizable	if $P^{-1}AP$ is diagonal for some invertible matrix P
similar to B	if $P^{-1}AP = B$ for some invertible matrix P
unitarily similar to B	if $U^*AU = B$ for some unitary matrix U
unitary	if $AA^* = A^*A = I$, i.e., $A^{-1} = A^*$
positive semidefinite	if $x^*Ax \geq 0$ for all vectors $x \in \mathbb{C}^n$
Hermitian	if $A = A^*$
normal	if $A^*A = AA^*$

$\lambda \in \mathbb{C}$ is an eigenvalue of $A \in \mathbb{M}_n$ if $Ax = \lambda x$ for some nonzero $x \in \mathbb{C}^n$.

Frequently Used Theorems

- **Cauchy–Schwarz inequality:** Let V be an inner product space over a number field ($\mathbb{R}$ or $\mathbb{C}$). Then for all vectors x and y in V

$$|(x,y)|^2 \le (x,x)(y,y).$$

  Equality holds if and only if x and y are linearly dependent.

- **Theorem on the eigenvalues of AB and BA:** Let A and B be $m \times n$ and $n \times m$ complex matrices, respectively. Then AB and BA have the same nonzero eigenvalues, counting multiplicity. As a consequence,

$$\operatorname{tr}(AB) = \operatorname{tr}(BA).$$

- **Schur triangularization theorem:** For any n-square matrix A, there exists an n-square unitary matrix U such that U^*AU is upper-triangular.

- **Jordan decomposition theorem:** For any n-square matrix A, there exists an n-square invertible complex matrix P such that

$$A = P^{-1}(J_1 \oplus J_2 \oplus \cdots \oplus J_k)P,$$

  where each J_i, $i = 1, 2, \ldots, k$, is a Jordan block.

- **Spectral decomposition theorem:** Let A be an n-square normal matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then there exists an n-square unitary matrix U such that

$$A = U^* \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)U.$$

  In particular, if A is positive semidefinite, then all $\lambda_i \ge 0$; if A is Hermitian, then all λ_i are real; and if A is unitary, then all $|\lambda_i| = 1$.

- **Singular value decomposition theorem:** Let A be an $m \times n$ complex matrix with rank r. Then there exist an m-square unitary matrix U and an n-square unitary matrix V such that

$$A = UDV,$$

  where D is the $m \times n$ matrix with (i,i)-entries being the singular values of A, $i = 1, 2, \ldots, r$, and other entries 0. If $m = n$, then D is diagonal.

CHAPTER 1

Elementary Linear Algebra Review

Introduction: We briefly review, mostly without proof, the basic concepts and results taught in an elementary linear algebra course. The subjects are vector spaces, basis and dimension, linear transformations and their eigenvalues, and inner product spaces.

1.1 Vector Spaces

Let V be a set of objects (elements) and $\mathbb{F}$ be a field, mostly the real number field $\mathbb{R}$ or the complex number field $\mathbb{C}$ throughout this book. The set V is called a *vector space* over $\mathbb{F}$ if the operations *addition*

$$u + v, \quad u, v \in V,$$

and *scalar multiplication*

$$cv, \quad c \in \mathbb{F}, v \in V,$$

are defined so that the addition is associative, is commutative, has an additive identity 0 and additive inverse $-v$ in V for each $v \in V$, and so that the scalar multiplication is distributive, is associative, and has an identity $1 \in \mathbb{F}$ for which $1v = v$ for every $v \in V$.

F. Zhang, *Matrix Theory: Basic Results and Techniques*, Universitext,
DOI 10.1007/978-1-4614-1099-7_1,

Put these in symbols:

1. $u + v \in V$ for all $u, v \in V$.
2. $cv \in V$ for all $c \in \mathbb{F}$ and $v \in V$.
3. $u + v = v + u$ for all $u, v \in V$.
4. $(u + v) + w = u + (v + w)$ for all $u, v, w \in V$.
5. There is an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$.
6. For each $v \in V$ there is an element $-v \in V$ so that $v+(-v) = 0$.
7. $c(u + v) = cu + cv$ for all $c \in \mathbb{F}$ and $u, v \in V$.
8. $(a + b)v = av + bv$ for all $a, b \in \mathbb{F}$ and $v \in V$.
9. $(ab)v = a(bv)$ for all $a, b \in \mathbb{F}$ and $v \in V$.
10. $1v = v$ for all $v \in V$.

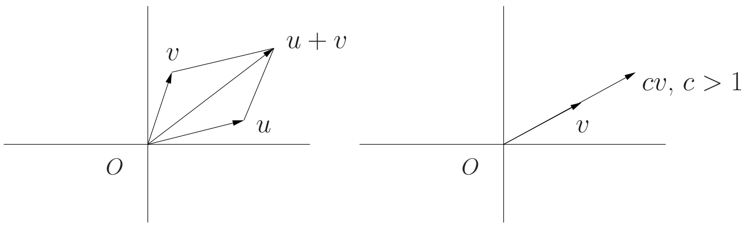

Figure 1.1: Vector addition and scalar multiplication

We call the elements of a vector space *vectors* and the elements of the field *scalars*. For instance, $\mathbb{R}^n$, the set of real column vectors

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \text{also written as} \quad (x_1, x_2, \ldots, x_n)^T$$

(T for transpose) is a vector space over $\mathbb{R}$ with respect to the addition

$$(x_1, x_2, \ldots, x_n)^T + (y_1, y_2, \ldots, y_n)^T = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)^T$$

and the scalar multiplication

$$c\,(x_1, x_2, \ldots, x_n)^T = (cx_1, cx_2, \ldots, cx_n)^T, \quad c \in \mathbb{R}.$$

Note that the real row vectors also form a vector space over $\mathbb{R}$; and they are essentially the same as the column vectors as far as vector spaces are concerned. For convenience, we may also consider $\mathbb{R}^n$ as a row vector space if no confusion is caused. However, in the matrix-vector product Ax, obviously x needs to be a column vector.

Let S be a nonempty subset of a vector space V over a field $\mathbb{F}$. Denote by $\operatorname{Span} S$ the collection of all finite linear combinations of the vectors in S; that is, $\operatorname{Span} S$ consists of all vectors of the form

$$c_1v_1 + c_2v_2 + \cdots + c_tv_t, \;\; t = 1, 2, \ldots, c_i \in \mathbb{F}, \; v_i \in S,$$

The set $\operatorname{Span} S$ is also a vector space over $\mathbb{F}$. If $\operatorname{Span} S = V$, then every vector in V can be expressed as a linear combination of vectors in S. In such cases we say that the set S *spans* the vector space V.

A set $S = \{v_1, v_2, \ldots, v_k\}$ is said to be *linearly independent* if

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k = 0$$

holds only when $c_1 = c_2 = \cdots = c_k = 0$. If there are also nontrivial solutions, i.e., not all c are zero, then S is *linearly dependent.*

For example, both $\{(1, 0), (0, 1), (1, 1)\}$ and $\{(1, 0), (0, 1)\}$ span $\mathbb{R}^2$. The first set is linearly dependent; the second one is linearly independent. The vectors $(1, 0)$ and $(1, 1)$ also span $\mathbb{R}^2$.

A *basis* of a vector space V is a linearly independent set that spans V. If V possesses a basis of an n-vector set $S = \{v_1, v_2, \ldots, v_n\}$, we say that V is of *dimension* n, written as $\dim V = n$. Conventionally, if $V = \{0\}$, we write $\dim V = 0$. If any finite set cannot span V, then V is infinite-dimensional and we write $\dim V = \infty$. Unless otherwise stated, we assume throughout the book that the vector spaces are finite-dimensional, as we mostly deal with finite matrices, even though some results hold for infinite-dimensional spaces.

For instance, $\mathbb{C}$ is a vector space of dimension 2 over $\mathbb{R}$ with basis $\{1, i\}$, where $i = \sqrt{-1}$, and of dimension 1 over $\mathbb{C}$ with basis $\{1\}$.

$\mathbb{C}^n$, the set of row (or column) vectors of n complex components, is a vector space over $\mathbb{C}$ having *standard basis*

$$e_1 = (1, 0, \ldots, 0, 0), \; e_2 = (0, 1, \ldots, 0, 0), \; \ldots, \; e_n = (0, 0, \ldots, 0, 1).$$

If $\{u_1, u_2, \ldots, u_n\}$ is a basis for a vector space V of dimension n, then every x in V can be uniquely expressed as a linear combination of the basis vectors:

$$x = x_1u_1 + x_2u_2 + \cdots + x_nu_n,$$

where the x_i are scalars. The n-tuple $(x_1, x_2, \ldots, x_n)$ is called the *coordinate* of vector x with respect to the basis.

Let V be a vector space of dimension n, and let $\{v_1, v_2, \ldots, v_k\}$ be a linearly independent subset of V. Then $k \leq n$, and it is not difficult to see that if $k < n$, then there exists a vector $v_{k+1} \in V$ such that the set $\{v_1, v_2, \ldots, v_k, v_{k+1}\}$ is linearly independent (Problem 16). It follows that the set $\{v_1, v_2, \ldots, v_k\}$ can be extended to a basis of V.

Let W be a subset of a vector space V. If W is also a vector space under the addition and scalar multiplication for V, then W is called a *subspace* of V. One may check (Problem 9) that W is a subspace if and only if W is closed under the addition and scalar multiplication.

For subspaces V_1 and V_2, the *sum* of V_1 and V_2 is defined to be

$$V_1 + V_2 = \{v_1 + v_2 : v_1 \in V_1,\ v_2 \in V_2\}.$$

It follows that the sum $V_1 + V_2$ is also a subspace. In addition, the intersection $V_1 \cap V_2$ is a subspace, and

$$V_1 \cap V_2 \subseteq V_i \subseteq V_1 + V_2, \quad i = 1, 2.$$

The sum $V_1 + V_2$ is called a *direct sum*, symbolized by $V_1 \oplus V_2$, if

$$v_1 + v_2 = 0, \quad v_1 \in V_1,\ v_2 \in V_2 \quad \Rightarrow \quad v_1 = v_2 = 0.$$

One checks that in the case of a direct sum, every vector in $V_1 \oplus V_2$ is uniquely written as a sum of a vector in V_1 and a vector in V_2.

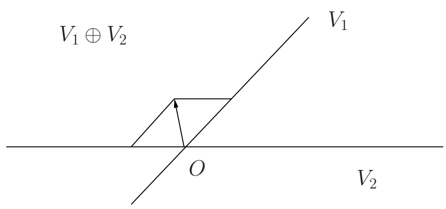

Figure 1.2: Direct sum

Theorem 1.1 (Dimension Identity) *Let V be a finite-dimensional vector space, and let V_1 and V_2 be subspaces of V. Then*

$$\dim V_1 + \dim V_2 = \dim(V_1 + V_2) + \dim(V_1 \cap V_2).$$

The proof is done by first choosing a basis $\{u_1, \ldots, u_k\}$ for $V_1 \cap V_2$, extending it to a basis $\{u_1, \ldots, u_k, v_{k+1}, \ldots, v_s\}$ for V_1 and a basis $\{u_1, \ldots, u_k, w_{k+1}, \ldots, w_t\}$ for V_2, and then showing that

$$\{u_1, \ldots, u_k, v_{k+1}, \ldots, v_s, w_{k+1}, \ldots, w_t\}$$

is a basis for $V_1 + V_2$.

It follows that subspaces V_1 and V_2 contain nonzero common vectors if the sum of their dimensions exceeds $\dim V$.

Problems

1. Show explicitly that $\mathbb{R}^2$ is a vector space over $\mathbb{R}$. Consider $\mathbb{R}^2$ over $\mathbb{C}$ with the usual addition. Define $c(x, y) = (cx, cy)$, $c \in \mathbb{C}$. Is $\mathbb{R}^2$ a vector space over $\mathbb{C}$? What if the "scalar multiplication" is defined as
$$c(x, y) = (ax + by, ax - by), \text{ where } c = a + bi, \; a, b \in \mathbb{R}?$$

2. Can a vector space have two different additive identities? Why?

3. Show that $\mathbb{F}_n[x]$, the collection of polynomials over a field $\mathbb{F}$ with degree at most n, is a vector space over $\mathbb{F}$ with respect to the ordinary addition and scalar multiplication of polynomials. Is $\mathbb{F}[x]$, the set of polynomials with any finite degree, a vector space over $\mathbb{F}$? What is the dimension of $\mathbb{F}_n[x]$ or $\mathbb{F}[x]$?

4. Determine whether the vectors $v_1 = 1 + x - 2x^2$, $v_2 = 2 + 5x - x^2$, and $v_3 = x + x^2$ in $\mathbb{F}_2[x]$ are linearly independent.

5. Show that $\{(1, i), \; (i, -1)\}$ is a linearly independent subset of $\mathbb{C}^2$ over the real $\mathbb{R}$ but not over the complex $\mathbb{C}$.

6. Determine whether $\mathbb{R}^2$, with the operations
$$(x_1, y_1) + (x_2, y_2) = (x_1 x_2, y_1 y_2)$$
and
$$c(x_1, y_1) = (cx_1, cy_1),$$
is a vector space over $\mathbb{R}$.

7. Let V be the set of all real numbers in the form

$$a + b\sqrt{2} + c\sqrt{5},$$

where a, b, and c are rational numbers. Show that V is a vector space over the rational number field $\mathbb{Q}$. Find $\dim V$ and a basis of V.

8. Let V be a vector space. If $u, v, w \in V$ are such that $au + bv + cw = 0$ for some scalars a, b, c, $ac \neq 0$, show that $\operatorname{Span}\{u, v\} = \operatorname{Span}\{v, w\}$.

9. Let V be a vector space over $\mathbb{F}$ and let W be a subset of V. Show that W is a subspace of V if and only if for all $u, v \in W$ and $c \in \mathbb{F}$

$$u + v \in W \quad \text{and} \quad cu \in W.$$

10. Is the set $\{(x,\, y) \in \mathbb{R}^2 : 2x - 3y = 0\}$ a subspace of $\mathbb{R}^2$? How about $\{(x,\, y) \in \mathbb{R}^2 : 2x - 3y = 1\}$? Give a geometric explanation.

11. Show that the set $\{(x,\, y - x,\, y) : x, y \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^3$. Find the dimension and a basis of the subspace.

12. Find a basis for $\operatorname{Span}\{u, v, w\}$, where $u = (1, 1, 0)$, $v = (1, 3, -1)$, and $w = (1, -1, 1)$. Find the coordinate of $(1, 2, 3)$ under the basis.

13. Let $W = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : x_3 = x_1 + x_2 \text{ and } x_4 = x_1 - x_2\}$.

    (a) Prove that W is a subspace of $\mathbb{R}^4$.

    (b) Find a basis for W. What is the dimension of W?

    (c) Prove that $\{c(1, 0, 1, 1) : c \in \mathbb{R}\}$ is a subspace of W.

    (d) Is $\{c(1, 0, 0, 0) : c \in \mathbb{R}\}$ a subspace of W?

14. Show that each of the following is a vector space over $\mathbb{R}$.

    (a) $C[a, b]$, the set of all (real-valued) continuous functions on $[a, b]$.

    (b) $C'(\mathbb{R})$, the set of all functions of continuous derivatives on $\mathbb{R}$.

    (c) The set of all even functions.

    (d) The set of all odd functions.

    (e) The set of all functions f that satisfy $f(0) = 0$.

    [Note: Unless otherwise stated, functions are added and multiplied by scalars in a usual way, i.e., $(f+g)(x) = f(x)+g(x)$, $(kf)(x) = kf(x)$.]

15. Show that if W is a subspace of vector space V of dimension n, then $\dim W \leq n$. Is it possible that $\dim W = n$ for a proper subspace W?

16. Let $\{u_1, \ldots, u_s\}$ and $\{v_1, \ldots, v_t\}$ be two sets of vectors. If $s > t$ and each u_i can be expressed as a linear combination of $v_1, \ldots, v_t$, show that $u_1, \ldots, u_s$ are linearly dependent.

17. Let V be a vector space over a field $\mathbb{F}$. Show that $cv = 0$, where $c \in \mathbb{F}$ and $v \in V$, if and only if $c = 0$ or $v = 0$. [Note: The scalar 0 and the vector 0 are usually different. For simplicity, here we use 0 for both. In general, one can easily tell from the text which is which.]

18. Let V_1 and V_2 be subspaces of a finite-dimensional space. Show that the sum $V_1 + V_2$ is a direct sum if and only if

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2.$$

Conclude that if $\{u_1, \ldots, u_s\}$ is a basis for V_1 and $\{v_1, \ldots, v_t\}$ is a basis for V_2, then $\{u_1, \ldots, u_s, v_1, \ldots, v_t\}$ is a basis for $V_1 \oplus V_2$.

19. If V_1, V_2, and W are subspaces of a finite-dimensional vector space V such that $V_1 \oplus W = V_2 \oplus W$, is it always true that $V_1 = V_2$?

20. Let V be a vector space of finite dimension over a field $\mathbb{F}$. If V_1 and V_2 are two subspaces of V such that $\dim V_1 = \dim V_2$, show that there exists a subspace W such that $V = V_1 \oplus W = V_2 \oplus W$.

21. Let V_1 and V_2 be subspaces of a vector space of finite dimension such that $\dim(V_1 + V_2) = \dim(V_1 \cap V_2) + 1$. Show that $V_1 \subseteq V_2$ or $V_2 \subseteq V_1$.

22. Let S_1, S_2, and S_3 be subspaces of a vector space of dimension n. Show that

$$(S_1 + S_2) \cap (S_1 + S_3) = S_1 + (S_1 + S_2) \cap S_3.$$

23. Let S_1, S_2, and S_3 be subspaces of a vector space of dimension n. Show that

$$\dim(S_1 \cap S_2 \cap S_3) \geq \dim S_1 + \dim S_2 + \dim S_3 - 2n.$$

1.2 Matrices and Determinants

An $m \times n$ *matrix* A over a field $\mathbb{F}$ is a rectangular array of m rows and n columns of entries in $\mathbb{F}$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Such a matrix, written as $A = (a_{ij})$, is said to be of *size* (or *order*) $m \times n$. Two matrices are considered to be equal if they have the same size and same corresponding entries in all positions.

The set of all $m \times n$ matrices over a field $\mathbb{F}$ is a vector space with respect to *matrix addition* by adding corresponding entries and to *scalar multiplication* by multiplying each entry of the matrix by the scalar. The dimension of the space is mn, and the matrices with one entry equal to 1 and 0 entries elsewhere form a basis. In the case of square matrices; that is, $m = n$, the dimension is n^2.

We denote by $\mathbb{M}_{m\times n}(\mathbb{F})$ the set of all $m \times n$ matrices over the field $\mathbb{F}$, and throughout the book we simply write $\mathbb{M}_n$ for the set of all complex n-square (i.e., $n \times n$) matrices.

The *product* AB of two matrices $A = (a_{ij})$ and $B = (b_{ij})$ is defined to be the matrix whose (i, j)-entry is given by

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

Thus, in order that AB make sense, the number of columns of A must be equal to the number of rows of B. Take, for example,

$$A = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & 4 & 5 \\ 6 & 0 & 8 \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} -3 & 4 & -3 \\ 12 & 0 & 16 \end{pmatrix}.$$

Note that BA is undefined.

Sometimes it is useful to write the matrix product AB, with $B = (b_1, b_2, \ldots, b_n)$, where the b_i are the column vectors of B, as

$$AB = (Ab_1, Ab_2, \ldots, Ab_n).$$

The *transpose* of an $m \times n$ matrix $A = (a_{ij})$ is an $n \times m$ matrix, denoted by A^T, whose (i,j)-entry is a_{ji}; and the *conjugate* of A is a matrix of the same size as A, symbolized by $\overline{A}$, whose (i,j)-entry is $\overline{a_{ij}}$. We denote the conjugate transpose $\overline{A}^T$ of A by A^*.

The $n \times n$ *identity* matrix I_n, or simply I, is the n-square matrix with all diagonal entries 1 and off-diagonal entries 0. A *scalar* matrix is a multiple of I, and a *zero* matrix 0 is a matrix with all entries 0. Note that two zero matrices may not be the same, as they may have different sizes. A square complex matrix $A = (a_{ij})$ is said to be

diagonal	if $a_{ij} = 0,\ i \neq j$,
upper-triangular	if $a_{ij} = 0,\ i > j$,
symmetric	if $A^T = A$,
Hermitian	if $A^* = A$,
normal	if $A^*A = AA^*$,
unitary	if $A^*A = AA^* = I$, and
orthogonal	if $A^TA = AA^T = I$.

A *submatrix* of a given matrix is an array lying in specified subsets of the rows and columns of the given matrix. For example,

$$C = \begin{pmatrix} 1 & 2 \\ 3 & \frac{1}{4} \end{pmatrix}$$

is a submatrix of

$$A = \begin{pmatrix} 0 & 1 & 2 \\ i & 3 & \frac{1}{4} \\ \pi & \sqrt{3} & -1 \end{pmatrix}$$

lying in rows one and two and columns two and three.

If we write $B = (0, i)$, $D = (\pi)$, and $E = (\sqrt{3}, -1)$, then

$$A = \begin{pmatrix} B & C \\ D & E \end{pmatrix}.$$

The right-hand side matrix is called a *partitioned* or *block* form of A, or we say that A is a partitioned (or block) matrix.

The manipulation of partitioned matrices is a basic technique in matrix theory. One can perform addition and multiplication of (appropriately) partitioned matrices as with ordinary matrices.

For instance, if A, B, C, X, Y, U, V are n-square matrices, then

$$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \begin{pmatrix} X & Y \\ U & V \end{pmatrix} = \begin{pmatrix} AX + BU & AY + BV \\ CU & CV \end{pmatrix}.$$

The block matrices of order 2×2 have appeared to be the most useful partitioned matrices. We primarily emphasize the techniques for block matrices of this kind in this book.

Elementary row operations for matrices are those that

i. Interchange two rows.

ii. Multiply a row by a nonzero constant.

iii. Add a multiple of a row to another row.

Elementary column operations are similarly defined, and similar operations on partitioned matrices are discussed in Section 2.1.

An n-square matrix is called an *elementary matrix* if it can be obtained from I_n by a single elementary row operation. Elementary operations can be represented by elementary matrices. Let E be the elementary matrix by performing an elementary row (or column) operation on I_m (or I_n for column). If the same elementary row (or column) operation is performed on an $m \times n$ matrix A, then the resulting matrix from A via the elementary row (or column) operation is given by the product EA (or AE, respectively).

For instance, by elementary row and column operations, the 2×3 matrix $A = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{smallmatrix}\right)$ is brought into $\left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right)$. Write in equations:

$$R_3 R_2 R_1 \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} C_1 C_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

where

$$R_1 = \begin{pmatrix} 1 & 0 \\ -4 & 1 \end{pmatrix}, \quad R_2 = \begin{pmatrix} 1 & 0 \\ 0 & -\frac{1}{3} \end{pmatrix}, \quad R_3 = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$$

and

$$C_1 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}.$$

This generalizes to the so-called *rank decomposition* as follows.

Theorem 1.2 *Let A be an $m \times n$ matrix over a field $\mathbb{F}$. Then there exist an $m \times m$ matrix P and an $n \times n$ matrix Q, both of which are products of elementary matrices with entries from $\mathbb{F}$, such that*

$$PAQ = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}. \tag{1.1}$$

The partitioned matrix in (1.1), written as $I_r \oplus 0$ and called a *direct sum* of I_r and 0, is uniquely determined by A. The size r of the identity I_r is the *rank* of A, denoted by $\operatorname{rank}(A)$. If $A = 0$, then $\operatorname{rank}(A) = 0$. Clearly $\operatorname{rank}(A^T) = \operatorname{rank}(\overline{A}) = \operatorname{rank}(A^*) = \operatorname{rank}(A)$.

An application of this theorem reveals the dimension of the *solution space* or *null space* of the linear equation system $Ax = 0$.

Theorem 1.3 *Let A be an $m \times n$ (real or complex) matrix of rank r. Let $\operatorname{Ker} A$ be the null space of A, i.e., $\operatorname{Ker} A = \{x : Ax = 0\}$. Then*

$$\dim \operatorname{Ker} A = n - r.$$

A notable fact about a linear equation system is that

$$Ax = 0 \quad \text{if and only if} \quad (A^*A)x = 0.$$

The *determinant* of a square matrix A, denoted by $\det A$, or $|A|$ as preferred if A is in a partitioned form, is a number associated with A. It can be defined in several different but equivalent ways. The one in terms of permutations is concise and sometimes convenient.

We say a permutation p on $\{1, 2, \ldots, n\}$ is *even* if p can be restored to natural order by an even number of interchanges. Otherwise, p is *odd*. For instance, consider the permutations on $\{1, 2, 3, 4\}$. The permutation $p = (2, 1, 4, 3)$; that is, $p(1) = 2$, $p(2) = 1$, $p(3) = 4$, $p(4) = 3$, is even because it will become $(1, 2, 3, 4)$ after interchanging 2 and 1 and 4 and 3 (two interchanges), whereas $(1, 4, 3, 2)$ is odd, for interchanging 4 and 2 gives (1, 2, 3, 4).

Let S_n be the set of all $(n!)$ permutations on $\{1, 2, \ldots, n\}$. For $p \in S_n$, define $\operatorname{sign}(p) = 1$ if p is even and $\operatorname{sign}(p) = -1$ if p is odd. Then the determinant of an n-square matrix $A = (a_{ij})$ is given by

$$\det A = \sum_{p \in S_n} \operatorname{sign}(p) \prod_{t=1}^{n} a_{tp(t)}.$$

Simply put, the determinant is the sum of all $(n!)$ possible "signed" products in which each product involves n entries of A belonging to different rows and columns. For $n = 2$, $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, $\det A = ad - bc$.

The determinant can be calculated by the *Laplace formula*

$$\det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A(1|j),$$

where $A(1|j)$ is a submatrix of A obtained by deleting row 1 and column j of A. This formula is referred to as the *Laplace expansion formula* along row 1. One can also expand a determinant along other rows or columns to get the same result. The determinant of a matrix has the following properties.

a. The determinant changes sign if two rows are interchanged.

b. The determinant is unchanged if a constant multiple of one row is added to another row.

c. The determinant is a linear function of any row when all the other rows are held fixed.

Similar properties are true for columns. Two often-used facts are

$$\det(AB) = \det A \det B, \quad A, B \in \mathbb{M}_n,$$

and

$$\begin{vmatrix} A & B \\ 0 & C \end{vmatrix} = \det A \det C, \quad A \in \mathbb{M}_n, \ C \in \mathbb{M}_m.$$

A square matrix A is said to be *invertible* or *nonsingular* if there exists a matrix B of the same size such that

$$AB = BA = I.$$

Such a matrix B, which can be proven to be unique, is called the *inverse* of A and denoted by A^{-1}. The inverse of A, when it exists, can be obtained from the *adjoint* of A, written as $\operatorname{adj}(A)$, whose (i, j)-entry is the *cofactor* of a_{ji}, that is, $(-1)^{j+i} \det A(j|i)$. In symbols,

$$A^{-1} = \frac{1}{\det A} \operatorname{adj}(A). \tag{1.2}$$

An effective way to find the inverse of a matrix A is to apply elementary row operations to the matrix (A, I) to get a matrix in the form (I, B). Then $B = A^{-1}$ (Problem 23).

If A is a square matrix, then $AB = I$ if and only if $BA = I$.

It is easy to see that $\operatorname{rank}(A) = \operatorname{rank}(PAQ)$ for invertible matrices P and Q of appropriate sizes (meaning that the involved operations for matrices can be performed). It can also be shown that the rank of A is the largest number of linearly independent columns (rows) of A. In addition, the rank of A is r if and only if there exists an r-square submatrix of A with nonzero determinant, but all $(r+1)$-square submatrices of A have determinant zero (Problem 24).

Theorem 1.4 *The following statements are equivalent for $A \in \mathbb{M}_n$.*

1. *A is invertible, i.e., $AB = BA = I$ for some $B \in \mathbb{M}_n$.*
2. *$AB = I$ (or $BA = I$) for some $B \in \mathbb{M}_n$.*
3. *A is of rank n.*
4. *A is a product of elementary matrices.*
5. *$Ax = 0$ has only the trivial solution $x = 0$.*
6. *$Ax = b$ has a unique solution for each $b \in \mathbb{C}^n$.*
7. $\det A \neq 0$.
8. *The column vectors of A are linearly independent.*
9. *The row vectors of A are linearly independent.*

Problems

1. Find the rank of $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 2 & 1 & 0 \end{pmatrix}$ by performing elementary operations.

2. Evaluate the determinants
$$\begin{vmatrix} 2 & -3 & 10 \\ 1 & 2 & -2 \\ 0 & 1 & -3 \end{vmatrix}, \quad \begin{vmatrix} 1+x & 1 & 1 \\ 1 & 1+y & 1 \\ 1 & 1 & 1+z \end{vmatrix}.$$

3. Show the 3×3 *Vandermonde* determinant identity
$$\begin{vmatrix} 1 & 1 & 1 \\ a_1 & a_2 & a_3 \\ a_1^2 & a_2^2 & a_3^2 \end{vmatrix} = (a_2 - a_1)(a_3 - a_1)(a_3 - a_2)$$
and evaluate the determinant
$$\begin{vmatrix} 1 & a & a^2 - bc \\ 1 & b & b^2 - ca \\ 1 & c & c^2 - ab \end{vmatrix}.$$

4. Let A be an n-square matrix and k be a scalar. Show that
$$\det(kA) = k^n \det A.$$

5. If A is a Hermitian (complex) matrix, show that $\det A$ is real.

6. If A an $n \times n$ real matrix, where n is odd, show that $A^2 \neq -I$.

7. Let $A \in \mathbb{M}_n$. Show that $A^* + A$ is Hermitian and $A^* - A$ is normal.

8. Let A and B be complex matrices of appropriate sizes. Show that
   (a) $\overline{AB} = \overline{A}\,\overline{B}$,
   (b) $(AB)^T = B^T A^T$,
   (c) $(AB)^* = B^* A^*$, and
   (d) $(AB)^{-1} = B^{-1} A^{-1}$ if A and B are invertible.

9. Show that matrices $\left(\begin{smallmatrix} 1 & i \\ i & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} i & i \\ i & -1 \end{smallmatrix}\right)$ are both symmetric, but one is normal and the other one is not normal.

10. Find the inverse of each of the following matrices.
$$\begin{pmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & b & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

11. If a, b, c, and d are complex numbers such that $ad - bc \neq 0$, show that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

12. Compute for every positive integer k,

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^k, \quad \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^k, \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^k, \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}^k.$$

13. Show that for any square matrices A and B of the same size,

$$A^*A - B^*B = \frac{1}{2}\big((A+B)^*(A-B) + (A-B)^*(A+B)\big).$$

14. If $AB = A + B$ for matrices A, B, show that A and B commute, i.e.,

$$AB = A + B \quad \Rightarrow \quad AB = BA.$$

15. Let A and B be n-square matrices such that $AB = BA$. Show that

$$(A+B)^k = A^k + kA^{k-1}B + \tfrac{k(k-1)}{2}A^{k-2}B^2 + \cdots + B^k.$$

16. Let A be a square complex matrix. Show that

$$I - A^{m+1} = (I - A)(I + A + A^2 + \cdots + A^m).$$

17. Let A, B, C, and D be n-square complex matrices. Compute

$$\begin{pmatrix} A & A^* \\ A^* & A \end{pmatrix}^2 \quad \text{and} \quad \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} D & -B \\ -C & A \end{pmatrix}.$$

18. Determine whether each of the following statements is true.
    - (a) The sum of Hermitian matrices is Hermitian.
    - (b) The product of Hermitian matrices is Hermitian.
    - (c) The sum of unitary matrices is unitary.
    - (d) The product of unitary matrices is unitary.
    - (e) The sum of normal matrices is normal.
    - (f) The product of normal matrices is normal.

19. Show that the solution set to the linear system $Ax = 0$ is a vector space of dimension $n - \operatorname{rank}(A)$ for any $m \times n$ matrix A over $\mathbb{R}$ or $\mathbb{C}$.

20. Let $A,\ B \in \mathbb{M}_n$. If $AB = 0$, show that $\operatorname{rank}(A) + \operatorname{rank}(B) \leq n$.

21. Let A and B be complex matrices with the same number of columns. If $Bx = 0$ whenever $Ax = 0$, show that

$$\operatorname{rank}(B) \leq \operatorname{rank}(A), \quad \operatorname{rank}\begin{pmatrix} A \\ B \end{pmatrix} = \operatorname{rank}(A),$$

and that $B = CA$ for some matrix C. When is C invertible?

22. Show that any two of the following three properties imply the third:

$$\text{(a)}\ \ A = A^*; \qquad \text{(b)}\ \ A^* = A^{-1}; \qquad \text{(c)}\ \ A^2 = I.$$

23. Let $A, B \in \mathbb{M}_n$. If $B(A, I) = (I, B)$, show that $B = A^{-1}$. Explain why A^{-1}, if it exists, can be obtained by row operations; that is, if

$$(A, I) \quad \text{row reduces to} \quad (I, B),$$

then matrix B is the inverse of A. Use this approach to find

$$\begin{pmatrix} 2 & 7 & 3 \\ 3 & 9 & 4 \\ 1 & 5 & 3 \end{pmatrix}^{-1}.$$

24. Show that the following statements are equivalent for $A \in \mathbb{M}_n$.

(a) $PAQ = \left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)$ for some invertible matrices P and Q.

(b) The largest number of column (row) vectors of A that constitute a linearly independent set is r.

(c) A contains an $r \times r$ nonsingular submatrix, and every $(r+1)$-square submatrix has determinant zero.

[Hint: View P and Q as sequences of elementary operations. Note that rank does not change under elementary operations.]

25. Prove Theorem 1.4.

26. Let A and B be $n \times n$ matrices. Show that for any $n \times n$ matrix X,

$$\operatorname{rank}\begin{pmatrix} A & X \\ 0 & B \end{pmatrix} \geq \operatorname{rank}(A) + \operatorname{rank}(B).$$

Discuss the cases where $X = 0$ and $X = I$, respectively.

1.3 Linear Transformations and Eigenvalues

Let V and W be vector spaces over a field $\mathbb{F}$. A map $\mathcal{A} : V \mapsto W$ is called a *linear transformation* from V to W if for all $u, v \in V$, $c \in \mathbb{F}$

$$\mathcal{A}(u + v) = \mathcal{A}(u) + \mathcal{A}(v)$$

and

$$\mathcal{A}(cv) = c\mathcal{A}(v).$$

It is easy to check that $\mathcal{A} : \mathbb{R}^2 \mapsto \mathbb{R}^2$, defined by

$$\mathcal{A}(x_1, x_2) = (x_1 + x_2, x_1 - x_2),$$

is a linear transformation and that the *differential operator* $\mathcal{D}_x$ from $C'[a, b]$, the set (space) of functions with continuous derivatives on the interval $[a, b]$, to $C[a, b]$, the set of continuous functions on $[a, b]$, defined by

$$\mathcal{D}_x(f) = \frac{df(x)}{dx}, \quad f \in C'[a, b],$$

is a linear transformation.

Let $\mathcal{A}$ be a linear transformation from V to W. The subset in W,

$$\mathrm{Im}(\mathcal{A}) = \{\mathcal{A}(v) : \; v \in V\},$$

is a subspace of W, called the *image* of $\mathcal{A}$, and the subset in V,

$$\mathrm{Ker}(\mathcal{A}) = \{v \in V : \; \mathcal{A}(v) = 0 \in W\},$$

is a subspace of V, called the *kernel* or *null space* of $\mathcal{A}$.

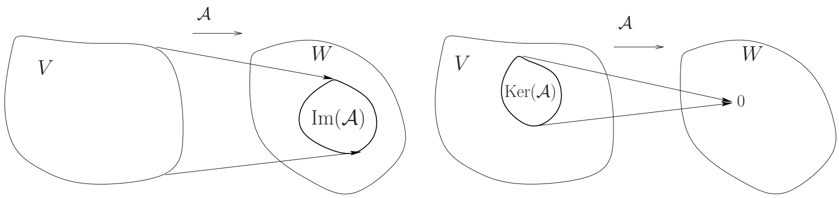

Figure 1.3: Image and kernel

Theorem 1.5 *Let $\mathcal{A}$ be a linear transformation from a vector space V of dimension n to a vector space W. Then*

$$\dim \operatorname{Im}(\mathcal{A}) + \dim \operatorname{Ker}(\mathcal{A}) = n.$$

This is seen by taking a basis $\{u_1, \ldots, u_s\}$ for $\operatorname{Ker}(\mathcal{A})$ and extending it to a basis $\{u_1, \ldots, u_s, v_1, \ldots, v_t\}$ for V, where $s + t = n$. It is easy to show that $\{\mathcal{A}(v_1), \ldots, \mathcal{A}(v_t)\}$ is a basis of $\operatorname{Im}(\mathcal{A})$.

Given an $m \times n$ matrix A with entries in $\mathbb{F}$, one can always define a linear transformation $\mathcal{A}$ from $\mathbb{F}^n$ to $\mathbb{F}^m$ by

$$\mathcal{A}(x) = Ax, \quad x \in \mathbb{F}^n. \tag{1.3}$$

Conversely, linear transformations can be represented by matrices. Consider, for example, $\mathcal{A} : \mathbb{R}^2 \mapsto \mathbb{R}^3$ defined by

$$\mathcal{A}(x_1,\, x_2)^T = (3x_1,\ 2x_1 + x_2,\ -x_1 - 2x_2)^T.$$

Then $\mathcal{A}$ is a linear transformation. We may write in the form

$$\mathcal{A}(x) = Ax,$$

where

$$x = (x_1,\, x_2)^T, \quad A = \begin{pmatrix} 3 & 0 \\ 2 & 1 \\ -1 & -2 \end{pmatrix}.$$

Let $\mathcal{A}$ be a linear transformation from V to W. Once the bases for V and W have been chosen, $\mathcal{A}$ has a unique matrix representation A as in (1.3) determined by the images of the basis vectors of V under $\mathcal{A}$, and there is a one-to-one correspondence between the linear transformations and their matrices. A linear transformation may have different matrices under different bases. In what follows we show that these matrices are similar when $V = W$. Two square matrices A and B of the same size are said to be *similar* if $P^{-1}AP = B$ for some invertible matrix P.

Let $\mathcal{A}$ be a linear transformation on a vector space V with a basis $\{u_1, \ldots, u_n\}$. Since each $\mathcal{A}(u_i)$ is a vector in V, we may write

$$\mathcal{A}(u_i) = \sum_{j=1}^{n} a_{ji} u_j, \quad i = 1, \ldots, n, \tag{1.4}$$

and call $A = (a_{ij})$ the *matrix of $\mathcal{A}$ under the basis* $\{u_1, \dots, u_n\}$.

Write (1.4) conventionally as

$$\mathcal{A}(u_1, \dots, u_n) = (\mathcal{A}(u_1), \dots, \mathcal{A}(u_n)) = (u_1, \dots, u_n)A.$$

Let $v \in V$. If $v = x_1 u_1 + \cdots + x_n u_n$, then

$$\mathcal{A}(v) = \sum_{i=1}^{n} x_i \mathcal{A}(u_i) = (\mathcal{A}(u_1), \dots, \mathcal{A}(u_n))x = (u_1, \dots, u_n)Ax,$$

where x is the column vector $(x_1, \dots, x_n)^T$. In case of $\mathbb{R}^n$ or $\mathbb{C}^n$ with the standard basis $u_1 = e_1, \dots, u_n = e_n$, we have

$$\mathcal{A}(v) = Ax.$$

Let $\{v_1, \dots, v_n\}$ also be a basis of V. Expressing each u_i as a linear combination of $v_1, \dots, v_n$ gives an $n \times n$ matrix B such that

$$(u_1, \dots, u_n) = (v_1, \dots, v_n)B.$$

It can be shown (Problem 10) that B is invertible since $\{u_1, \dots, u_n\}$ is a linearly independent set. It follows by using (1.4) that

$$\begin{aligned} \mathcal{A}(v_1, \dots, v_n) &= \mathcal{A}((u_1, \dots, u_n)B^{-1}) \\ &= (u_1, \dots, u_n)AB^{-1} \\ &= (v_1, \dots, v_n)(BAB^{-1}). \end{aligned}$$

This says that the matrices of a linear transformation under different bases $\{u_1, \dots, u_n\}$ and $\{v_1, \dots, v_n\}$ are similar.

Given a linear transformation on a vector space, it is a central theme of linear algebra to find a basis of the vector space so that the matrix of a linear transformation is as simple as possible, in the sense that the matrix contains more zeros or has a particular structure. In the words of matrices, the given matrix is reduced to a canonical form via similarity. This is discussed in Chapter 3.

Let $\mathcal{A}$ be a linear transformation on a vector space V over $\mathbb{C}$. A nonzero vector $v \in V$ is called an *eigenvector* of $\mathcal{A}$ belonging to an *eigenvalue* $\lambda \in \mathbb{C}$ if

$$\mathcal{A}(v) = \lambda v, \quad v \neq 0.$$

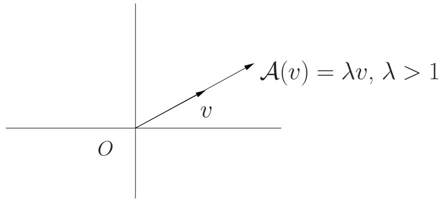

Figure 1.4: Eigenvalue and eigenvector

If, for example, $\mathcal{A}$ is defined on $\mathbb{R}^2$ by

$$\mathcal{A}(x, y) = (y, x),$$

then $\mathcal{A}$ has two eigenvalues, 1 and -1. What are the eigenvectors?

If λ_1 and λ_2 are different eigenvalues of $\mathcal{A}$ with respective eigenvectors x_1 and x_2, then x_1 and x_2 are linearly independent, for if

$$l_1 x_1 + l_2 x_2 = 0 \tag{1.5}$$

for some scalars l_1 and l_2, then applying $\mathcal{A}$ to both sides yields

$$l_1 \lambda_1 x_1 + l_2 \lambda_2 x_2 = 0. \tag{1.6}$$

Multiplying both sides of (1.5) by λ_1, we have

$$l_1 \lambda_1 x_1 + l_2 \lambda_1 x_2 = 0. \tag{1.7}$$

Subtracting (1.6) from (1.7) results in

$$l_2(\lambda_1 - \lambda_2) x_2 = 0.$$

It follows that $l_2 = 0$, and thus $l_1 = 0$ from (1.5).

This can be generalized by induction to the following statement.

If α_{i_j} are linearly independent eigenvectors corresponding to an eigenvalue λ_i, then the set of all eigenvectors α_{i_j} for these eigenvalues λ_i together is linearly independent. Simply put:

Theorem 1.6 *The eigenvectors belonging to different eigenvalues are linearly independent.*

Let $\mathcal{A}$ be a linear transformation on a vector space V of dimension n. If $\mathcal{A}$ happens to have n linearly independent eigenvectors belonging to (not necessarily distinct) eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then $\mathcal{A}$, under the basis formed by the corresponding eigenvectors, has a diagonal matrix representation

$$\begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

To find eigenvalues and eigenvectors, one needs to convert

$$\mathcal{A}(v) = \lambda v$$

under a basis into a linear equation system

$$Ax = \lambda x.$$

Therefore, the eigenvalues of $\mathcal{A}$ are those $\lambda \in \mathbb{F}$ such that

$$\det(\lambda I - A) = 0,$$

and the eigenvectors of $\mathcal{A}$ are the vectors whose coordinates under the basis are the solutions to the equation system $Ax = \lambda x$.

Suppose A is an $n \times n$ complex matrix. The polynomial in λ,

$$p_A(\lambda) = \det(\lambda I_n - A), \tag{1.8}$$

is called the *characteristic polynomial* of A, and the zeros of the polynomial are called the *eigenvalues of* A. It follows that every n-square matrix has n eigenvalues over $\mathbb{C}$ (including repeated ones).

The *trace* of an n-square matrix A, denoted by $\operatorname{tr} A$, is defined to be the sum of the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A, that is,

$$\operatorname{tr} A = \lambda_1 + \cdots + \lambda_n.$$

It is easy to see from (1.8) by expanding the determinant that

$$\operatorname{tr} A = a_{11} + \cdots + a_{nn}$$

and

$$\det A = \prod_{i=1}^{n} \lambda_i.$$

Let $\mathcal{A}$ be a linear transformation on a vector space V. Let W be a subspace of V. If for every $w \in W$, $\mathcal{A}(w) \in W$, we say that W is *invariant* under $\mathcal{A}$. Obviously $\{0\}$ and V are invariant under $\mathcal{A}$. It is easy to check that $\operatorname{Ker} \mathcal{A}$ and $\operatorname{Im} \mathcal{A}$ are invariant under $\mathcal{A}$ too.

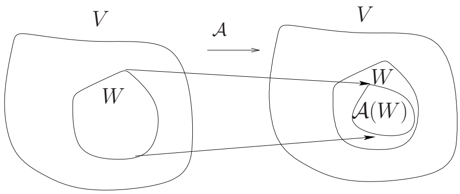

Figure 1.5: Invariant subspace

Let V be a vector space over a field. Consider all linear transformations (*operators*) on V and denote the set by $L(V)$. Then $L(V)$ is a vector space under the following addition and scalar multiplication:

$$(\mathcal{A} + \mathcal{B})(v) = \mathcal{A}(v) + \mathcal{B}(v), \quad (k\mathcal{A})(v) = k\mathcal{A}(v).$$

The zero vector in $L(V)$ is the zero transformation. And for every $\mathcal{A} \in L(V)$, $-\mathcal{A}$ is the operator $(-\mathcal{A})(v) = -(\mathcal{A}(v))$. For two operators $\mathcal{A}$ and $\mathcal{B}$ on V, define the *product* of $\mathcal{A}$ and $\mathcal{B}$ by

$$(\mathcal{A}\mathcal{B})(v) = \mathcal{A}(\mathcal{B}(v)), \quad v \in V.$$

Then $\mathcal{A}\mathcal{B}$ is again a linear transformation on V. The *identity* transformation $\mathcal{I}$ is the one such that $\mathcal{I}(v) = v$ for all $v \in V$.

Problems

1. Show that the map $\mathcal{A}$ from $\mathbb{R}^3$ to $\mathbb{R}^3$ defined by

$$\mathcal{A}(x, y, z) = (x + y, x - y, z)$$

is a linear transformation. Find its matrix under the standard basis.

2. Find the dimensions of $\operatorname{Im}(\mathcal{A})$ and $\operatorname{Ker}(\mathcal{A})$, and find their bases for the linear transformation $\mathcal{A}$ on $\mathbb{R}^3$ defined by

$$\mathcal{A}(x, y, z) = (x - 2z, y + z, 0).$$

3. Define a linear transformation $\mathcal{A} : \mathbb{R}^2 \mapsto \mathbb{R}^2$ by

$$\mathcal{A}(x, y) = (y, 0).$$

   (a) Find $\operatorname{Im}(\mathcal{A})$ and $\operatorname{Ker}(\mathcal{A})$.

   (b) Find a matrix representation of $\mathcal{A}$.

   (c) Verify that $\dim \mathbb{R}^2 = \dim \operatorname{Im}(\mathcal{A}) + \dim \operatorname{Ker}(\mathcal{A})$.

   (d) Is $\operatorname{Im}(\mathcal{A}) + \operatorname{Ker}(\mathcal{A})$ a direct sum?

   (e) Does $\mathbb{R}^2 = \operatorname{Im}(\mathcal{A}) + \operatorname{Ker}(\mathcal{A})$?

4. Find the eigenvalues and eigenvectors of the differential operator $\mathcal{D}_x$.

5. Find the eigenvalues and corresponding eigenvectors of the matrix

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

6. Find the eigenvalues of the matrix

$$A = \begin{pmatrix} 4 & -2 & -1 & 0 \\ -2 & 4 & 0 & -1 \\ -1 & 0 & 4 & -2 \\ 0 & -1 & -2 & 4 \end{pmatrix}.$$

7. Let λ be an eigenvalue of $\mathcal{A}$ on a vector space V, and let

$$V_\lambda = \{v \in V : \mathcal{A}(v) = \lambda v\},$$

called the *eigenspace* of λ. Show that V_λ is an *invariant subspace* of V under $\mathcal{A}$; that is, it is a subspace and $\mathcal{A}(v) \in V_\lambda$ for every $v \in V_\lambda$.

8. Define linear transformations $\mathcal{A}$ and $\mathcal{B}$ on $\mathbb{R}^2$ by

$$\mathcal{A}(x, y) = (x + y, y), \quad \mathcal{B}(x, y) = (x + y, x - y).$$

Find all eigenvalues of $\mathcal{A}$ and $\mathcal{B}$ and their eigenspaces.

9. Let $p(x) = \det(xI - A)$ be the characteristic polynomial of matrix $A \in \mathbb{M}_n$. If λ is an eigenvalue of A such that $p(x) = (x - \lambda)^k q(x)$ for some polynomial $q(x)$ with $q(\lambda) \neq 0$, show that
$$k \geq \dim V_\lambda.$$
[Note: Such a k is known as the *algebraic multiplicity* of λ; $\dim V_\lambda$ is the *geometric multiplicity* of λ. When we say multiplicity of λ, we usually mean the former unless otherwise stated.]

10. Let $\{u_1, \ldots, u_n\}$ and $\{v_1, \ldots, v_n\}$ be two bases of a vector space V. Show that there exists an invertible matrix B such that
$$(u_1, \ldots, u_n) = (v_1, \ldots, v_n)B.$$

11. Let $\{u_1, \ldots, u_n\}$ be a basis for a vector space V and let $\{v_1, \ldots, v_k\}$ be a set of vectors in V. If $v_i = \sum_{j=1}^n a_{ij}u_j$, $i = 1, \ldots, k$, show that
$$\dim \operatorname{Span}\{v_1, \ldots, v_k\} = \operatorname{rank}(A), \quad \text{where } A = (a_{ij}).$$

12. Show that similar matrices have the same trace and determinant.

13. Let v_1 and v_2 be eigenvectors of matrix A belonging to different eigenvalues λ_1 and λ_2, respectively. Show that $v_1 + v_2$ is not an eigenvector of A. How about $av_1 + bv_2$, $a,\ b \in \mathbb{R}$?

14. Let $A \in \mathbb{M}_n$ and let $S \in \mathbb{M}_n$ be nonsingular. If the first column of $S^{-1}AS$ is $(\lambda, 0, \ldots, 0)^T$, show that λ is an eigenvalue of A and that the first column of S is an eigenvector of A belonging to λ.

15. Let $x \in \mathbb{C}^n$. Find the eigenvalues and eigenvectors of the matrices
$$A_1 = xx^* \quad \text{and} \quad A_2 = \begin{pmatrix} 0 & x^* \\ x & 0 \end{pmatrix}.$$

16. If each row sum (i.e., the sum of all entries in a row) of matrix A is 1, show that 1 is an eigenvalue of A.

17. If λ is an eigenvalue of $A \in \mathbb{M}_n$, show that λ^2 is an eigenvalue of A^2 and that if A is invertible, then λ^{-1} is an eigenvalue of A^{-1}.

18. If $x \in \mathbb{C}^n$ is an eigenvector of $A \in \mathbb{M}_n$ belonging to the eigenvalue λ, show that for any $y \in \mathbb{C}^n$, $\lambda + y^*x$ is an eigenvalue of $A + xy^*$.

19. A *minor* of a matrix $A \in \mathbb{M}_n$ is the determinant of a square submatrix of A. Show that

$$\det(\lambda I - A) = \lambda^n - \delta_1 \lambda^{n-1} + \delta_2 \lambda^{n-2} - \cdots + (-1)^n \det A,$$

where δ_i is the sum of all principal minors of order i, $i = 1, 2, \ldots, n-1$. [Note: A principal minor is the determinant of a submatrix indexed by the same rows and columns, called a *principal submatrix*.]

20. A linear transformation $\mathcal{A}$ on a vector space V is said to be *invertible* if there exists a linear transformation $\mathcal{B}$ such that $\mathcal{AB} = \mathcal{BA} = \mathcal{I}$, the identity. If $\dim V < \infty$, show that the following are equivalent.

    (a) $\mathcal{A}$ is invertible.

    (b) If $\mathcal{A}(x) = 0$, then $x = 0$; that is, $\operatorname{Ker}(\mathcal{A}) = \{0\}$.

    (c) If $\{u_1, \ldots, u_n\}$ is a basis for V, then so is $\{\mathcal{A}u_1, \ldots, \mathcal{A}u_n\}$.

    (d) $\mathcal{A}$ is one-to-one.

    (e) $\mathcal{A}$ is onto; that is, $\operatorname{Im}(\mathcal{A}) = V$.

    (f) $\mathcal{A}$ has a nonsingular matrix representation under some basis.

21. Let $\mathcal{A}$ be a linear transformation on a vector space of dimension n with matrix representation A. Show that

$$\dim \operatorname{Im}(\mathcal{A}) = \operatorname{rank}(A) \quad \text{and} \quad \dim \operatorname{Ker}(\mathcal{A}) = n - \operatorname{rank}(A).$$

22. Let $\mathcal{A}$ and $\mathcal{B}$ be linear transformations on a finite-dimensional vector space V having the same image; that is, $\operatorname{Im}(\mathcal{A}) = \operatorname{Im}(\mathcal{B})$. If

$$V = \operatorname{Im}(\mathcal{A}) \oplus \operatorname{Ker}(\mathcal{A}) = \operatorname{Im}(\mathcal{B}) \oplus \operatorname{Ker}(\mathcal{B}),$$

does it follow that $\operatorname{Ker}(\mathcal{A}) = \operatorname{Ker}(\mathcal{B})$?

23. Consider the vector space $\mathbb{F}[x]$ of all polynomials over $\mathbb{F}(= \mathbb{R}$ or $\mathbb{Q})$. For $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in \mathbb{F}[x]$, define

$$\mathcal{S}(f(x)) = \frac{a_n}{n+1} x^{n+1} + \frac{a_{n-1}}{n} x^n + \cdots + \frac{a_1}{2} x^2 + a_0 x$$

and

$$\mathcal{T}(f(x)) = n a_n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \cdots + a_1.$$

Compute $\mathcal{ST}$ and $\mathcal{TS}$. Does $\mathcal{ST} = \mathcal{TS}$?

24. Define $\mathcal{P} : \mathbb{C}^n \mapsto \mathbb{C}^n$ by $\mathcal{P}(x) = (0, 0, x_3, \ldots, x_n)$. Show that $\mathcal{P}$ is a linear transformation and $\mathcal{P}^2 = \mathcal{P}$. What is $\operatorname{Ker}(\mathcal{P})$?

25. Let $\mathcal{A}$ be a linear transformation on a finite-dimensional vector space V, and let W be a subspace of V. Denote

$$\mathcal{A}(W) = \{\mathcal{A}(w) : \ w \in W\}.$$

Show that $\mathcal{A}(W)$ is a subspace of V. Furthermore, show that

$$\dim(\mathcal{A}(W)) + \dim(\operatorname{Ker}(\mathcal{A}) \cap W) = \dim W.$$

26. Let V be a vector space of dimension n over $\mathbb{C}$ and let $\{u_1, \ldots, u_n\}$ be a basis of V. For $x = x_1u_1 + \cdots + x_nu_n \in V$, define

$$\mathcal{T}(x) = (x_1, \ldots, x_n) \in \mathbb{C}^n.$$

Show that $\mathcal{T}$ is an *isomorphism*, or $\mathcal{T}$ is one-to-one, onto, and satisfies

$$\mathcal{T}(ax + by) = a\mathcal{T}(x) + b\mathcal{T}(y), \quad x, \, y \in V, \ a, \, b \in \mathbb{C}.$$

27. Let V be the vector space of all sequences

$$(c_1, c_2, \ldots), \quad c_i \in \mathbb{C}, \ i = 1, 2, \ldots.$$

Define a linear transformation on V by

$$\mathcal{S}(c_1, c_2, \cdots) = (0, c_1, c_2, \cdots).$$

Show that $\mathcal{S}$ has no eigenvalues. Moreover, if we define

$$\mathcal{S}^*(c_1, c_2, c_3, \cdots) = (c_2, c_3, \cdots),$$

then $\mathcal{S}^*\mathcal{S}$ is the identity, but $\mathcal{S}\mathcal{S}^*$ is not.

28. Let $\mathcal{A}$ be a linear operator on a vector space V of dimension n. Let

$$V_0 = \cup_{i=1}^{\infty} \operatorname{Ker}(\mathcal{A}^i), \quad V_1 = \cap_{i=1}^{\infty} \operatorname{Im}(\mathcal{A}^i).$$

Show that V_0 and V_1 are invariant under $\mathcal{A}$ and that $V = V_0 \oplus V_1$.

1.4 Inner Product Spaces

A vector space V over the number field $\mathbb{C}$ or $\mathbb{R}$ is called an *inner product space* if it is equipped with an *inner product* $(\ ,\ )$ satisfying for all $u, v, w \in V$ and scalar c,

1. $(u, u) \geq 0$, and $(u, u) = 0$ if and only if $u = 0$,
2. $(u + v, w) = (u, w) + (v, w)$,
3. $(cu, v) = c(u, v)$, and
4. $\overline{(u, v)} = (v, u)$.

$\mathbb{C}^n$ is an inner product space over $\mathbb{C}$ with the inner product

$$(x, y) = y^* x = \overline{y_1}\, x_1 + \cdots + \overline{y_n}\, x_n.$$

An inner product space over $\mathbb{R}$ is usually called a *Euclidean space.*

The *Cauchy–Schwarz inequality* for an inner product space is one of the most useful inequalities in mathematics.

Theorem 1.7 (Cauchy–Schwarz Inequality) *Let V be an inner product space. Then for all vectors x and y in V,*

$$|(x, y)|^2 \leq (x, x)(y, y).$$

Equality holds if and only if x and y are linearly dependent.

The proof of this can be done in a number of different ways. The most common proof is to consider the quadratic function in t

$$(x + ty, x + ty) \geq 0$$

and to derive the inequality from the nonpositive discriminant. One may also obtain the inequality from $(z, z) \geq 0$ by setting

$$z = y - \frac{(y, x)}{(x, x)} x, \quad x \neq 0,$$

and showing that $(z, x) = 0$ and then $(z, z) = (z, y) \geq 0$.

A matrix proof is left as an exercise for the reader (Problem 13).

For any vector x in an inner product space, the positive square root of (x, x) is called the *length* or *norm* of the vector x and is denoted by $\|x\|$; that is,

$$\|x\| = \sqrt{(x, x)}.$$

Thus, the Cauchy–Schwarz inequality is rewritten as

$$|(x, y)| \le \|x\| \, \|y\|.$$

Theorem 1.8 *For all vectors x and y in an inner product space,*

$$\text{i. } \|x\| \ge 0; \quad \text{ii. } \|cx\| = |c| \|x\|,\ c \in \mathbb{C}; \quad \text{iii. } \|x + y\| \le \|x\| + \|y\|.$$

The last inequality is referred to as the *triangle inequality.*

A *unit* vector is a vector whose length is 1. For any nonzero vector u, $\frac{1}{\|u\|}u$ is a unit vector. Two vectors x and y are said to be *orthogonal* if $(x, y) = 0$. An *orthogonal set* is a set in which any two of the vectors are orthogonal. Such a set is further said to be *orthonormal* if every vector in the set is of length 1.

For example, $\{v_1, v_2\}$ is an orthonormal set in $\mathbb{R}^2$, where

$$v_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \quad v_2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right).$$

The column (row) vectors of a unitary matrix are orthonormal.

Let S be a subset of an inner product space V. Denote by $S^\perp$ the collection of the vectors in V that are orthogonal to all vectors in S; that is,

$$S^\perp = \{v \in V : \ (v, s) = 0 \ \text{ for all } s \in S\}.$$

It is easy to see that $S^\perp$ is a subspace of V. If S contains only one element, say x, we simply use $x^\perp$ for $S^\perp$. For two subsets S_1 and S_2, if $(x, y) = 0$ for all $x \in S_1$ and $y \in S_2$, we write $S_1 \perp S_2$.

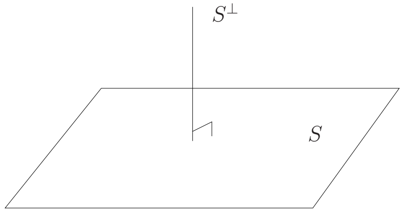

Figure 1.6: Orthogonality

As we saw in the first section, a set of linearly independent vectors of a vector space of finite dimension can be extended to a basis for the vector space. Likewise a set of orthogonal vectors of an inner product space can be extended to an orthogonal basis of the space. The same is true for a set of orthonormal vectors. Consider $\mathbb{C}^n$, for instance. Let u_1 be a unit vector in $\mathbb{C}^n$. Pick a unit vector u_2 in $u_1^\perp$ if $n \geq 2$. Then u_1 and u_2 are orthonormal. Now if $n \geq 3$, let u_3 be a unit vector in $(\text{Span}\{u_1, u_2\})^\perp$ (equivalently, $(u_1, u_3) = 0$ and $(u_2, u_3) = 0$). Then u_1, u_2, u_3 are orthonormal. Continuing this way, one obtains an orthonormal basis for the inner product space. We summarize this as a theorem for $\mathbb{C}^n$, which will be freely and frequently used in the future.

Theorem 1.9 *If $u_1, \ldots, u_k$ are k linearly independent column vectors in $\mathbb{C}^n$, $1 \leq k < n$, then there exist $n - k$ column vectors $u_{k+1}, \ldots, u_n$ in $\mathbb{C}^n$ such that the matrix*

$$P = (u_1, \ldots, u_k, u_{k+1}, \ldots, u_n)$$

is invertible. Furthermore, if $u_1, \ldots, u_k$ are orthonormal, then there exist unit $n - k$ vectors $u_{k+1}, \ldots, u_n$ in $\mathbb{C}^n$ such that the matrix

$$U = (u_1, \ldots, u_k, u_{k+1}, \ldots, u_n)$$

is unitary. In particular, for any unit vector u in $\mathbb{C}^n$, there exists a unitary matrix that contains u as its first column.

If $\{u_1, \ldots, u_n\}$ is an orthonormal basis of an inner product space V over $\mathbb{C}$, and if x and y are two vectors in V expressed as

$$x = x_1 u_1 + \cdots + x_n u_n, \quad y = y_1 u_1 + \cdots + y_n u_n,$$

then $x_i = (x, u_i)$, $y_i = (y, u_i)$ for $i = 1, \ldots, n$,

$$\|x\| = (|x_1|^2 + \cdots + |x_n|^2)^{1/2} \tag{1.9}$$

and

$$(x, y) = \overline{y_1}\, x_1 + \cdots + \overline{y_n}\, x_n. \tag{1.10}$$

For $A \in \mathbb{M}_n$ and with the standard basis $e_1, \ldots, e_n$ of $\mathbb{C}^n$, we have

$$\operatorname{tr} A = \sum_{i=1}^{n} (Ae_i, e_i)$$

and for $x \in \mathbb{C}^n$

$$(Ax,\, x) = x^* Ax = \sum_{i,\, j=1}^{n} a_{ij}\overline{x_i}x_j.$$

Upon computation, we have

$$(Ax,\, y) = (x,\, A^* y)$$

and, with $\operatorname{Im} A = \{Ax :\, x \in \mathbb{C}^n\}$ and $\operatorname{Ker} A = \{x \in \mathbb{C}^n :\, Ax = 0\}$,

$$\operatorname{Ker} A^* = (\operatorname{Im} A)^{\perp}, \quad \operatorname{Im} A^* = (\operatorname{Ker} A)^{\perp}. \tag{1.11}$$

$\mathbb{M}_n$ is an inner product space with the inner product

$$(A, B)_{\mathbb{M}} = \operatorname{tr}(B^* A), \quad A,\ B \in \mathbb{M}_n.$$

It is immediate by the Cauchy–Schwarz inequality that

$$|\operatorname{tr}(AB)|^2 \le \operatorname{tr}(A^* A)\operatorname{tr}(B^* B)$$

and that

$$\operatorname{tr}(A^* A) = 0 \quad \text{if and only if } A = 0.$$

We end this section by presenting an inequality of the angles between vectors in a Euclidean space. This inequality is intuitive and obvious in $\mathbb{R}^2$ and $\mathbb{R}^3$. The good part of this theorem is the idea in its proof of reducing the problem to $\mathbb{R}^2$ or $\mathbb{R}^3$.

Let V be an inner product space over $\mathbb{R}$. For any nonzero vectors x and y, define the (measure of) *angle* between x and y by

$$\angle_{x,\, y} = \cos^{-1} \frac{(x,\, y)}{\|x\|\|y\|}.$$

Theorem 1.10 *For any nonzero vectors x, y, and z in a Euclidean space (i.e., an inner product space V over $\mathbb{R}$),*

$$\angle_{x,z} \le \angle_{x,y} + \angle_{y,z}.$$

Equality occurs if and only if $y = ax + bz, \ a, b \ge 0$.

Proof. Because the inequality involves only the vectors x, y, and z, we may focus on the subspace $\operatorname{Span}\{x, y, z\}$, which has dimension at most 3. We can further choose an orthonormal basis (a unit vector in the case of dimension one) for this subspace. Let x, y, and z have coordinate vectors α, β, and γ under the basis, respectively. Then the inequality holds if and only if it holds for real vectors α, β, and γ, due to (1.9) and (1.10). Thus, the problem is reduced to $\mathbb{R}$, $\mathbb{R}^2$, or $\mathbb{R}^3$ depending on whether the dimension of $\operatorname{Span}\{x, y, z\}$ is 1, 2, or 3, respectively. For $\mathbb{R}$, the assertion is trivial. For $\mathbb{R}^2$ or $\mathbb{R}^3$, a simple graph will do the job. ∎

Problems

1. If V is an inner product space over $\mathbb{C}$, show that for $x, y \in V$, $c \in \mathbb{C}$,

$$(x, cy) = \bar{c}(x, y) \quad \text{and} \quad (x, y)(y, x) = |(x, y)|^2.$$

2. Find all vectors in $\mathbb{R}^2$ (with the usual inner product) that are orthogonal to $(1, 1)$. Is $(1, 1)$ a unit vector?

3. Show that in an inner product space over $\mathbb{R}$ or $\mathbb{C}$

$$(x, y) = 0 \quad \Rightarrow \quad \|x + y\|^2 = \|x\|^2 + \|y\|^2$$

and that the converse is true over $\mathbb{R}$ but not over $\mathbb{C}$.

4. Let V be an inner product space over $\mathbb{R}$ and let $x, y \in V$. Show that

$$\|x\| = \|y\| \quad \Rightarrow \quad (x + y, x - y) = 0.$$

5. Show that for any two vectors x and y in an inner product space

$$\|x - y\|^2 + \|x + y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

6. Is $[\ ,\ ]$ defined as $[x, y] = x_1 y_1 + \cdots + x_n y_n$ an inner product for $\mathbb{C}^n$?

7. For what diagonal $D \in \mathbb{M}_n$ is $[x, y] = y^* D x$ an inner product for $\mathbb{C}^n$?

8. Let $\{u_1, \ldots, u_n\}$ be an orthonormal basis of an inner product space V. Show that for $x \in V$

$$(u_i, x) = 0, \quad i = 1, \ldots, n, \quad \Leftrightarrow \quad x = 0,$$

and that for $x,\ y \in V$

$$(u_i, x) = (u_i,\ y), \quad i = 1, \ldots, n, \quad \Leftrightarrow \quad x = y.$$

9. Let $\{v_1, \ldots, v_n\}$ be a set of vectors in an inner product space V. Denote the matrix with entries (v_i, v_j) by G. Show that $\det G = 0$ if and only if $v_1, \ldots, v_n$ are linearly dependent.

10. Let A be an n-square complex matrix. Show that

$$\operatorname{tr}(AX) = 0 \ \text{ for every } X \in \mathbb{M}_n \quad \Leftrightarrow \quad A = 0.$$

11. Let $A \in \mathbb{M}_n$. Show that for any unit vector $x \in \mathbb{C}^n$,

$$|x^* A x|^2 \le x^* A^* A x.$$

12. Let $A = (a_{ij})$ be a complex matrix. Show that

$$\operatorname{tr}(A^* A) = \operatorname{tr}(A A^*) = \sum_{i,\,j} |a_{ij}|^2.$$

13. Use the fact that $\operatorname{tr}(A^* A) \ge 0$ with equality if and only if $A = 0$ to show the Cauchy–Schwarz inequality for $\mathbb{C}^n$ by taking $A = xy^* - yx^*$.

14. Let A and B be complex matrices of the same size. Show that

$$|\operatorname{tr}(A^* B)| \le \big(\operatorname{tr}(A^* A) \operatorname{tr}(B^* B)\big)^{1/2} \le \frac{1}{2}\big(\operatorname{tr}(A^* A) + \operatorname{tr}(B^* B)\big).$$

If A and B are both n-square, then $\operatorname{tr}(A^* B)$ on the left-hand side may be replaced by $\operatorname{tr}(AB)$, even though $\operatorname{tr}(A^* B) \ne \operatorname{tr}(AB)$. Why?

15. Let A and B be n-square complex matrices. If, for every $x \in \mathbb{C}^n$,

$$(Ax,\ x) = (Bx,\ x),$$

does it follow that $A = B$? What if $x \in \mathbb{R}^n$?

16. Show that for any n-square complex matrix A

$$\mathbb{C}^n = \operatorname{Im} A \oplus (\operatorname{Im} A)^\perp = \operatorname{Im} A \oplus \operatorname{Ker} A^*$$

and that $\mathbb{C}^n = \operatorname{Im} A \oplus \operatorname{Ker} A$ if and only if $\operatorname{rank}(A^2) = \operatorname{rank}(A)$.

17. Let θ_i and λ_i be positive numbers and $\sum_{i=1}^n \theta_i = 1$. Show that

$$1 \leq \left(\sum_{i=1}^n \theta_i\lambda_i\right)\left(\sum_{i=1}^n \theta_i\lambda_i^{-1}\right).$$

18. Let $\{u_1, \ldots, u_n\}$ be an orthonormal basis of an inner product space V. Show that $x_1, \ldots, x_m$ in V are pairwise orthogonal if and only if

$$\sum_{k=1}^n (x_i, u_k)\overline{(x_j, u_k)} = 0, \quad i \neq j.$$

19. If $\{u_1, \ldots, u_k\}$ is an orthonormal set in an inner product space V of dimension n, show that $k \leq n$ and for any $x \in V$,

$$\|x\|^2 \geq \sum_{i=1}^k |(x, u_i)|^2.$$

20. Let $\{u_1, \ldots, u_n\}$ and $\{v_1, \ldots, v_n\}$ be two orthonormal bases of an inner product space. Show that there exists a unitary U such that

$$(u_1, \ldots, u_n) = (v_1, \ldots, v_n)U.$$

21. (*Gram–Schmidt Orthonormalization*) Let $x_1, x_2, \ldots, x_n$ be linearly independent vectors in an inner product space. Let $y_1 = x_1$ and define $z_1 = \|y_1\|^{-1}y_1$; let $y_2 = x_2 - (x_2, z_1)z_1$ and define $z_2 = \|y_2\|^{-1}y_2$. Then z_1, z_2 are orthonormal. Continue this process inductively: $y_k = x_k - (x_k, z_{k-1})z_{k-1} - (x_k, z_{k-2})z_{k-2} - \cdots - (x_k, z_1)z_1$, $z_k = \|y_k\|^{-1}y_k$. Show that the vectors $z_1, z_2, \ldots, z_n$ are orthonormal.

22. Prove or disprove for unit vectors u, v, w in an inner product space

$$|(u, w)| \leq |(u, v)| + |(v, w)|.$$

23. Let V be an inner product space and $\|\cdot\|$ be the *induced norm*, that is, $\|u\| = \sqrt{(u, u)}$, $u \in V$. Show that for vectors x and y in V,

$$\|x + y\| = \|x\| + \|y\|$$

if and only if one of the vectors is a nonnegative multiple of the other.

24. If the angle between nonzero vectors x and y in an inner product space over $\mathbb{R}$ or $\mathbb{C}$ is defined by $<_{x,y} = \cos^{-1} \frac{|(x, y)|}{\|x\|\|y\|}$, show that $x \perp y$ if and only if $<_{x,y} = \frac{\pi}{2}$. Explain why the law of cosines does not hold.

25. Let V be a (not necessarily inner product) vector space over a field $\mathbb{F}$ ($= \mathbb{R}$ or $\mathbb{C}$). We say that V is a *normed space* if it is equipped with a function $\|\cdot\| : V \mapsto \mathbb{R}$, called a *vector norm*, satisfying

$$\text{(i)}\ \ \|x\| \geq 0, \quad \text{(ii)}\ \ \|cx\| = |c|\,\|x\|, \quad \text{(iii)}\ \ \|x+y\| \leq \|x\| + \|y\|,$$

for all $x, y \in V$, $c \in \mathbb{F}$, and $\|x\| = 0$ if and only if $x = 0$. Show that

(a) If $x \to 0$ in $V = \mathbb{R}^n$ entrywise, then $\|x\| \to 0$.

(b) $|\,\|x\| - \|y\|\,| \leq \|x - y\|$ for all vectors x and y in V.

(c) Give an example showing it is possible that $\|x+y\| = \|x\| + \|y\|$ for some vectors x and y that are linearly independent.

26. Let V_1 and V_2 be subsets of an inner product space V. Show that

$$V_1 \subseteq V_2 \ \Rightarrow\ V_2^\perp \subseteq V_1^\perp.$$

27. Let V_1 and V_2 be subspaces of an inner product space V. Show that

$$(V_1 + V_2)^\perp = V_1^\perp \cap V_2^\perp$$

and

$$(V_1 \cap V_2)^\perp = V_1^\perp + V_2^\perp.$$

28. Let V_1 and V_2 be subspaces of an inner product space V of dimension n. If $\dim V_1 > \dim V_2$, show that there exists a subspace V_3 such that

$$V_3 \subset V_1, \quad V_3 \perp V_2, \quad \dim V_3 \geq \dim V_1 - \dim V_2.$$

Give a geometric explanation of this in $\mathbb{R}^3$.

29. Let A and B be $m \times n$ complex matrices. Show that

$$\operatorname{Im} A \perp \operatorname{Im} B \quad \Leftrightarrow \quad A^*B = 0.$$

30. Let A be an n-square complex matrix. Show that for any x, $y \in \mathbb{C}^n$

$$4(Ax, y) = (As, s) - (At, t) + i(Au, u) - i(Av, v),$$

where $s = x + y$, $t = x - y$, $u = x + iy$, and $v = x - iy$.

31. Let u be a nonzero vector in an inner product space V. If $v_1, v_2, \ldots, v_k$ are vectors in V such that (i) $(v_i, u) > 0$ for all i and (ii) $(v_i, v_j) \leq 0$ whenever $i \neq j$, show that $v_1, v_2, \ldots, v_k$ are linearly independent.

32. Show that for any nonzero vectors x and y in $\mathbb{C}^n$

$$\|x - y\| \geq \frac{1}{2}(\|x\| + \|y\|)\left\| \frac{1}{\|x\|}x - \frac{1}{\|y\|}y \right\|.$$

CHAPTER 2

Partitioned Matrices, Rank, and Eigenvalues

Introduction: We begin with the elementary operations on partitioned (block) matrices, followed by discussions of the inverse and rank of the sum and product of matrices. We then present four different proofs of the theorem that the products AB and BA of matrices A and B of sizes $m \times n$ and $n \times m$, respectively, have the same nonzero eigenvalues. At the end of this chapter we discuss the often-used matrix technique of continuity argument and the tool for localizing eigenvalues by means of the Geršgorin discs.

2.1 Elementary Operations of Partitioned Matrices

The manipulation of partitioned matrices is a basic tool in matrix theory. The techniques for manipulating partitioned matrices resemble those for ordinary numerical matrices. We begin by considering a 2×2 matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad a,\ b,\ c,\ d \in \mathbb{C}.$$

An application of an elementary row operation, say, adding the second row multiplied by -3 to the first row, can be represented by the

F. Zhang, *Matrix Theory: Basic Results and Techniques*, Universitext,
DOI 10.1007/978-1-4614-1099-7_2,

matrix multiplication

$$\begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a-3c & b-3d \\ c & d \end{pmatrix}.$$

Elementary row or column operations for matrices play an important role in elementary linear algebra. These operations (Section 1.2) can be generalized to partitioned matrices as follows.

I. Interchange two block rows (columns).

II. Multiply a block row (column) from the left (right) by a nonsingular matrix of appropriate size.

III. Multiply a block row (column) by a matrix from the left (right), then add it to another row (column).

Write in matrices, say, for type III elementary row operations,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \to \begin{pmatrix} A & B \\ C+XA & D+XB \end{pmatrix},$$

where $A \in \mathbb{M}_m$, $D \in \mathbb{M}_n$, and X is $n \times m$. Note that A is multiplied by X from the left (when row operations are performed).

Generalized elementary matrices are those obtained by applying a single elementary operation to the identity matrix. For instance,

$$\begin{pmatrix} 0 & I_m \\ I_n & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} I_m & 0 \\ X & I_n \end{pmatrix}$$

are generalized elementary matrices of type I and type III.

Theorem 2.1 *Let G be the generalized elementary matrix obtained by performing an elementary row (column) operation on I. If that same elementary row (column) operation is performed on a block matrix A, then the resulting matrix is given by the product GA (AG).*

Proof. We show the case of 2×2 partitioned matrices Because we deal with this type of partitioned matrix most of the time. An argument for the general case is similar.

Let A, B, C, and D be matrices, where A and D are m- and n-square, respectively. Suppose we apply a type III operation, say, adding the first row times an $n \times m$ matrix E from the left to the second row, to the matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}. \tag{2.1}$$

Then we have, by writing in equation,

$$\begin{pmatrix} A & B \\ C+EA & D+EB \end{pmatrix} = \begin{pmatrix} I_m & 0 \\ E & I_n \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \quad \blacksquare$$

As an application, suppose that A is invertible. By successively applying suitable elementary row and column block operations, we can change the matrix (2.1) so that the lower-left and upper-right submatrices become 0. More precisely, we can make the lower-left and upper-right submatrices 0 by subtracting the first row multiplied by CA^{-1} from the the second row, and by subtracting the first column multiplied by $A^{-1}B$ from the second column. In symbols,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \to \begin{pmatrix} A & B \\ 0 & D-CA^{-1}B \end{pmatrix} \to \begin{pmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{pmatrix},$$

and in equation form,

$$\begin{pmatrix} I_m & 0 \\ -CA^{-1} & I_n \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I_m & -A^{-1}B \\ 0 & I_n \end{pmatrix}$$
$$= \begin{pmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{pmatrix}. \tag{2.2}$$

Note that by taking determinants,

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = \det A \ \det(D - CA^{-1}B).$$

The method of manipulating block matrices by elementary operations and the corresponding generalized elementary matrices as in (2.2) is used repeatedly in this book.

For practice, we now consider expressing the block matrix

$$\begin{pmatrix} A & B \\ 0 & A^{-1} \end{pmatrix} \tag{2.3}$$

as a product of block matrices of the forms

$$\begin{pmatrix} I & X \\ 0 & I \end{pmatrix}, \quad \begin{pmatrix} I & 0 \\ Y & I \end{pmatrix}.$$

In other words, we want to get a matrix in the above form by performing type III operations on the block matrix in (2.3).

Add the first row of (2.3) times A^{-1} to the second row to get

$$\begin{pmatrix} A & B \\ I & A^{-1} + A^{-1}B \end{pmatrix}.$$

Add the second row multiplied by $I - A$ to the first row to get

$$\begin{pmatrix} I & A^{-1} + A^{-1}B - I \\ I & A^{-1} + A^{-1}B \end{pmatrix}.$$

Subtract the first row from the second row to get

$$\begin{pmatrix} I & A^{-1} + A^{-1}B - I \\ 0 & I \end{pmatrix},$$

which is in the desired form. Putting these steps in identity, we have

$$\begin{pmatrix} I & 0 \\ -I & I \end{pmatrix}\begin{pmatrix} I & I - A \\ 0 & I \end{pmatrix}\begin{pmatrix} I & 0 \\ A^{-1} & I \end{pmatrix}\begin{pmatrix} A & B \\ 0 & A^{-1} \end{pmatrix}$$
$$= \begin{pmatrix} I & A^{-1} + A^{-1}B - I \\ 0 & I \end{pmatrix}.$$

Therefore,

$$\begin{aligned} \begin{pmatrix} A & B \\ 0 & A^{-1} \end{pmatrix} &= \begin{pmatrix} I & 0 \\ A^{-1} & I \end{pmatrix}^{-1}\begin{pmatrix} I & I - A \\ 0 & I \end{pmatrix}^{-1} \\ &\quad \times \begin{pmatrix} I & 0 \\ -I & I \end{pmatrix}^{-1}\begin{pmatrix} I & A^{-1} + A^{-1}B - I \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} I & 0 \\ -A^{-1} & I \end{pmatrix}\begin{pmatrix} I & A - I \\ 0 & I \end{pmatrix} \\ &\quad \times \begin{pmatrix} I & 0 \\ I & I \end{pmatrix}\begin{pmatrix} I & A^{-1} + A^{-1}B - I \\ 0 & I \end{pmatrix} \end{aligned}$$

is a product of type III generalized elementary matrices.

Problems

1. Let $E = E[i(c) \to j]$ denote the elementary matrix obtained from I_n by adding row i times c to row j.

   (a) Show that $E^* = E[j(\bar{c}) \to i]$.

   (b) Show that $E^{-1} = E[i(-c) \to j]$.

   (c) How is E obtained via an elementary column operation?

2. Show that

$$A(B, C) = (AB, AC) \text{ for } A \in \mathbb{M}_{m\times n},\ B \in \mathbb{M}_{n\times p},\ C \in \mathbb{M}_{n\times q};$$

$$\begin{pmatrix} A \\ B \end{pmatrix} C = \begin{pmatrix} AC \\ BC \end{pmatrix} \text{ for } A \in \mathbb{M}_{p\times n},\ B \in \mathbb{M}_{q\times n},\ C \in \mathbb{M}_{n\times m}.$$

3. Let X be any $n \times m$ complex matrix. Show that

$$\begin{pmatrix} I_m & 0 \\ X & I_n \end{pmatrix}^{-1} = \begin{pmatrix} I_m & 0 \\ -X & I_n \end{pmatrix}.$$

4. Show that for any n-square complex matrix X,

$$\begin{pmatrix} X & I_n \\ I_n & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & I_n \\ I_n & -X \end{pmatrix}.$$

   Does it follow that

$$\begin{pmatrix} 0 & I_m \\ I_n & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & I_m \\ I_n & 0 \end{pmatrix}?$$

5. Show that every 2×2 matrix of determinant 1 is the product of some matrices of the following types, with $y \neq 0$:

$$\begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}, \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} y & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & y \end{pmatrix}.$$

6. Let X and Y be matrices with the same number of rows. Multiply

$$\begin{pmatrix} X & Y \\ 0 & 0 \end{pmatrix}\begin{pmatrix} X^* & 0 \\ Y^* & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} X^* & 0 \\ Y^* & 0 \end{pmatrix}\begin{pmatrix} X & Y \\ 0 & 0 \end{pmatrix}.$$

7. Let X and Y be complex matrices of the same size. Verify that

$$\begin{pmatrix} I + XX^* & X + Y \\ X^* + Y^* & I + Y^*Y \end{pmatrix} = \begin{pmatrix} I & X \\ Y^* & I \end{pmatrix}\begin{pmatrix} I & Y \\ X^* & I \end{pmatrix} =$$
$$\begin{pmatrix} X & I \\ I & Y^* \end{pmatrix}\begin{pmatrix} X^* & I \\ I & Y \end{pmatrix} = \begin{pmatrix} X \\ I \end{pmatrix}(X^*, I) + \begin{pmatrix} I \\ Y^* \end{pmatrix}(I, Y).$$

8. Let $a_1, a_2, \ldots, a_n$ be complex numbers, $a = (-a_2, \ldots, -a_n)$, and

$$A = \begin{pmatrix} 0 & I_{n-1} \\ -a_1 & a \end{pmatrix}.$$

Find $\det A$. Show that A is invertible if $a_1 \neq 0$ and that

$$A^{-1} = \begin{pmatrix} \frac{1}{a_1}a & -\frac{1}{a_1} \\ I_{n-1} & 0 \end{pmatrix}.$$

9. Let $a_1, a_2, \ldots, a_n$ be nonzero complex numbers. Find

$$\begin{pmatrix} 0 & a_1 & 0 & \ldots & 0 \\ 0 & 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & a_{n-1} \\ a_n & 0 & 0 & \ldots & 0 \end{pmatrix}^{-1}.$$

10. Show that the block matrices $\left(\begin{smallmatrix} I_m & A \\ 0 & I_n \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} I_m & B \\ 0 & I_n \end{smallmatrix}\right)$ commute.

11. Show that a *generalized elementary matrix* $\left(\begin{smallmatrix} I & X \\ 0 & I \end{smallmatrix}\right)$ can be written as the product of the same type of elementary matrices with only one nonzero off-diagonal entry. [Hint: See how to get (i, j)-entry x_{ij} in the matrix by a type iii elementary operation from Section 1.2.]

12. Let A and B be nonsingular matrices. Find $\left(\begin{smallmatrix} A & C \\ 0 & B \end{smallmatrix}\right)^{-1}$.

13. Let A and B be m- and n-square matrices, respectively. Show that

$$\begin{pmatrix} A & * \\ 0 & B \end{pmatrix}^k = \begin{pmatrix} A^k & * \\ 0 & B^k \end{pmatrix},$$

where the $*$ are some matrices, and that if A and B are invertible,

$$\begin{pmatrix} A & * \\ 0 & B \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & * \\ 0 & B^{-1} \end{pmatrix}.$$

14. Let A and B be $n \times n$ complex matrices. Show that

$$\begin{vmatrix} 0 & A \\ B & 0 \end{vmatrix} = (-1)^n \det A \det B$$

and that if A and B are invertible, then

$$\begin{pmatrix} 0 & A \\ B & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & B^{-1} \\ A^{-1} & 0 \end{pmatrix}.$$

15. Let A and B be $n \times n$ matrices. Apply elementary operations to $\left(\begin{smallmatrix} A & 0 \\ I & B \end{smallmatrix}\right)$ to get $\left(\begin{smallmatrix} AB & 0 \\ B & I \end{smallmatrix}\right)$. Deduce $\det(AB) = \det A \det B$.

16. Let A and B be $n \times n$ matrices. Apply elementary operations to $\left(\begin{smallmatrix} I & A \\ B & I \end{smallmatrix}\right)$ to get $\left(\begin{smallmatrix} I-AB & 0 \\ 0 & I \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} I & 0 \\ 0 & I-AB \end{smallmatrix}\right)$. Conclude that $I - AB$ and $I - BA$ have the same rank for any $A,\ B \in \mathbb{M}_n$.

17. Let A, $B \in \mathbb{M}_n$. Show that $\left(\begin{smallmatrix} A & B \\ B & A \end{smallmatrix}\right)$ is similar to $\left(\begin{smallmatrix} A+B & 0 \\ 0 & A-B \end{smallmatrix}\right)$.

18. Let A and B be $n \times n$ matrices. Apply elementary operations to $\left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right)$ to get $\left(\begin{smallmatrix} A+B & B \\ B & B \end{smallmatrix}\right)$ and derive the rank inequality

$$\operatorname{rank}(A+B) \le \operatorname{rank}(A) + \operatorname{rank}(B).$$

19. Let A be a square complex matrix partitioned as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad A_{11} \in \mathbb{M}_m,\ A_{22} \in \mathbb{M}_n.$$

Show that for any $B \in \mathbb{M}_m$

$$\begin{vmatrix} BA_{11} & BA_{12} \\ A_{21} & A_{22} \end{vmatrix} = \det B\ \det A$$

and for any $n \times m$ matrix C

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{21} + CA_{11} & A_{22} + CA_{12} \end{vmatrix} = \det A.$$

20. Let A be a square complex matrix. Show that

$$\begin{vmatrix} I & A \\ A^* & I \end{vmatrix} = 1 - \sum M_1^* M_1 + \sum M_2^* M_2 - \sum M_3^* M_3 + \cdots,$$

where each M_k is a minor of order $k = 1, 2, \ldots$. [Hint: Reduce the left-hand side to $\det(I - A^*A)$ and use Problem 19 of Section 1.3.]

2.2 The Determinant and Inverse of Partitioned Matrices

Let M be a square complex matrix partitioned as

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where A and D are m- and n-square matrices, respectively. We discuss the determinants and inverses of matrices in this form. The results are fundamental and used almost everywhere in matrix theory, such as matrix computation and matrix inequalities. The methods of continuity and finding inverses deserve special attention.

Theorem 2.2 *Let M be a square matrix partitioned as above. Then*

$$\det M = \det A \ \det(D - CA^{-1}B), \quad \text{if } A \text{ is invertible,}$$

and

$$\det M = \det(AD - CB), \quad \text{if } AC = CA.$$

Proof. When A^{-1} exists, it is easy to verify (see also (2.2)) that

$$\begin{pmatrix} I_m & 0 \\ -CA^{-1} & I_n \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A & B \\ 0 & D - CA^{-1}B \end{pmatrix}.$$

By taking determinants for both sides, we have

$$\begin{aligned} \det M &= \begin{vmatrix} A & B \\ 0 & D - CA^{-1}B \end{vmatrix} \\ &= \det A \ \det(D - CA^{-1}B). \end{aligned}$$

For the second part, if A and C commute, then A, B, C, and D are of the same size. We show the identity by the so-called continuity argument method.

First consider the case where A is invertible. Following the above argument and using the fact that

$$\det(XY) = \det X \det Y$$

for any two square matrices X and Y of the same size, we have

$$\begin{aligned}\det M &= \det A \det(D - CA^{-1}B) \\ &= \det(AD - ACA^{-1}B) \\ &= \det(AD - CAA^{-1}B) \\ &= \det(AD - CB).\end{aligned}$$

Now assume that A is singular. Since $\det(A + \epsilon I)$ as a polynomial in ϵ has a finite number of zeros, we may choose $\delta > 0$ such that

$$\det(A + \epsilon I) \neq 0 \quad \text{whenever } 0 < \epsilon < \delta;$$

that is, $A + \epsilon I$ is invertible for all $\epsilon \in (0, \delta)$. Denote

$$M_\epsilon = \begin{pmatrix} A + \epsilon I & B \\ C & D \end{pmatrix}.$$

Noticing further that $A + \epsilon I$ and C commute, we have

$$\det M_\epsilon = \det((A + \epsilon I)D - CB) \quad \text{whenever } 0 < \epsilon < \delta.$$

Observe that both sides of the above equation are continuous functions of ϵ. Letting $\epsilon \to 0^+$ gives that

$$\det M = \det(AD - CB). \quad \blacksquare$$

Note that the identity need not be true if $AC \neq CA$.

We now turn our attention to the inverses of partitioned matrices.

Theorem 2.3 *Suppose that the partitioned matrix*

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

is invertible and that the inverse is conformally partitioned as

$$M^{-1} = \begin{pmatrix} X & Y \\ U & V \end{pmatrix},$$

where A, D, X, and V are square matrices. Then

$$\det A = \det V \det M. \tag{2.4}$$

Proof. The identity (2.4) follows immediately by taking the determinants of both sides of the matrix identity

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} I & Y \\ 0 & V \end{pmatrix} = \begin{pmatrix} A & 0 \\ C & I \end{pmatrix}. \quad \blacksquare$$

Note that the identity matrices I in the proof may have different sizes. By Theorem 2.3, A is singular if and only if V is singular.

Theorem 2.4 *Let M and M^{-1} be as defined in Theorem 2.3. If A is a nonsingular principal submatrix of M, then*

$$\begin{aligned} X &= A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}, \\ Y &= -A^{-1}B(D - CA^{-1}B)^{-1}, \\ U &= -(D - CA^{-1}B)^{-1}CA^{-1}, \\ V &= (D - CA^{-1}B)^{-1}. \end{aligned}$$

Proof. As we know from elementary linear algebra (Theorem 1.2), every invertible matrix can be written as a product of elementary matrices; so can M^{-1}. Furthermore, since

$$M^{-1}(M, I) = (I,\ M^{-1}),$$

this says that we can obtain the inverse of M by performing row operations on (M, I) to get (I, M^{-1}) (Problem 23, Section 1.2).

We now apply row operations to the augmented block matrix

$$\begin{pmatrix} A & B & I & 0 \\ C & D & 0 & I \end{pmatrix}.$$

Multiply row 1 by A^{-1} (from the left) to get

$$\begin{pmatrix} I & A^{-1}B & A^{-1} & 0 \\ C & D & 0 & I \end{pmatrix}.$$

Subtract row 1 multiplied by C from row 2 to get

$$\begin{pmatrix} I & A^{-1}B & A^{-1} & 0 \\ 0 & D - CA^{-1}B & -CA^{-1} & I \end{pmatrix}.$$

Multiply row 2 by $(D - CA^{-1}B)^{-1}$ (which exists; why?) to get

$$\begin{pmatrix} I & A^{-1}B & A^{-1} & 0 \\ 0 & I & -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}.$$

By subtracting row 2 times $A^{-1}B$ from row 1, we get the inverse of the partitioned matrix M in the form

$$\begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}.$$

Comparing to the inverse M^{-1} in the block form, we have X, Y, U, and V with the desired expressions in terms of A, B, C, and D. ■

A similar discussion for a nonsingular D implies

$$X = (A - BD^{-1}C)^{-1}.$$

It follows that

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}. \tag{2.5}$$

Below is a direct proof for (2.5). More similar identities are derived by means of partitioned matrices in Section 6.4 of Chapter 6.

Theorem 2.5 *Let $A \in \mathbb{M}_m$ and $B \in \mathbb{M}_n$ be nonsingular matrices and let C and D be $m \times n$ and $n \times m$ matrices, respectively. If the matrix $A + CBD$ is nonsingular, then*

$$(A + CBD)^{-1} = A^{-1} - A^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1}. \tag{2.6}$$

Proof. Note that $B^{-1} + DA^{-1}C$ is nonsingular, since (Problem 5)

$$\begin{aligned} \det(B^{-1} + DA^{-1}C) &= \det B^{-1} \det(I_n + BDA^{-1}C) \\ &= \det B^{-1} \det(I_m + A^{-1}CBD) \\ &= \det B^{-1} \det A^{-1} \det(A + CBD) \neq 0. \end{aligned}$$

We now prove (2.6) by a direct verification:

$$
\begin{aligned}
&(A+CBD)\Big(A^{-1}-A^{-1}C(B^{-1}+DA^{-1}C)^{-1}DA^{-1}\Big)\\
&\quad = I_m - C(B^{-1}+DA^{-1}C)^{-1}DA^{-1} + CBDA^{-1}\\
&\qquad - CBDA^{-1}C(B^{-1}+DA^{-1}C)^{-1}DA^{-1}\\
&\quad = I_m - C\Big((B^{-1}+DA^{-1}C)^{-1} - B\\
&\qquad + BDA^{-1}C(B^{-1}+DA^{-1}C)^{-1}\Big)DA^{-1}\\
&\quad = I_m - C\Big((I_n + BDA^{-1}C)(B^{-1}+DA^{-1}C)^{-1} - B\Big)DA^{-1}\\
&\quad = I_m - C\Big(B(B^{-1}+DA^{-1}C)(B^{-1}+DA^{-1}C)^{-1} - B\Big)DA^{-1}\\
&\quad = I_m - C(B-B)DA^{-1}\\
&\quad = I_m. \quad \blacksquare
\end{aligned}
$$

A great number of matrix identities involving inverses can be derived from (2.6). The following two are immediate when the involved inverses exist:

$$(A+B)^{-1} = A^{-1} - A^{-1}(B^{-1}+A^{-1})^{-1}A^{-1}$$

and

$$(A+UV^*)^{-1} = A^{-1} - A^{-1}U(I+V^*A^{-1}U)^{-1}V^*A^{-1}.$$

Problems

1. Let A be an $n \times n$ nonsingular matrix, $a \in \mathbb{C}$, α^T, $\beta \in \mathbb{C}^n$. Prove

$$(\det A)^{-1} \begin{vmatrix} a & \alpha \\ \beta & A \end{vmatrix} = a - \alpha A^{-1}\beta.$$

2. Refer to Theorem 2.2 and assume $AC = CA$. Does it follow that

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = \det(AD - BC)?$$

3. For matrices A, B, C of appropriate sizes, evaluate the determinants

$$\begin{vmatrix} A & I_n \\ I_m & 0 \end{vmatrix}, \quad \begin{vmatrix} 0 & A \\ A^{-1} & 0 \end{vmatrix}, \quad \begin{vmatrix} 0 & A \\ B & C \end{vmatrix}.$$

4. Let A, B, and C be n-square complex matrices. Show that

$$\begin{vmatrix} I_n & A \\ B & C \end{vmatrix} = \det(C - BA).$$

5. Let A and B be $m \times n$ and $n \times m$ matrices, respectively. Show that

$$\begin{vmatrix} I_n & B \\ A & I_m \end{vmatrix} = \begin{vmatrix} I_m & A \\ B & I_n \end{vmatrix}$$

and conclude that

$$\det(I_m - AB) = \det(I_n - BA).$$

Is it true that

$$\operatorname{rank}(I_m - AB) = \operatorname{rank}(I_n - BA)?$$

6. Can any two of the following expressions be identical for general complex square matrices A, B, C, D of the same size?

$$\det(AD - CB), \ \det(AD - BC), \ \det(DA - CB), \ \det(DA - BC),$$

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix}.$$

7. If A is an invertible matrix, show that

$$\operatorname{rank}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \operatorname{rank}(A) + \operatorname{rank}(D - CA^{-1}B).$$

In particular,

$$\operatorname{rank}\begin{pmatrix} I_n & I_n \\ X & Y \end{pmatrix} = n + \operatorname{rank}(X - Y).$$

8. Does it follow from the identity (2.4) that any principal submatrix (A) of a singular matrix (M) is singular?

9. Find the determinant and the inverse of the $2m \times 2m$ block matrix

$$A = \begin{pmatrix} aI_m & bI_m \\ cI_m & dI_m \end{pmatrix}, \quad ad - bc \neq 0.$$

10. If U is a unitary matrix partitioned as $U = \left(\begin{smallmatrix} u & x \\ y & U_1 \end{smallmatrix}\right)$, where $u \in \mathbb{C}$, show that $|u| = |\det U_1|$. What if U is real orthogonal?

11. Find the inverses, if they exist, for the matrices

$$\begin{pmatrix} A & I \\ I & 0 \end{pmatrix}, \quad \begin{pmatrix} I & X \\ Y & Z \end{pmatrix}.$$

12. Let A be an n-square nonsingular matrix. Write

$$A = B + iC, \quad A^{-1} = F + iG,$$

where B, C, F, G are real matrices, and set

$$D = \begin{pmatrix} B & -C \\ C & B \end{pmatrix}.$$

Show that D is nonsingular and that the inverse of D is

$$\begin{pmatrix} F & -G \\ G & F \end{pmatrix}.$$

In addition, D is normal if A is normal, and orthogonal if A is unitary.

13. Let A and C be m- and n-square invertible matrices, respectively. Show that for any $m \times n$ matrix B and $n \times m$ matrix D,

$$\det(A + BCD) = \det A \det C \det(C^{-1} + DA^{-1}B).$$

What can be deduced from this identity if $A = I_m$ and $C = I_n$?

14. Let A and B be real square matrices of the same size. Show that

$$\begin{vmatrix} A & -B \\ B & A \end{vmatrix} = |\det(A + iB)|^2.$$

15. Let A and B be complex square matrices of the same size. Show that

$$\begin{vmatrix} A & B \\ B & A \end{vmatrix} = \det(A + B)\det(A - B).$$

Also show that the eigenvalues of the 2×2 block matrix on the left-hand side consist of those of $A + B$ and $A - B$.

16. Let A and B be n-square matrices. For any integer k (positive or negative if A is invertible), find the $(1, 2)$ block of the matrix

$$\begin{pmatrix} A & B \\ 0 & I \end{pmatrix}^k.$$

17. Let B and C be complex matrices with the same number of rows, and let $A = (B, C)$. Show that

$$\begin{vmatrix} I & B^* \\ B & AA^* \end{vmatrix} = \det(CC^*)$$

and that if $C^*B = 0$ then

$$\det(A^*A) = \det(B^*B)\det(C^*C).$$

18. Let $A \in \mathbb{M}_n$. Show that there exists a diagonal matrix D with diagonal entries ± 1 such that $\det(A + D) \neq 0$. [Hint: Show by induction.]

19. If $I + A$ is nonsingular, show that $(I + A)^{-1}$ and $I - A$ commute and

$$(I + A)^{-1} + (I + A^{-1})^{-1} = I.$$

20. Show that for any $m \times n$ complex matrix A

$$(I + A^*A)^{-1}A^*A = A^*A(I + A^*A)^{-1} = I - (I + A^*A)^{-1}.$$

21. Let A and B be $m \times n$ and $n \times m$ matrices, respectively. If $I_n + BA$ is nonsingular, show that $I_m + AB$ is nonsingular and that

$$(I_n + BA)^{-1}B = B(I_m + AB)^{-1}.$$

22. Let A and B be $m \times n$ and $n \times m$ matrices, respectively. If the involved inverses exist, show that

$$(I - AB)^{-1} = I + A(I - BA)^{-1}B.$$

Conclude that if $I - AB$ is invertible, then so is $I - BA$. In particular,

$$(I + AA^*)^{-1} = I - A(I + A^*A)^{-1}A^*.$$

23. Let $A \in \mathbb{M}_n$ and $\alpha, \beta \in \mathbb{C}$. If the involved inverses exist, show that

$$(\alpha I - A)^{-1} - (\beta I - A)^{-1} = (\beta - \alpha)(\alpha I - A)^{-1}(\beta I - A)^{-1}.$$

24. Show that for any $x, y \in \mathbb{C}^n$

$$\operatorname{adj}(I - xy^*) = xy^* + (1 - y^*x)I.$$

25. Let $u, v \in \mathbb{C}^n$ with $v^*u \neq 0$. Write $v^*u = p^{-1} + q^{-1}$. Show that

$$(I - puv^*)^{-1} = I - quv^*.$$

26. Let u and v be column vectors in $\mathbb{C}^n$ such that $v^*A^{-1}u$ is not equal to -1. Show that $A + uv^*$ is invertible and, with $\delta = 1 + v^*A^{-1}u$,

$$(A + uv^*)^{-1} = A^{-1} - \delta^{-1}A^{-1}uv^*A^{-1}.$$

27. Let $M = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$. Assume that the inverses involved exist, and denote

$$S = D - CA^{-1}B, \quad T = A - BD^{-1}C.$$

Show that each of the following expressions is equal to M^{-1}.

(a) $\begin{pmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{pmatrix}$.

(b) $\begin{pmatrix} T^{-1} & -T^{-1}BD^{-1} \\ -D^{-1}CT^{-1} & D^{-1} + D^{-1}CT^{-1}BD^{-1} \end{pmatrix}$.

(c) $\begin{pmatrix} T^{-1} & -A^{-1}BS^{-1} \\ -D^{-1}CT^{-1} & S^{-1} \end{pmatrix}$.

(d) $\begin{pmatrix} T^{-1} & (C - DB^{-1}A)^{-1} \\ (B - AC^{-1}D)^{-1} & S^{-1} \end{pmatrix}$.

(e) $\begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} \begin{pmatrix} A^{-1} & 0 \\ 0 & S^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -CA^{-1} & I \end{pmatrix}$.

(f) $\begin{pmatrix} A^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} A^{-1}B \\ -I \end{pmatrix} S^{-1}(CA^{-1}, -I)$.

28. Deduce the following inverse identities from the previous problem.

$$\begin{aligned} (A - BD^{-1}C)^{-1} &= A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} \\ &= -C^{-1}D(B - AC^{-1}D)^{-1} \\ &= -(C - DB^{-1}A)^{-1}DB^{-1} \\ &= C^{-1}D(D - CA^{-1}B)^{-1}CA^{-1} \\ &= A^{-1}B(D - CA^{-1}B)^{-1}DB^{-1}. \end{aligned}$$

2.3 The Rank of Product and Sum

This section is concerned with the ranks of the product AB and the sum $A+B$ in terms of the ranks of matrices A and B.

Matrix rank is one of the most important concepts. In the previous chapter we defined the rank of a matrix A to be the nonnegative number r in the matrix $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$, which is obtained through elementary operations on A. The rank of matrix A, denoted by $\operatorname{rank}(A)$, is uniquely determined by A. $\operatorname{rank}(A) = 0$ if and only if $A = 0$.

The rank of a matrix can be defined in many different but equivalent ways (see Problem 24 of Section 1.2). For instance, it can be defined by *row rank* or *column rank*. The row (column) rank of a matrix A is the dimension of the vector space spanned by the rows (columns) of A. The row rank and column rank of a matrix are equal. Here is why: let the row rank of an $m \times n$ matrix A be r and the column rank be c. We show $r = c$. Suppose that columns $C_1, C_2, \ldots, C_c$ of A are linearly independent and span the column space of A. For the jth column A_j of A, $j = 1, 2, \ldots, n$, we can write

$$A_j = d_{1j}C_1 + d_{2j}C_2 + \cdots + d_{cj}C_c = Cd_j,$$

where $C = (C_1, C_2, \ldots, C_c)$ and $d_j = (d_{1j}, d_{2j}, \ldots, d_{cj})^T$.

Let $D = (d_{ij}) = (d_1, d_2, \ldots, d_n)$. Then $A = CD$ and D is $c \times n$. It follows that every row of A is a linear combination of the rows of D. Thus, $r \le c$. A similar argument on A^T shows $c \le r$.

We can also see that row rank equals column rank through the relation that the dimension of the column space, i.e., $\operatorname{Im} A$, is the rank of A. Recall that the kernel, or the null space, and the image of an $m \times n$ matrix A, viewed as a linear transformation, are, respectively,

$$\operatorname{Ker} A = \{x \in \mathbb{C}^n : \ Ax = 0\}, \quad \operatorname{Im} A = \{Ax : \ x \in \mathbb{C}^n\}.$$

Denote the row and column ranks of matrix X by $rr(X)$ and $cr(X)$, respectively. Then $cr(X) = \dim \operatorname{Im} X$. Since $Ax = 0 \Leftrightarrow A^*Ax = 0$,

$$cr(A) = cr(A^*A) = \dim \operatorname{Im}(A^*A) \le \dim \operatorname{Im}(A^*) = cr(A^*).$$

Hence $cr(A^*) \leq cr((A^*)^*) = cr(A)$. So $cr(A) = cr(A^*)$. However, $cr(A) = rr(A^T)$; we have $cr(A) = cr(A^*) = rr(\bar{A}) = rr(A)$ (over $\mathbb{C}$).

Theorem 2.6 (Sylvester) *Let A and B be complex matrices of sizes $m \times n$ and $n \times p$, respectively. Then*

$$\operatorname{rank}(AB) = \operatorname{rank}(B) - \dim(\operatorname{Im} B \cap \operatorname{Ker} A). \tag{2.7}$$

Consequently,

$$\operatorname{rank}(A) + \operatorname{rank}(B) - n \leq \operatorname{rank}(AB) \leq \min\{\operatorname{rank}(A), \operatorname{rank}(B)\}.$$

Proof. Recall from Theorem 1.5 that if $\mathcal{A}$ is a linear transformation on an n-dimensional vector space, then

$$\dim \operatorname{Im}(\mathcal{A}) + \dim \operatorname{Ker}(\mathcal{A}) = n.$$

Viewing A as a linear transformation on $\mathbb{C}^n$, we have (Problem 1)

$$\operatorname{rank}(A) = \dim \operatorname{Im} A.$$

For the rank of AB, we think of A as a linear transformation on the vector space $\operatorname{Im} B$. Then its image is $\operatorname{Im}(AB)$ and its null space is $\operatorname{Im} B \cap \operatorname{Ker} A$. We thus have

$$\dim \operatorname{Im}(AB) + \dim(\operatorname{Im} B \cap \operatorname{Ker} A) = \dim \operatorname{Im} B.$$

The identity (2.7) then follows.

For the inequalities, the second one is immediate from (2.7), and the first one is due to the fact that

$$\dim \operatorname{Im} A + \dim(\operatorname{Im} B \cap \operatorname{Ker} A) \leq n. \quad \blacksquare$$

For the product of three matrices, we have

$$\operatorname{rank}(ABC) \geq \operatorname{rank}(AB) + \operatorname{rank}(BC) - \operatorname{rank}(B). \tag{2.8}$$

A pure matrix proof of (2.8) goes as follows. Note that

$$\operatorname{rank}\begin{pmatrix} 0 & X \\ Y & Z \end{pmatrix} \geq \operatorname{rank}(X) + \operatorname{rank}(Y)$$

for any matrix Z of appropriate size, and that equality holds if $Z = 0$. The inequality (2.8) then follows from the matrix identity

$$\begin{pmatrix} I & -A \\ 0 & I \end{pmatrix} \begin{pmatrix} 0 & AB \\ BC & B \end{pmatrix} \begin{pmatrix} I & 0 \\ -C & I \end{pmatrix} = \begin{pmatrix} -ABC & 0 \\ 0 & B \end{pmatrix}.$$

Theorem 2.7 *Let A and B be $m \times n$ matrices and denote by C and D, respectively, the partitioned matrices*

$$C = (I_m,\ I_m), \quad D = \begin{pmatrix} A \\ B \end{pmatrix}.$$

Then

$$\begin{aligned} \operatorname{rank}(A+B) &= \operatorname{rank}(A) + \operatorname{rank}(B) - \dim(\operatorname{Ker} C \cap \operatorname{Im} D) \\ &\quad - \dim(\operatorname{Im} A^* \cap \operatorname{Im} B^*). \end{aligned} \tag{2.9}$$

In particular,

$$\operatorname{rank}(A+B) \leq \operatorname{rank}(A) + \operatorname{rank}(B).$$

Proof. Write

$$A + B = (I_m, I_m) \begin{pmatrix} A \\ B \end{pmatrix} = CD.$$

Utilizing the previous theorem, we have

$$\operatorname{rank}(A+B) = \operatorname{rank}(D) - \dim(\operatorname{Im} D \cap \operatorname{Ker} C). \tag{2.10}$$

However,

$$\begin{aligned} \operatorname{rank}(D) &= \operatorname{rank}(D^*) = \operatorname{rank}(A^*,\ B^*) \\ &= \dim \operatorname{Im}(A^*,\ B^*) \quad \text{(by Problem 9)} \\ &= \dim(\operatorname{Im} A^* + \operatorname{Im} B^*) \\ &= \dim \operatorname{Im} A^* + \dim \operatorname{Im} B^* \\ &\quad - \dim(\operatorname{Im} A^* \cap \operatorname{Im} B^*) \\ &= \operatorname{rank}(A^*) + \operatorname{rank}(B^*) - \dim(\operatorname{Im} A^* \cap \operatorname{Im} B^*) \\ &= \operatorname{rank}(A) + \operatorname{rank}(B) - \dim(\operatorname{Im} A^* \cap \operatorname{Im} B^*). \end{aligned}$$

Substituting this into (2.10) reveals (2.9). ∎

Problems

1. Show that $\operatorname{rank}(A) = \dim \operatorname{Im} A$ for any complex matrix A.

2. Let $A, B \in \mathbb{M}_n$. If $\operatorname{rank}(A) < \frac{n}{2}$ and $\operatorname{rank}(B) < \frac{n}{2}$, show that $\det(A + \lambda B) = 0$ for some complex number λ.

3. If $B \in \mathbb{M}_n$ is invertible, show that $\operatorname{rank}(AB) = \operatorname{rank}(A)$ for every $m \times n$ matrix A. Is the converse true?

4. Is it true that the sum of two singular matrices is singular? How about the product?

5. Let A be an $m \times n$ matrix. Show that if $\operatorname{rank}(A) = m$, then there is an $n \times m$ matrix B such that $AB = I_m$, and that if $\operatorname{rank}(A) = n$, then there is an $m \times n$ matrix B such that $BA = I_n$.

6. Let A be an $m \times n$ matrix. Show that for any $s \times m$ matrix X with columns linearly independent and any $n \times t$ matrix Y with rows linearly independent,
$$\operatorname{rank}(A) = \operatorname{rank}(XA) = \operatorname{rank}(AY) = \operatorname{rank}(XAY).$$

7. For matrices A and B, show that if $\operatorname{rank}(AB) = \operatorname{rank}(B)$, then
$$ABX = ABY \quad \Leftrightarrow \quad BX = BY,$$
and if $\operatorname{rank}(AB) = \operatorname{rank}(A)$, then
$$XAB = YAB \quad \Leftrightarrow \quad XA = YA.$$

8. Let A be an $m \times n$ matrix over a field $\mathbb{F}$, $\operatorname{rank}(A) = r$. Show that for any positive integer k, $r \le k \le n$, there exists an $n \times n$ matrix B over $\mathbb{F}$ such that $AB = 0$ and $\operatorname{rank}(A) + \operatorname{rank}(B) = k$,

9. For any matrices A and B of the same size, show that
$$\operatorname{Im}(A, B) = \operatorname{Im} A + \operatorname{Im} B.$$

10. Let A be an $m \times n$ matrix. Show that for any $n \times m$ matrix B,
$$\dim \operatorname{Im} A + \dim \operatorname{Ker} A = \dim \operatorname{Im}(BA) + \dim \operatorname{Ker}(BA).$$

11. Let A and B be $m \times n$ and $n \times m$ matrices, respectively. Show that
$$\det(AB) = 0 \quad \text{if } m > n.$$

12. For matrices A and B of the same size, show that

$$|\operatorname{rank}(A) - \operatorname{rank}(B)| \le \operatorname{rank}(A \pm B).$$

13. Let A and B be n-square complex matrices. Show that

$$\operatorname{rank}(AB - I) \le \operatorname{rank}(A - I) + \operatorname{rank}(B - I).$$

14. Let $A \in \mathbb{M}_n$. If $A^2 = I$, show that

$$\operatorname{rank}(A + I) + \operatorname{rank}(A - I) = n.$$

15. Show that if $A \in \mathbb{M}_n$ and $A^2 = A$, then $\operatorname{rank}(A) + \operatorname{rank}(I_n - A) = n$.

16. Let $A, B \in \mathbb{M}_n$. Show that

    (a) $\operatorname{rank}(A - ABA) = \operatorname{rank}(A) + \operatorname{rank}(I_n - BA) - n$.

    (b) If $A + B = I_n$ and $\operatorname{rank}(A) + \operatorname{rank}(B) = n$, then $A^2 = A$, $B^2 = B$, and $AB = 0 = BA$.

17. If A is an n-square matrix with rank r, show that there exist an n-square matrix B of rank $n - r$ such that $AB = 0$.

18. Let A and B be $n \times n$ matrices over a field $\mathbb{F}$ having null spaces W_1 and W_2, respectively. (i). If $AB = 0$, show that $\dim W_1 + \dim W_2 \ge n$. (ii). Show that $W_1 = W_2$ if and only if $A = PB$ and $B = QA$ for some $n \times n$ matrices P and Q.

19. Let A be an $m \times n$ matrix with rank n. If $m > n$, show that there is a matrix B of size $(m - n) \times m$ and a matrix C of size $m \times (m - n)$, both of rank $m - n$, such that $BA = 0$ and (A, C) is nonsingular.

20. Let $\mathcal{A}$ be a linear transformation on a finite-dimensional vector space. Show that

$$\operatorname{Ker}(\mathcal{A}) \subseteq \operatorname{Ker}(\mathcal{A}^2) \subseteq \operatorname{Ker}(\mathcal{A}^3) \subseteq \cdots$$

and that

$$\operatorname{Im}(\mathcal{A}) \supseteq \operatorname{Im}(\mathcal{A}^2) \supseteq \operatorname{Im}(\mathcal{A}^3) \supseteq \cdots.$$

Further show that there are finite proper inclusions in each chain.

21. If A is an $m \times n$ complex matrix, $\operatorname{Im}(A)$ is in fact the space spanned by the column vectors of A, called the *column space* of A and denoted by $\mathcal{C}(A)$. Similarly, the row vectors of A span the *row space*, symbolized by $\mathcal{R}(A)$. Let A and B be two matrices. Show that

$$\mathcal{C}(A) \subseteq \mathcal{C}(B) \quad \Leftrightarrow \quad A = BC,$$

for some matrix C, and

$$\mathcal{R}(A) \subseteq \mathcal{R}(B) \quad \Leftrightarrow \quad A = RB,$$

for some matrix R. [Note: $\mathcal{C}(A) = \operatorname{Im} A$ and $\mathcal{R}(A) = \operatorname{Im} A^T$.]

22. Let A and B be $m \times n$ matrices. Show that

$$\mathcal{C}(A+B) \subseteq \mathcal{C}(A) + \mathcal{C}(B)$$

and that the following statements are equivalent.

(a) $\mathcal{C}(A) \subseteq \mathcal{C}(A+B)$.
(b) $\mathcal{C}(B) \subseteq \mathcal{C}(A+B)$.
(c) $\mathcal{C}(A+B) = \mathcal{C}(A) + \mathcal{C}(B)$.

23. Prove or disprove, for any n-square matrices A and B, that

$$\operatorname{rank} \begin{pmatrix} A \\ B \end{pmatrix} = \operatorname{rank}(A, B).$$

24. Let A be an $m \times n$ matrix and B be a $p \times n$ matrix. Show that

$$\operatorname{Ker} A \cap \operatorname{Ker} B = \operatorname{Ker} \begin{pmatrix} A \\ B \end{pmatrix}.$$

25. Let A and B be matrices of the same size. Show the rank inequalities

$$\operatorname{rank}(A+B) \le \operatorname{rank} \begin{pmatrix} A \\ B \end{pmatrix} \le \operatorname{rank}(A) + \operatorname{rank}(B)$$

and

$$\operatorname{rank}(A+B) \le \operatorname{rank}(A, B) \le \operatorname{rank}(A) + \operatorname{rank}(B)$$

by writing

$$A + B = (A, B) \begin{pmatrix} I \\ I \end{pmatrix} = (I, I) \begin{pmatrix} A \\ B \end{pmatrix}.$$

Additionally, show that

$$\operatorname{rank} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \le \operatorname{rank}(A) + \operatorname{rank}(B) + \operatorname{rank}(C) + \operatorname{rank}(D).$$

26. Let A, B, and C be complex matrices of the same size. Show that

$$\operatorname{rank}(A, B, C) \le \operatorname{rank}(A, B) + \operatorname{rank}(B, C) - \operatorname{rank}(B).$$

2.4 The Eigenvalues of AB and BA

For square matrices A and B of the same size, the product matrices AB and BA need not be equal, or even similar. For instance, if

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix},$$

then

$$AB = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{but} \quad BA = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}.$$

Note that in this example both AB $(= 0)$ and BA $(\neq 0)$ have only repeated eigenvalue 0 (twice, referred to as *multiplicity* of 0). Is this a coincidence, or can we construct an example such that AB has only zero eigenvalues but BA has some nonzero eigenvalues?

The following theorem gives a negative answer to the question. This is a very important result in matrix theory.

Theorem 2.8 *Let A and B be $m \times n$ and $n \times m$ complex matrices, respectively. Then AB and BA have the same nonzero eigenvalues, counting multiplicity. If $m = n$, A and B have the same eigenvalues.*

Proof 1. Use determinants. Notice that

$$\begin{pmatrix} I_m & -A \\ 0 & \lambda I_n \end{pmatrix} \begin{pmatrix} \lambda I_m & A \\ B & I_n \end{pmatrix} = \begin{pmatrix} \lambda I_m - AB & 0 \\ \lambda B & \lambda I_n \end{pmatrix}$$

and that

$$\begin{pmatrix} I_m & 0 \\ -B & \lambda I_n \end{pmatrix} \begin{pmatrix} \lambda I_m & A \\ B & I_n \end{pmatrix} = \begin{pmatrix} \lambda I_m & A \\ 0 & \lambda I_n - BA \end{pmatrix}.$$

By taking determinants and equating the right-hand sides, we obtain

$$\lambda^n \det(\lambda I_m - AB) = \lambda^m \det(\lambda I_n - BA). \tag{2.11}$$

Thus, $\det(\lambda I_m - AB) = 0$ if and only if $\det(\lambda I_n - BA) = 0$ when $\lambda \neq 0$. It is immediate that AB and BA have the same nonzero eigenvalues, including multiplicity (by factorization).

Proof 2. Use matrix similarity. Consider the block matrix

$$\begin{pmatrix} 0 & 0 \\ B & 0 \end{pmatrix}.$$

Add the second row multiplied by A from the left to the first row:

$$\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix}.$$

Do the similar operation for columns to get

$$\begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}.$$

Write, in equation form,

$$\begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix}\begin{pmatrix} 0 & 0 \\ B & 0 \end{pmatrix} = \begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix}$$

and

$$\begin{pmatrix} 0 & 0 \\ B & 0 \end{pmatrix}\begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}.$$

It follows that

$$\begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix}^{-1}\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix}\begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix};$$

that is, matrices

$$\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$$

are similar. Thus, matrices AB and BA have the same nonzero eigenvalues, counting multiplicity. (Are AB and BA similar?)

Proof 3. Use the continuity argument. We first deal with the case where $m = n$. If A is nonsingular, then

$$BA = A^{-1}(AB)A.$$

Thus, AB and BA are similar and have the same eigenvalues.

If A is singular, let δ be such a positive number that $\epsilon I + A$ is nonsingular for every ϵ, $0 < \epsilon < \delta$. Then

$$(\epsilon I + A)B \quad \text{and} \quad B(\epsilon I + A)$$

are similar and have the same characteristic polynomials. Therefore,

$$\det\left(\lambda I_n - (\epsilon I_n + A)B\right) = \det\left(\lambda I_n - B(\epsilon I_n + A)\right), \quad 0 < \epsilon < \delta.$$

Since both sides are continuous functions of ϵ, letting $\epsilon \to 0$ gives

$$\det(\lambda I_n - AB) = \det(\lambda I_n - BA).$$

Thus, AB and BA have the same eigenvalues.

For the case where $m \neq n$, assume $m < n$. Augment A and B by zero rows and zero columns, respectively, so that

$$A_1 = \begin{pmatrix} A \\ 0 \end{pmatrix}, \quad B_1 = (B, 0)$$

are n-square matrices. Then

$$A_1B_1 = \begin{pmatrix} AB & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad B_1A_1 = BA.$$

It follows that A_1B_1 and B_1A_1, consequently AB and BA, have the same nonzero eigenvalues, counting multiplicity.

Proof 4. Treat matrices as operators. It must be shown that if $\lambda I_m - AB$ is singular, so is $\lambda I_n - BA$, and vice versa. It may be assumed that $\lambda = 1$, by multiplying $\frac{1}{\lambda}$ otherwise.

If $I_m - AB$ is invertible, let $X = (I_m - AB)^{-1}$. We compute

$$\begin{aligned}(I_n - BA)(I_n + BXA) &= I_n + BXA - BA - BABXA \\ &= I_n + (BXA - BABXA) - BA \\ &= I_n + B(I_m - AB)XA - BA \\ &= I_n + BA - BA \\ &= I_n.\end{aligned}$$

Thus, $I_n - BA$ is invertible. Note that this approach gives no information on the multiplicity of the nonzero eigenvalues. ∎

As a side product, using (2.11), we have (see also Problem 5 of Section 2.2), for any $m \times n$ matrix A and $n \times m$ matrix B,

$$\det(I_m + AB) = \det(I_n + BA).$$

Note that $I_m + AB$ is invertible if and only if $I_n + BA$ is invertible.

Problems

1. Show that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for any $m \times n$ matrix A and $n \times m$ matrix B. In particular, $\operatorname{tr}(A^*A) = \operatorname{tr}(AA^*)$.

2. For any square matrices A and B of the same size, show that

$$\operatorname{tr}(A+B)^2 = \operatorname{tr} A^2 + 2\operatorname{tr}(AB) + \operatorname{tr} B^2.$$

Does it follow that $(A+B)^2 = A^2 + 2AB + B^2$?

3. Let A and B be square matrices of the same size. Show that

$$\det(AB) = \det(BA).$$

Does this hold if A and B are not square? Is it true that

$$\operatorname{rank}(AB) = \operatorname{rank}(BA)?$$

4. Let A and B be $m \times n$ and $n \times m$ complex matrices, respectively, with $m < n$. If the eigenvalues of AB are $\lambda_1, \ldots, \lambda_m$, what are the eigenvalues of BA?

5. Let A and B be $n \times n$ matrices. Show that for every integer $k \geq 1$,

$$\operatorname{tr}(AB)^k = \operatorname{tr}(BA)^k.$$

Does

$$\operatorname{tr}(AB)^k = \operatorname{tr}(A^k B^k)?$$

6. Show that for any $x, y \in \mathbb{C}^n$

$$\det(I_n + xy^*) = 1 + y^*x.$$

7. Compute the determinant

$$\begin{vmatrix} 1 + x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & 1 + x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \vdots & \vdots \\ x_ny_1 & x_ny_2 & \cdots & 1 + x_ny_n \end{vmatrix}.$$

8. If A, B, and C are three complex matrices of appropriate sizes, show that ABC, CAB, and BCA have the same nonzero eigenvalues. Is it true that ABC and CBA have the same nonzero eigenvalues?

9. Do A^* and A have the same nonzero eigenvalues? How about A^*A and AA^*? Show by example that $\det(AA^*) \neq \det(A^*A)$ in general.

10. Let $A, B \in \mathbb{M}_n$. Show that $\begin{pmatrix} 0 & A \\ B & 0 \end{pmatrix}$ is similar to $\begin{pmatrix} 0 & B \\ A & 0 \end{pmatrix}$ via $\begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$.

11. For any square matrices A and B of the same size, show that

$$A^2 + B^2 \quad \text{and} \quad \begin{pmatrix} A^2 & AB \\ BA & B^2 \end{pmatrix}$$

have the same nonzero eigenvalues. Further show that the latter block matrix must have zero eigenvalues. How many of them?

12. Let $A \in \mathbb{M}_n$. Find the eigenvalues of $\begin{pmatrix} A & A \\ A & A \end{pmatrix}$ in terms of those of A.

13. Let A be a 3×2 matrix and B be a 2×3 matrix such that

$$AB = \begin{pmatrix} 8 & 2 & -2 \\ 2 & 5 & 4 \\ -2 & 4 & 5 \end{pmatrix}.$$

Find the ranks of AB and $(AB)^2$ and show that $BA = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}$.

14. Let A be an $m \times n$ complex matrix and $M = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}$. Show that

   (a) M is a Hermitian matrix.

   (b) The eigenvalues of M are

$$-\sigma_1, \ldots, -\sigma_r, \overbrace{0, \ldots, 0}^{m+n-2r}, \sigma_r, \ldots, \sigma_1,$$

   where $\sigma_1 \geq \cdots \geq \sigma_r$ are the positive square roots of the nonzero eigenvalues of A^*A, called *singular values* of A.

   (c) $\det M = \det(-A^*A) = (-1)^n |\det A|^2$ if A is n-square.

   (d) $2\begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} = \begin{pmatrix} B & A \\ A^* & C \end{pmatrix} + \begin{pmatrix} -B & A \\ A^* & C \end{pmatrix}$ for any matrices B, C.

15. Let A and B be matrices of sizes $m \times n$ and $n \times m$, respectively. Do AB and BA have the same nonzero singular values?

. ———— ⊙ ———— .

2.5 The Continuity Argument and Matrix Functions

One of the most frequently used techniques in matrix theory is the *continuity argument.* A good example of this is to show, as we saw in the previous section, that matrices AB and BA have the same set of eigenvalues when A and B are both square matrices of the same size. It goes as follows. First consider the case where A is invertible and conclude that AB and BA are similar due to the fact that

$$AB = A(BA)A^{-1}.$$

If A is singular, consider $A + \epsilon I$. Choose $\delta > 0$ such that $A + \epsilon I$ is invertible for all ϵ, $0 < \epsilon < \delta$. Thus, $(A + \epsilon I)B$ and $B(A + \epsilon I)$ have the same set of eigenvalues for every $\epsilon \in (0, \delta)$.

Equate the characteristic polynomials to get

$$\det(\lambda I - (A + \epsilon I)B) = \det(\lambda I - B(A + \epsilon I)), \quad 0 < \epsilon < \delta.$$

Since both sides are continuous functions of ϵ, letting $\epsilon \to 0^+$ gives

$$\det(\lambda I - AB) = \det(\lambda I - BA).$$

Thus, AB and BA have the same eigenvalues.

The proof was done in three steps:

1. Show that the assertion is true for the nonsingular A.
2. Replace singular A by nonsingular $A + \epsilon I$.
3. Use continuity of a function in ϵ to get the desired conclusion.

We have used and will more frequently use the following theorem.

Theorem 2.9 *Let A be an $n \times n$ matrix. If A is singular, then there exists a $\delta > 0$ such that $A + \epsilon I$ is nonsingular for all $\epsilon \in (0, \delta)$.*

Proof. The polynomial $\det(\lambda I + A)$ in λ has at most n zeros. If they are all 0, we can take δ to be any positive number. Otherwise, let δ be the smallest nonzero λ in modulus. This δ serves the purpose. ∎

A continuity argument is certainly an effective way for many matrix problems when a singular matrix is involved. The setting in which the technique is used is rather important. Sometimes the result for nonsingular matrices may be invalid for the singular case. Here is an example for which the continuity argument fails.

Theorem 2.10 *Let C and D be n-square matrices such that*

$$CD^T + DC^T = 0.$$

If D is nonsingular, then for any n-square matrices A and B

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = \det(AD^T + BC^T).$$

The identity is invalid in general if D is singular.

Proof. It is easy to verify that

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} D^T & 0 \\ C^T & I \end{pmatrix} = \begin{pmatrix} AD^T + BC^T & B \\ 0 & D \end{pmatrix}.$$

Taking determinants of both sides results in the desired identity.

For an example of the singular case, we take A, B, C, and D to be, respectively,

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

where D is singular. It is easy to see by a simple computation that the determinant identity does not hold. ∎

The continuity argument may be applied to more general functions of matrices. For instance, the trace and determinant depend continuously on the entries of a matrix. These are easy to see as the trace is the sum of the main diagonal entries and the determinant is the sum of all products of (different) diagonal entries. So we may simply say that the trace and determinant are continuous functions of (the entries of) the matrix.

We have used the term *matrix function*. What is a matrix function after all? A *matrix function*, $f(A)$, or *function of a matrix* can have several different meanings. It can be an operation on a matrix producing a scalar, such as $\operatorname{tr} A$ and $\det A$; it can be a mapping from a matrix space to a matrix space, like $f(A) = A^2$; it can also be entrywise operations on the matrix, for instance, $g(A) = (a_{ij}^2)$. In this book we use the term matrix function in a general (loose) sense; that is, a matrix function is a mapping $f : A \mapsto f(A)$ as long as $f(A)$ is well defined, where $f(A)$ is a scalar or a matrix (or a vector).

Given a square matrix A, the square of A, A^2, is well defined. How about a *square root* of A? Take $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, for example. There is no matrix B such that $B^2 = A$. After a moment's consideration, one may realize that this thing is nontrivial. In fact, generalizing a function $f(z)$ of a scalar variable $z \in \mathbb{C}$ to a matrix function $f(A)$ is a serious business and it takes great effort.

Most of the terminology in calculus can be defined for square matrices. For instance, a *matrix sequence* (or *series*) is convergent if it is convergent entrywise. As an example,

$$\begin{pmatrix} \frac{1}{k} & \frac{k-1}{k} \\ 0 & \frac{1}{k} \end{pmatrix} \to \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \text{as } k \to \infty.$$

For differentiation and integration, let $A(t) = (a_{ij}(t))$ and denote

$$\frac{d}{dt}(A(t)) = \left(\frac{d}{dt}a_{ij}(t)\right), \quad \int A(t)dt = \left(\int a_{ij}(t)dt\right).$$

That is, by differentiating or integrating a matrix we mean to perform the operation on the matrix entrywise. It can be shown that the product rule for derivatives in calculus holds for matrices whereas the power rule does not. Now one is off to a good start working on matrix calculus, which is useful for differential equations. Interested readers may pursue and explore more in this direction.

Problems

1. Why did the continuity argument fail Theorem 2.10?

2. Let C and D be real matrices such that $CD^T + DC^T = 0$. Show that if C is *skew-symmetric* (i.e., $C^T = -C$), then so is DC.

3. Show that A has no square root. How about B and C, where

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}?$$

4. Let $A = \left(\begin{smallmatrix} 2 & 1 \\ 2 & 3 \end{smallmatrix}\right)$. Find a 2×2 matrix X so that $X^2 = A$.

5. Use a continuity argument to show that for any $A, B \in \mathbb{M}_n$
$$\operatorname{adj}(AB) = \operatorname{adj}(B)\operatorname{adj}(A).$$

6. Show that $A_\epsilon = P_\epsilon J_\epsilon P_\epsilon^{-1}$ if $\epsilon \neq 0$, where
$$A_\epsilon = \begin{pmatrix} \epsilon & 0 \\ 1 & 0 \end{pmatrix}, \quad P_\epsilon = \begin{pmatrix} 0 & \epsilon \\ 1 & 1 \end{pmatrix}, \quad J_\epsilon = \begin{pmatrix} 0 & 0 \\ 0 & \epsilon \end{pmatrix}.$$
What happens to the matrix identity if $\epsilon \to 0$? Is A_0 similar to J_0?

7. Explain why $\operatorname{rank}(A^2) \leq \operatorname{rank}(A)$. Discuss whether a continuity argument can be used to show the inequality.

8. Show that the eigenvalues of A are independent of ϵ, where
$$A = \begin{pmatrix} \epsilon - 1 & -1 \\ \epsilon^2 - \epsilon + 1 & -\epsilon \end{pmatrix}.$$

9. Denote by $\sigma_{\max}$ and $\sigma_{\min}$, $\sigma_{\max} \geq \sigma_{\min}$, the singular values of matrix $A = \left(\begin{smallmatrix} 1 & \epsilon \\ \epsilon & 1 \end{smallmatrix}\right)$, $\epsilon > 0$. Show that $\lim_{\epsilon \to 1^-} \sigma_{\max}/\sigma_{\min} = +\infty$.

10. Let A be a nonsingular matrix with $A^{-1} = B = (b_{ij})$. Show that b_{ij} are continuous functions of a_{ij}, the entries of A, and that if $\lim_{t \to 0} A(t) = A$ and $\det A \neq 0$ (this condition is necessary), then
$$\lim_{t \to 0} \big(A(t)\big)^{-1} = A^{-1}.$$
Conclude that
$$\lim_{\lambda \to 0} (A - \lambda I)^{-1} = A^{-1}$$
and for any $m \times n$ matrix X and $n \times m$ matrix Y, independent of ϵ,
$$\lim_{\epsilon \to 0} \begin{pmatrix} I_m & \epsilon X \\ \epsilon Y & I_n \end{pmatrix}^{-1} = I_{m+n}.$$

11. Let $A \in \mathbb{M}_n$. If $|\lambda| < 1$ for all eigenvalues λ of A, show that
$$(I - A)^{-1} = \sum_{k=1}^{\infty} A^k = I + A + A^2 + A^3 + \cdots.$$

12. Let $A = \left(\begin{smallmatrix} -1 & 1 \\ 0 & -1 \end{smallmatrix}\right)$. Show that $\sum_{k=1}^{\infty} \frac{1}{k^2} A^k$ is convergent.

13. Let $p(x)$, $q(x)$ be polynomials and $A \in \mathbb{M}_n$ be such that $q(A)$ is invertible. Show that $p(A)(q(A))^{-1} = (q(A))^{-1}p(A)$. Conclude that $(I-A)^{-1}(I+A^2) = (I+A^2)(I-A)^{-1}$ when A has no eigenvalue 1.

14. Let n be a positive number and x be a real number. Let

$$A = \begin{pmatrix} 1 & -\frac{x}{n} \\ \frac{x}{n} & 1 \end{pmatrix}.$$

Show that

$$\lim_{x\to 0}\left(\lim_{n\to\infty} \frac{1}{x}(I - A^n)\right) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

[Hint: $A = cP$ for some constant c and orthogonal matrix P.]

15. For any square matrix X, show that $e^X = \sum_{k=0}^{\infty} \frac{1}{k!}X^k$ is well defined; that is, the series always converges. Let A, $B \in \mathbb{M}_n$. Show that

    (a) If $A = 0$, then $e^A = I$.

    (b) If $A = I$, then $e^A = eI$.

    (c) If $AB = BA$, then $e^{A+B} = e^A e^B = e^B e^A$.

    (d) If A is invertible, then $e^{-A} = (e^A)^{-1}$.

    (e) If A is invertible, then $e^{ABA^{-1}} = Ae^BA^{-1}$.

    (f) If λ is an eigenvalue of A, then e^λ is an eigenvalue of e^A.

    (g) $\det e^A = e^{\operatorname{tr} A}$.

    (h) $(e^A)^* = e^{A^*}$.

    (i) If A is Hermitian, then e^{iA} is unitary.

    (j) If A is real skew-symmetric, then e^A is (real) orthogonal.

16. Let $A \in \mathbb{M}_n$ and $t \in \mathbb{R}$. Show that $\frac{d}{dt}e^{tA} = Ae^{tA} = e^{tA}A$.

17. Let $A(t) = \begin{pmatrix} e^{2t} & t \\ 1+t & \sin t \end{pmatrix}$, where $t \in \mathbb{R}$. Find $\int_0^1 A(t)dt$ and $\frac{d}{dt}A(t)$.

2.6 Localization of Eigenvalues: The Geršgorin Theorem

Is there a way to locate in the complex plane the eigenvalues of a matrix? The Geršgorin theorem is a celebrated result on this; it ensures that the eigenvalues of a matrix lie in certain discs in the complex plane centered at the diagonal entries of the matrix.

Before proceeding, we note that the eigenvalues of a matrix are continuous functions of the matrix entries. To see this, as an example, we examine the 2×2 case for

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

A computation gives the eigenvalues of A

$$\lambda = \frac{1}{2}\left[a_{11} + a_{22} \pm \sqrt{(a_{11} - a_{22})^2 + 4a_{21}a_{12}}\,\right],$$

which are obviously continuous functions of the entries of A.

In general, the eigenvalues of a matrix depend continuously on the entries of the matrix. This follows from the continuous dependence of the zeros of a polynomial on its coefficients, which invokes the theory of polynomials with real or complex coefficients. Simply put: small changes in the coefficients of a polynomial can lead only to small changes in any zero. As a result, the eigenvalues of a (real or complex) matrix depend continuously upon the entries of the matrix. The idea of the proof of this goes as follows. Obviously, the coefficients of the characteristic polynomial depend continuously on the entries of the matrix. It remains to show that the roots of a polynomial depend continuously on the coefficients. Consider, without loss of generality, the zero root case.

Let $p(x) = x^n + a_1 x^{n-1} + \cdots + a_n$ with $p(0) = 0$. Then $a_n = 0$. For any positive number ϵ, if $q(x) = x^n + b_1 x^{n-1} + \cdots + b_n$ is a polynomial such that $|b_i - a_i| < \epsilon$, $i = 1, \ldots, n$, then the roots $x_1, \ldots, x_n$ of $q(x)$ satisfy $|x_1 \cdots x_n| = |b_n| < \epsilon$. It follows that $|x_i| < \sqrt[n]{\epsilon}$ for some i. This means that $q(x)$ has an eigenvalue "close" to 0.

Theorem 2.11 *The eigenvalues of a matrix are continuous functions of the entries of the matrix.*

Because singular values are the square roots of the eigenvalues of certain matrices, singular values are also continuous functions of the entries of the matrix. It is readily seen that determinant and trace are continuous functions of the entries of the matrix too.

Theorem 2.12 (Geršgorin) *Let $A = (a_{ij}) \in \mathbb{M}_n$ and let*

$$r_i = \sum_{j=1,\, j\neq i}^{n} |a_{ij}|, \quad i = 1, 2, \ldots, n.$$

Then all the eigenvalues of A lie in the union of n closed discs

$$\cup_{i=1}^{n} \{z \in \mathbb{C} : |z - a_{ii}| \leq r_i\}.$$

Furthermore, if a union of k of these n discs forms a connected region that is disjoint from the remaining $n-k$ discs, then there are exactly k eigenvalues of A in this region (counting algebraic multiplicities).

Let us see an example before proving the theorem. For

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0.25 & 2 & 0.25 \\ 0.25 & 0 & 3 \end{pmatrix},$$

there are three Geršgorin discs:

$$G_1 = \{z \in \mathbb{C} : \ |z - 1| \leq 1\},$$

$$G_2 = \{z \in \mathbb{C} : \ |z - 2| \leq 0.5\},$$

$$G_3 = \{z \in \mathbb{C} : \ |z - 3| \leq 0.25\}.$$

The first two discs are connected and disjoint with G_3. Thus there are two eigenvalues of A in $G_1 \cup G_2$, and one eigenvalue in G_3.

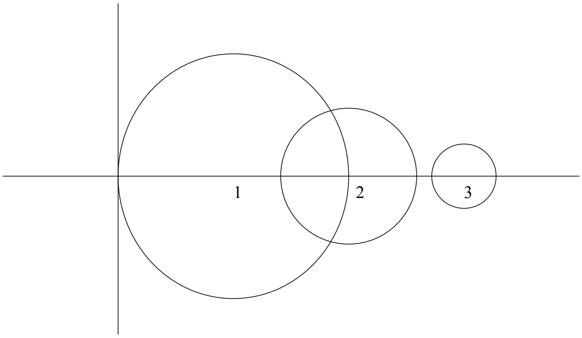

Figure 2.7: Geršgorin discs

Proof of Theorem 2.12. Let λ be an eigenvalue of A, and let x be an eigenvector of A corresponding to λ. Suppose that x_p is the largest component of x in absolute value; that is,

$$|x_p| \geq |x_i|, \quad i = 1, 2, \ldots, n.$$

Then $x_p \neq 0$. The equation $Ax = \lambda x$ gives

$$\sum_{j=1}^{n} a_{pj} x_j = \lambda x_p$$

or

$$x_p(\lambda - a_{pp}) = \sum_{j=1,\, j \neq p}^{n} a_{pj} x_j.$$

By taking absolute value, we have

$$\begin{aligned} |x_p||\lambda - a_{pp}| &= \Big| \sum_{j=1,\, j \neq p}^{n} a_{pj} x_j \Big| \\ &\leq \sum_{j=1,\, j \neq p}^{n} |a_{pj} x_j| \\ &= \sum_{j=1,\, j \neq p}^{n} |a_{pj}||x_j| \end{aligned}$$

$$\begin{aligned} &\le |x_p| \sum_{j=1,\, j\neq p}^{n} |a_{pj}| \\ &= |x_p| r_p. \end{aligned}$$

It follows that $|\lambda - a_{pp}| \le r_p$, and that λ lies in a closed disc centered at a_{pp} with radius r_p.

To prove the second part, we assume that the first k discs, centered at $a_{11}, \ldots, a_{kk}$, form a connected union G which is disjoint from other discs. Let $A = D + B$, where $D = \operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn})$, and

$$A_\epsilon = D + \epsilon B, \quad \epsilon \in [0, 1]$$

and denote r_i' for A_ϵ as r_i for A. Then $r_i' = \epsilon r_i$. It is immediate that for every $\epsilon \in [0,1]$ the set G contains the union G_ϵ of the first k discs of A_ϵ, where

$$G_\epsilon = \cup_{i=1}^{k} \{z \in \mathbb{C} : \ |z - a_{ii}| \le r_i'\}.$$

Consider the eigenvalues of A_0 and A_ϵ:

$$\lambda_i(A_0) = a_{ii}, \quad \lambda_i(A_\epsilon), \quad i = 1, 2, \ldots, k, \ \ \epsilon > 0.$$

Because the eigenvalues are continuous functions of the entries of A and because for each $i = 1, 2, \ldots, k$,

$$\lambda_i(A_0) \in G_\epsilon \subseteq G, \quad \text{for all } \epsilon \in [0, 1],$$

we have that each $\lambda_i(A_0)$ is joined to some $\lambda_i(A_1) = \lambda_i(A)$ by the continuous curve

$$\{\lambda_i(A_\epsilon) : \ 0 \le \epsilon \le 1\} \subseteq G.$$

Thus for each $\epsilon \in [0,1]$, there are at least k eigenvalues of A_ϵ in G_ϵ, and G contains at least k eigenvalues of A (not necessarily different). The remaining $n - k$ eigenvalues of A_0 start outside the connected set G, and those eigenvalues of A lie outside G. ∎

An application of this theorem to A^T gives a version of the theorem for the columns of A.

Problems

1. Apply the Geršgorin theorem to the matrix

$$\begin{pmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \\ 1 & 2 & 4 \end{pmatrix}.$$

2. Show by the Geršgorin theorem that A has three different eigenvalues:

$$A = \begin{pmatrix} 1 & \frac{1}{2} & \frac{i}{2} \\ \frac{1}{2} & 3 & 0 \\ 0 & 1 & 5 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & i & 0 \\ \frac{1}{2} & 4 & \frac{i}{2} \\ 1 & 0 & 7 \end{pmatrix}.$$

   Conclude that A is diagonalizable, i.e., $S^{-1}AS$ is diagonal for some S.

3. Construct a 4×4 complex matrix so that it contains no zero entries and that the four different eigenvalues of the matrix lie in the discs centered at 1, -1, i, and $-i$, all with diameter 1.

4. Illustrate the Geršgorin theorem by the matrix

$$A = \begin{pmatrix} 2 & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & 1+2i & 0 & \frac{1}{4} \\ -\frac{1}{2} & \frac{1}{4} & -1 & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{2} & -2-2i \end{pmatrix}.$$

5. State and prove the first part of Geršgorin theorem for columns.

6. Let $A \in \mathbb{M}_n$ and $D = \operatorname{diag}(A)$; that is, the diagonal matrix of A. Denote $B = A - D$. If λ is an eigenvalue of A and it is not a diagonal entry of A, show that 1 is an eigenvalue of $(\lambda I - D)^{-1}B$.

7. Let $A = (a_{ij}) \in \mathbb{M}_n$. Show that for any eigenvalue λ of A

$$|\lambda| \geq \min_i \{|a_{ii}| - r_i\}, \quad \text{where } r_i = \sum_{j=1,\, j\neq i}^{n} |a_{ij}|.$$

   Derive that

$$|\det A| \geq (\min_i \{|a_{ii}| - r_i\})^n.$$

8. Let $A = (a_{ij}) \in \mathbb{M}_n$. Show that for any eigenvalue λ of A

$$|\lambda| \leq \min\left\{ \max_i \sum_{s=1}^{n} |a_{is}|,\ \max_j \sum_{t=1}^{n} |a_{tj}| \right\}.$$

9. Let $A = (a_{ij}) \in \mathbb{M}_n$. Show that for any eigenvalue λ of A

$$|\lambda| \leq n \max_{i,j} |a_{ij}|.$$

10. Let $A = (a_{ij}) \in \mathbb{M}_n$. Show that for any positive numbers $d_1, \ldots, d_n$ and for any eigenvalue λ of A

$$|\lambda| \leq \min \left\{ \max_i \frac{1}{d_i} \sum_{s=1}^n d_s |a_{is}|, \ \max_j \frac{1}{d_j} \sum_{t=1}^n d_t |a_{tj}| \right\}.$$

11. **(Levy–Desplanques)** A matrix $A \in \mathbb{M}_n$ is said to be *strictly diagonally dominant* if

$$|a_{ij}| > \sum_{i=1,\ i \neq j}^n |a_{ij}|, \quad j = 1, 2, \ldots, n.$$

Show that a strictly diagonally dominant matrix is nonsingular.

12. Let $A = (a_{ij}) \in \mathbb{M}_n$ and $R_i = \sum_{j=1}^n |a_{ij}|$, $i = 1, 2, \ldots, n$. Suppose that A does not have a zero row. Show that

$$\operatorname{rank}(A) \geq \sum_{i=1}^n \frac{|a_{ii}|}{R_i}.$$

13. Let $A = (a_{ij}) \in \mathbb{M}_n$. Denote $\delta = \min_{i \neq j} |a_{ii} - a_{jj}|$ and $\epsilon = \max_{i \neq j} |a_{ij}|$. If $\delta > 0$ and $\epsilon \leq \delta/4n$, show that each Geršgorin disc contains exactly one eigenvalue of A.

14. Let $f(x) = x^n + a_1 x^{n-1} + \cdots + a_n$ be any monic polynomial. Show that all the roots of f are bounded (in absolute value) by

$$\gamma = 2 \max_{1 \leq k \leq n} |a_k|^{1/k}.$$

[Hint: Consider $|f(x)/x^n|$ and show that $|f(x)/x^n| > 0$ if $|x| > \gamma$.]

CHAPTER 3

Matrix Polynomials and Canonical Forms

Introduction: This chapter is devoted to matrix decompositions. The main studies are on the Schur decomposition, spectral decomposition, singular value decomposition, Jordan decomposition, and numerical range. Attention is also paid to the polynomials that annihilate matrices, especially the minimal and characteristic polynomials, and to the similarity of a complex matrix to a real matrix. At the end we introduce three important matrix operations: the Hadamard product, the Kronecker product, and compound matrices.

3.1 Commuting Matrices

Matrices do not commute in general. One may easily find two square matrices A and B of the same size such that $AB \neq BA$. Any square matrix A, however, commutes with polynomials in A.

A question arises: If a matrix B commutes with A, is it true that B can be expressed as a polynomial in A? The answer is negative, by taking A to be the $n \times n$ identity matrix I and B to be an $n \times n$ nondiagonal matrix. For some sorts of matrices, nevertheless, we have the following result.

Theorem 3.1 *Let A and B be $n \times n$ matrices such that $AB = BA$. If all the eigenvalues of A are distinct, then B can be expressed uniquely as a polynomial in A with degree no more than $n-1$.*

F. Zhang, *Matrix Theory: Basic Results and Techniques*, Universitext,
DOI 10.1007/978-1-4614-1099-7_3,

Proof. To begin, recall from Theorem 1.6 that the eigenvectors belonging to different eigenvalues are linearly independent. Thus, a matrix with distinct eigenvalues has a set of linearly independent eigenvectors that form a basis of $\mathbb{C}^n$ (or $\mathbb{R}^n$ if the matrix is real).

Let $u_1, u_2, \ldots, u_n$ be the eigenvectors corresponding to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A, respectively. Set

$$P = (u_1, u_2, \ldots, u_n).$$

Then P is an invertible matrix and

$$P^{-1}AP = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

Let

$$P^{-1}AP = C \quad \text{and} \quad P^{-1}BP = D.$$

It follows from $AB = BA$ that $CD = DC$.

The diagonal entries of C are distinct, therefore D must be a diagonal matrix too (Problem 1). Let $D = \operatorname{diag}(\mu_1, \mu_2, \ldots, \mu_n)$.

Consider the linear equation system of unknowns $x_0, x_1, \ldots, x_{n-1}$:

$$\begin{gathered} x_0 + \lambda_1 x_1 + \cdots + \lambda_1^{n-1} x_{n-1} = \mu_1, \\ x_0 + \lambda_2 x_1 + \cdots + \lambda_2^{n-1} x_{n-1} = \mu_2, \\ \vdots \\ x_0 + \lambda_n x_1 + \cdots + \lambda_n^{n-1} x_{n-1} = \mu_n. \end{gathered}$$

Because the coefficient matrix is a Vandermonde matrix that is nonsingular when $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct (see Problem 3 of Section 1.2 or Theorem 5.9 in Chapter 5), the equation system has a unique solution, say, $(a_0, a_1, \ldots, a_{n-1})$.

Define a polynomial with degree no more than $n-1$ by

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}.$$

It follows that

$$p(\lambda_i) = \mu_i, \quad i = 1, 2, \ldots, n.$$

It is immediate that $p(A) = B$ since $p(C) = D$ and that this polynomial $p(x)$ is unique for the solution to the system is unique. ∎

Such a method of finding a polynomial $p(x)$ of the given pairs is referred to as *interpolation*. The proof also shows that there exists an invertible matrix P such that $P^{-1}AP$ and $P^{-1}BP$ are both diagonal.

Theorem 3.2 *Let A and B be square matrices of the same size. If $AB = BA$, then there exists a unitary matrix U such that U^*AU and U^*BU are both upper-triangular.*

Proof. We use induction on n. If $n = 1$, we have nothing to show. Suppose that the assertion is true for $n - 1$.

For the case of n, we consider matrices as linear transformations. Note that if A is a linear transformation on a finite-dimensional vector space V over $\mathbb{C}$, then A has at least one eigenvector in V, for

$$Ax = \lambda x, \quad x \neq 0, \quad \text{if and only if} \quad \det(\lambda I - A) = 0,$$

which has a solution in $\mathbb{C}$.

For each eigenvalue μ of B, consider the eigenspace of B

$$V_\mu = \{v \in \mathbb{C}^n : \ Bv = \mu v\}.$$

If A and B commute, then for every $v \in V_\mu$,

$$B(Av) = (BA)v = (AB)v = A(Bv) = A(\mu v) = \mu(Av).$$

Thus, $Av \in V_\mu$; that is, V_μ is an invariant subspace of A. As a linear transformation on V_μ, A has an eigenvalue λ and a corresponding eigenvector v_1 in V_μ. Put in symbols,

$$Av_1 = \lambda v_1, \quad Bv_1 = \mu v_1, \quad v_1 \in V_\mu.$$

We may assume that v_1 is a unit vector. Extend v_1 to a unitary matrix U_1; that is, U_1 is a unitary matrix whose first column is v_1. By computation, we have

$$U_1^*AU_1 = \begin{pmatrix} \lambda & \alpha \\ 0 & C \end{pmatrix} \quad \text{and} \quad U_1^*BU_1 = \begin{pmatrix} \mu & \beta \\ 0 & D \end{pmatrix},$$

where C, $D \in \mathbb{M}_{n-1}$, and α and β are some row vectors.

It follows from $AB = BA$ that $CD = DC$. The induction hypothesis guarantees a unitary matrix $U_2 \in \mathbb{M}_{n-1}$ such that $U_2^* C U_2$ and $U_2^* D U_2$ are both upper-triangular. Let

$$U = U_1 \begin{pmatrix} 1 & 0 \\ 0 & U_2 \end{pmatrix}.$$

Then U, a product of two unitary matrices, is unitary, and U^*AU and U^*BU are both upper-triangular. ∎

Problems

1. Let A be a diagonal matrix with different diagonal entries. If B is a matrix such that $AB = BA$, show that B is also diagonal.

2. Let $A, B \in \mathbb{M}_n$ and let A have n distinct eigenvalues. Show that $AB = BA$ if and only if there exists a set of n linearly independent vectors as the eigenvectors of A and B.

3. Let $A = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right)$. Show that $AB = BA$. Find a unitary matrix U such that both U^*AU and U^*BU are upper-triangular. Show that such a U cannot be a real matrix.

4. Give an example of matrices A and B for which $AB = BA$, λ is an eigenvalue of A, μ is an eigenvalue of B, but $\lambda + \mu$ is not an eigenvalue of $A + B$, and $\lambda\mu$ is not an eigenvalue of AB.

5. Let $f(x)$ be a polynomial and let A be an n-square matrix. Show that for any n-square invertible matrix P,

$$f(P^{-1}AP) = P^{-1}f(A)P$$

   and that there exists a unitary matrix U such that both U^*AU and $U^*f(A)U$ are upper-triangular.

6. Show that the adjoints, inverses, sums, products, and polynomials of upper-triangular matrices are upper-triangular.

7. Show that every square matrix is a sum of two commuting matrices.

8. If A and B are two matrices such that $AB = I_m$ and $BA = I_n$, show that $m = n$, $AB = BA = I$, and $B = A^{-1}$.

9. Let A, B, and C be matrices such that $AB = CA$. Show that for any polynomial $f(x)$,

$$Af(B) = f(C)A.$$

10. Is it true that any linear transformation on a vector space over $\mathbb{R}$ has at least a real eigenvalue?

11. Show that if $AB = BA = 0$ then $\operatorname{rank}(A+B) = \operatorname{rank}(A) + \operatorname{rank}(B)$.

12. Let $A_1, A_2, \ldots, A_k \in \mathbb{M}_n$ be commuting matrices, i.e., $A_i A_j = A_j A_i$ for all i, j. Show that there exists a unitary matrix $U \in \mathbb{M}_n$ such that all $U^* A_i U$ are upper-triangular.

13. Let A and B be commuting matrices. If A has k distinct eigenvalues, show that B has at least k linearly independent eigenvectors. Does it follow that B has k distinct eigenvalues?

14. Let A and B be n-square matrices. If $AB = BA$, what are the eigenvalues of $A + B$ and AB in terms of those of A and B?

15. Let A, $B \in \mathbb{M}_n$. If $AB = BA$, find the eigenvalues of the matrix
$$\begin{pmatrix} A & B \\ B & -A \end{pmatrix}.$$

16. What matrices in $\mathbb{M}_n$ commute with all diagonal matrices? With all Hermitian matrices? With all matrices in $\mathbb{M}_n$?

17. Let A and B be square complex matrices. If A commutes with B and B^*, show that $A + A^*$ commutes with $B + B^*$.

18. Show that Theorem 3.2 holds for more than two commuting matrices.

19. What conclusion can be drawn from Theorem 3.2 if B is assumed to be the identity matrix?

20. Let A and B be complex matrices. Show that
$$AB = A + B \quad \Rightarrow \quad AB = BA.$$

21. Let A and B be n- and m-square matrices, respectively, with $m \le n$. If $AP = PB$ for an $n \times m$ matrix P with columns linearly independent, show that every eigenvalue of B is an eigenvalue of A.

22. If A and B are nonsingular matrices such that $AB - BA$ is singular, show that 1 is an eigenvalue of $A^{-1}B^{-1}AB$.

23. Let A and B be $n \times n$ matrices such that $\operatorname{rank}(AB - BA) \le 1$. Show that A and B have a common eigenvector. Find a common eigenvector (probably belonging to different eigenvalues) for
$$A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

24. Let $S = \{AB - BA : A, B \in \mathbb{M}_n\}$. Show that $\operatorname{Span} S$ is a subspace of $\mathbb{M}_n$ and that
$$\dim(\operatorname{Span} S) = n^2 - 1.$$

25. Let A and B be $2n \times 2n$ matrices partitioned conformally as
$$A = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$
If A commutes with B, show that A_{11} and A_{22} commute with B_{11} and B_{22}, respectively, and that for any polynomial $f(x)$
$$f(A_{11})B_{12} = B_{12}f(A_{22}), \quad f(A_{22})B_{21} = B_{21}f(A_{11}).$$
In particular, if $A_{11} = a_1 I$ and $A_{22} = a_2 I$ with $a_1 \neq a_2$, then $B_{12} = B_{21} = 0$, and thus $B = B_{11} \oplus B_{22}$. The same conclusion follows if $f(A_{11}) = 0$ (or singular) and $f(A_{22})$ is nonsingular (respectively, 0).

26. Let S be the $n \times n$ *backward identity* matrix; that is,
$$S = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}.$$
Show that $S^{-1} = S^T = S$; equivalently, $S^T = S$, $S^2 = I$, $S^T S = I$. What is $\det S$? When $n = 3$, compute SAS for $A = (a_{ij}) \in \mathbb{M}_3$.

27. Let $A = (a_{ij})$ be an $n \times n$ matrix and S be the $n \times n$ backward identity. Denote $\hat{A} = (\hat{a}_{ij}) = SAS$. Show that $\hat{a}_{ij} = a_{n-i+1,n-j+1}$. [Note: The matrix SAS can be obtained by any of the following methods: (1) Relist all the rows in the reverse order then relist all the columns of the resulting matrix; (2) Flip the matrix along the main diagonal then flip the resulting matrix along the backward diagonal $a_{1n}, \dots, a_{n1}$; or (3) Rotate the matrix 180° in the plane; that is, hold the upper-left corner, then rotate the paper by 180°.]

If A is one of the following matrices, show that the other one is SAS.
$$\begin{pmatrix} + & - & \times & \div \\ \oplus & \ominus & \otimes & \odot \\ \diamond & \bowtie & \infty & \star \\ = & \equiv & | & \| \end{pmatrix}, \qquad \begin{pmatrix} \| & | & \equiv & = \\ \star & \infty & \bowtie & \diamond \\ \odot & \otimes & \ominus & \oplus \\ \div & \times & - & + \end{pmatrix}.$$

. ——————— $\odot$ ——————— .

3.2 Matrix Decompositions

Factorizations of matrices into some special sorts of matrices via similarity are of fundamental importance in matrix theory. We study the following decompositions of matrices in this section: the Schur decomposition, spectral decomposition, singular value decomposition, and polar decomposition. We also continue our study of Jordan decomposition in later sections.

Theorem 3.3 (Schur Decomposition) *Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of $A \in \mathbb{M}_n$. Then there exists a unitary matrix $U \in \mathbb{M}_n$ such that U^*AU is an upper-triangular matrix. In symbols,*

$$U^*AU = \begin{pmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

Proof. This theorem follows from Theorem 3.2. We present a pure matrix proof below without using the theory of vector spaces.

If $n = 1$, there is nothing to show. Suppose the statement is true for matrices with sizes less than n. We show by induction that it is true for the matrices of size n.

Let x_1 be a unit eigenvector of A belonging to eigenvalue λ_1:

$$Ax_1 = \lambda_1 x_1, \quad x_1 \neq 0.$$

Extend x_1 to a unitary matrix $S = (x_1,\ y_2,\ \ldots,\ y_n)$. Then

$$\begin{aligned} AS &= (Ax_1, Ay_2, \ldots, Ay_n) \\ &= (\lambda_1 x_1, Ay_2, \ldots, Ay_n) \\ &= S(u, S^{-1}Ay_2, \ldots, S^{-1}Ay_n), \end{aligned}$$

where $u = (\lambda_1, 0, \ldots, 0)^T$. Thus, we can write

$$S^*AS = \begin{pmatrix} \lambda_1 & v \\ 0 & B \end{pmatrix},$$

where v is a row vector and $B \in \mathbb{M}_{n-1}$.

Applying the induction hypothesis on B, we have a unitary matrix T of size $n-1$ such that T^*BT is upper-triangular. Let

$$U = S \begin{pmatrix} 1 & 0 \\ 0 & T \end{pmatrix}.$$

Then U, a product of two unitary matrices, is unitary, and U^*AU is upper-triangular. It is obvious that the diagonal entries λ_i of the upper-triangular matrix are the eigenvalues of A. ∎

A weaker statement is that of *triangularization.* For every $A \in \mathbb{M}_n$ there exists an invertible P such that $P^{-1}AP$ is upper-triangular.

Schur triangularization is one of the most important theorems in linear algebra and matrix theory. It is used repeatedly in this book. As an application, we see by taking the conjugate transpose that any Hermitian matrix A (i.e., $A^* = A$) is unitarily diagonalizable. The same is true for normal matrices A, because the matrix identity $A^*A = AA^*$, together with the Schur decomposition of A, implies the desired decomposition form of A (Problem 4).

A *positive semidefinite matrix* $A \in \mathbb{M}_n$, by definition, $x^*Ax \geq 0$ for all $x \in \mathbb{C}^n$ (instead of $\mathbb{R}^n$), has a similar structure. To see this, it suffices to show that a positive semidefinite matrix is necessarily Hermitian. This goes as follows.

Since $x^*Ax \geq 0$ for every $x \in \mathbb{C}^n$, we have, by taking x to be the column vector with the sth component 1, the tth component $c \in \mathbb{C}$, and 0 elsewhere,

$$x^*Ax = a_{ss} + a_{tt}|c|^2 + a_{ts}\bar{c} + a_{st}c \geq 0.$$

It follows that each diagonal entry a_{ss} is nonnegative by putting $c = 0$ and that $a_{st} = \overline{a_{ts}}$ or $A^* = A$ by putting $c = 1$, i, respectively.

Note that if A is real and $x^TAx \geq 0$ for all real vectors x, A need not be symmetric. Let $A = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Then $x^TAx = 0$ for all $x \in \mathbb{R}^2$.

It is immediate that the eigenvalues λ of a positive semidefinite A are nonnegative since $x^*Ax = \lambda x^*x$ for any eigenvector x of λ.

We summarize these discussions in the following theorem.

Theorem 3.4 (Spectral Decomposition) *Let A be an n-square complex matrix with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$. Then A is normal if and only if A is unitarily diagonalizable; that is, there exists a unitary matrix U such that*

$$U^*AU = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

In particular, A is Hermitian if and only if the λ_i are all real and is positive semidefinite if and only if the λ_i are all nonnegative.

As a result, by taking the square roots of the λ_i in the decomposition, we see that for any positive semidefinite matrix A, there exists a positive semidefinite matrix B such that $A = B^2$. We show such a matrix B is unique (see also Section 7.1) and call it a *square root* of A, denoted by $A^{1/2}$. In addition, we write $A \geq 0$ if A is positive semidefinite and $A > 0$ if A is positive definite; that is, $x^*Ax > 0$ for all nonzero $x \in \mathbb{C}^n$. For two Hermitian matrices A and B of the same size, we write $A \geq B$ if $A - B \geq 0$ and $A > B$ if $A - B > 0$.

Theorem 3.5 (Uniqueness of Square Root) *Let A be an n-square positive semidefinite matrix. Then there is a unique n-square positive semidefinite matrix B such that $B^2 = A$.*

Proof. Let $A = U^* \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)U$ be a spectral decomposition of A. Take $B = U^* \operatorname{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}\,)U$. Then $B \geq 0$ and $B^2 = A$. For uniqueness, let C also be a positive semidefinite matrix such that $C^2 = A$ and, by the spectral decomposition, write $C = V^* \operatorname{diag}(\mu_1, \mu_2, \dots, \mu_n)V$ (actually $\mu_i = \sqrt{\lambda_i}\,$). Then $B^2 = C^2 = A$ implies that $U^* \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)U = V^* \operatorname{diag}(\mu_1^2, \mu_2^2, \dots, \mu_n^2)V$; that is, $W \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \operatorname{diag}(\mu_1^2, \mu_2^2, \dots, \mu_n^2)W$, where $W = (w_{ij}) = VU^*$. This results in $w_{ij}\lambda_j = \mu_i^2 w_{ij}$ for all i, j. It follows that $w_{ij}\sqrt{\lambda_j} = \mu_i w_{ij}$. Therefore, $W \operatorname{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}\,) = \operatorname{diag}(\mu_1, \mu_2, \dots, \mu_n)W$, which reveals $B = C$. ∎

Such a B in the theorem is called the *square root* of A and is denoted by $A^{1/2}$. The proof shows that if $A = U^* \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)U$, where all $\lambda_i > 0$, then $A^{1/2} = U^* \operatorname{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}\,)U$. Likewise, $A^{1/3} = U^* \operatorname{diag}(\sqrt[3]{\lambda_1}, \sqrt[3]{\lambda_2}, \dots, \sqrt[3]{\lambda_n}\,)U$. Similarly, for any $r > 0$, one may verify the following A^r is well defined (Problem 30):

$$A^r = U^* \operatorname{diag}(\lambda_1^r, \lambda_2^r, \dots, \lambda_n^r)U.$$

The *singular values* of a matrix A are defined to be the nonnegative square roots of the eigenvalues of A^*A, which is positive semidefinite, for $x^*(A^*A)x = (Ax)^*(Ax) \geq 0$. If we denote by σ_i a singular value and by λ_i an eigenvalue, we may simply write

$$\sigma_i(A) = \sqrt{\lambda_i(A^*A)}.$$

Theorem 3.6 (Singular Value Decomposition) *Let A be an $m \times n$ matrix with nonzero singular values $\sigma_1, \sigma_2, \dots, \sigma_r$. Then there exist an $m \times m$ unitary U and an $n \times n$ unitary V such that*

$$A = U \begin{pmatrix} D_r & 0 \\ 0 & 0 \end{pmatrix} V, \tag{3.1}$$

where the block matrix is of size $m \times n$ and $D_r = \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$.

Proof. If A is a number c, say, then the absolute value $|c|$ is the singular value of A, and $A = |c|e^{i\theta}$ for some $\theta \in \mathbb{R}$. If A is a nonzero row or column vector, say, $A = (a_1, \dots, a_n)$, then σ_1 is the norm (length) of the vector A. Let V be a unitary matrix with the first row the unit vector $(\frac{1}{\sigma_1}a_1, \dots, \frac{1}{\sigma_1}a_n)$. Then $A = (\sigma_1, 0, \dots, 0)V$.

We now assume $m > 1$, $n > 1$, and $A \neq 0$. Let u_1 be a unit eigenvector of A^*A belonging to σ_1^2; that is,

$$(A^*A)u_1 = \sigma_1^2 u_1, \quad u_1^* u_1 = 1.$$

Let

$$v_1 = \frac{1}{\sigma_1} A u_1.$$

Then v_1 is a unit vector and a simple computation gives $u_1^* A^* v_1 = \sigma_1$.

Let P and Q be unitary matrices with u_1 and v_1 as their first columns, respectively. Then, with $A^*v_1 = \sigma_1 u_1$ and $u_1^* A^* = (Au_1)^* = \sigma_1 v_1^*$, we see the first column of P^*A^*Q is $(\sigma_1, 0, \dots, 0)^T$ and the first row is $(\sigma_1, 0, \dots, 0)$. It follows that

$$P^*A^*Q = \begin{pmatrix} \sigma_1 & 0 \\ 0 & B \end{pmatrix} \quad \text{or} \quad A = Q \begin{pmatrix} \sigma_1 & 0 \\ 0 & B^* \end{pmatrix} P^*$$

for some $(n-1) \times (m-1)$ matrix B. The assertion follows by repeating the process (or by induction) on B^*. ∎

Apparently the rank of A equals r as U and V are nonsingular. If A is a real matrix, U and V can be chosen to be real. Besides, when A is an $n \times n$ matrix, then U and V are $n \times n$ unitary matrices. By inserting VV^* between U and the block matrix in (3.1), we have $A = UVV^*DV = WP$, where $W = UV$ is unitary and $P = V^*DV$ is positive semidefinite. Since $A^*A = PW^*WP = P^2$, we see $P = (A^*A)^{1/2}$, which is uniquely determined by the matrix A. We denote

$$|A| = (A^*A)^{1/2}$$

and call it the *modulus* of A. Note that $|A|$ is positive semidefinite.

Theorem 3.7 (Polar Decomposition) *For any square matrix A, there exist unitary matrices U and V such that*

$$A = W|A| = |A^*|V.$$

The polar decomposition may be generalized for rectangular matrices with partial unitary matrices (Problem 24). The polar decomposition was proven by the singular value decomposition. One may prove the latter using the polar decomposition and the spectral decomposition.

Problems

1. Let B and D be square matrices (of possibly different sizes). Let

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}.$$

Show that every eigenvalue of B or D is an eigenvalue of A.

2. Find a matrix P so that $P^{-1}AP$ is diagonal; then compute A^5, where

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

3. Show that the complex symmetric A is not diagonalizable, where

$$A = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}.$$

That is, $P^{-1}AP$ is not diagonal for any invertible matrix P.

4. If A is an upper-triangular matrix such that $A^*A = AA^*$, show that A is in fact a diagonal matrix.

5. Let A be an n-square complex matrix. Show that

$$x^*Ax = 0 \text{ for all } x \in \mathbb{C}^n \quad \Leftrightarrow \quad A = 0$$

and

$$x^TAx = 0 \text{ for all } x \in \mathbb{R}^n \quad \Leftrightarrow \quad A^T = -A.$$

6. Let $A \in \mathbb{M}_n$. If $x^*Ax \in \mathbb{R}$ for all $x \in \mathbb{C}^n$, show that $A^* = A$ by the previous problem or by making use of the spectral theorem on $A - A^*$. (Note that $A - A^*$ is skew-Hermitian, thus normal.)

7. Let $A \in \mathbb{M}_n$. Show that if λ is an eigenvalue of A, then λ^k is an eigenvalue of A^k, and that $\alpha \in \mathbb{C}$ is an eigenvalue of $f(A)$ if and only if $\alpha = f(\lambda)$ for some eigenvalue λ of A, where f is a polynomial.

8. Show that if $A \in \mathbb{M}_n$ has n distinct eigenvalues, then A is diagonalizable. Does the converse hold?

9. Let A be an n-square positive semidefinite matrix. Show that

   (a) $X^*AX \geq 0$ for every $n \times m$ matrix X.

   (b) Every principal submatrix of A is positive semidefinite.

10. Let

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

   (a) Show that $A^3 = B^3 = C^3 = 0$, where $C = \lambda A + \mu B,\ \lambda,\ \mu \in \mathbb{C}$.

   (b) Does there exist an integer k such that $(AB)^k = 0$?

   (c) Does there exist a nonsingular matrix P such that $P^{-1}AP$ and $P^{-1}BP$ are both upper-triangular?

11. Let A be an n-square matrix and let P be a nonsingular matrix of the same size such that $P^{-1}AP$ is upper-triangular. Write

$$P^{-1}AP = D - U,$$

where D is a diagonal matrix and U is an upper-triangular matrix with 0 on the diagonal. Show that if A is invertible, then

$$A^{-1} = PD^{-1}\Big(I + UD^{-1} + (UD^{-1})^2 + \cdots + (UD^{-1})^{n-1}\Big)P^{-1}.$$

12. Let A be an $n \times n$ real matrix. If all eigenvalues of A are real, then there exists an $n \times n$ real orthogonal matrix Q such that $Q^T AQ$ is upper-triangular. What if the eigenvalues of A are not real?

13. Let A be an n-square complex matrix. Show that

    (a) (*QR Factorization*) There exist a unitary matrix Q and an upper-triangular matrix R such that $A = QR$. Furthermore Q and R can be chosen to be real if A is real.

    (b) (*LU Factorization*) If all the leading principal minors of A are nonzero, then $A = LU$, where L and U are lower- and upper-triangular matrices, respectively.

14. Let X_{ii} denote the (i,i)-entry of matrix X. If $A \geq 0$, show that

$$(A^{1/2})_{ii} \leq (A_{ii})^{1/2}.$$

15. Let $A \in \mathbb{M}_n$ have rank r. Show that A is normal if and only if

$$A = \sum_{i=1}^{r} \lambda_i u_i u_i^*,$$

where λ_i are complex numbers and u_i are column vectors of a unitary matrix. Further show that A is Hermitian if and only if all λ_i are real, and A is positive semidefinite if and only if all λ_i are nonnegative.

16. Let A be an $m \times n$ complex matrix with rank r. Show that

    (a) A has r nonzero singular values.

    (b) A has at most r nonzero eigenvalues (in case $m = n$).

    (c) $A = U_r D_r V_r$, where U_r is an $m \times r$ matrix, $D_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$, V_r is an $r \times n$ matrix, all of rank r, and $U_r^* U_r = V_r V_r^* = I_r$.

    (d) $A = \sum_{i=1}^{r} \sigma_i u_i v_i^*$, where the σ_i are the singular values of A, and u_i and v_i are column vectors of some unitary matrices.

17. Let A and B be upper-triangular matrices with positive diagonal entries. If $A = UB$ for some unitary U, show that $U = I$ and $A = B$.

18. Let A be a square matrix. Show that A has a zero singular value if and only if A has a zero eigenvalue. Does it follow that the number of zero singular values is equal to that of the zero eigenvalues?

19. Show that two Hermitian matrices are (unitarily) similar if and only if they have the same set of eigenvalues.

20. Let $A = P_1U_1 = P_2U_2$ be two polar decompositions of A. Show that $P_1 = P_2$ and $U_1 = U_2$ if A is nonsingular. What if A is singular?

21. Let P be a positive definite matrix and U and V be unitary matrices such that $UP = PV$. Show that $U = V$.

22. Show that if A is an n-square complex matrix, then there exist nonsingular matrices P and Q such that $(PA)^2 = PA$, $(AQ)^2 = AQ$.

23. Let $A \in \mathbb{M}_n$. Show that $|A| = UA$ for some unitary matrix U.

24. State and show the polar decomposition for rectangular matrices.

25. Let A be an $m \times n$ complex matrix of rank r. Show that $A = ST$ for some $m \times r$ matrix S and $r \times n$ matrix T; both have rank r.

26. Show that $|\lambda_1 \cdots \lambda_n| = \sigma_1 \cdots \sigma_n$ for any $A \in \mathbb{M}_n$, where the λ_i and σ_i are the eigenvalues and singular values of A, respectively.

27. What can be said about $A \in \mathbb{M}_n$ if all its singular values are equal? Can the same conclusion be drawn in the case of eigenvalues?

28. For any n-square complex matrix A, show that
$$\operatorname{tr}\left(\frac{A^* + A}{2}\right) \le \operatorname{tr}\left((A^*A)^{1/2}\right).$$

29. If A is a matrix with eigenvalues λ_i and singular values σ_i, show that
$$\sum_i |\lambda_i|^2 \le \operatorname{tr}(A^*A) = \operatorname{tr}(AA^*) = \sum_{i,j} |a_{ij}|^2 = \sum_i \sigma_i^2.$$
Equality occurs if and only if A is normal. Use this and the matrix with $(i, i+1)$ entry $\sqrt{x_i}$, $i = 1, 2, \ldots, n-1$, $(n, 1)$ entry $\sqrt{x_n}$, and 0 elsewhere to show the arithmetic mean–geometric mean inequality
$$\left(\prod_{i=1}^n x_i\right)^{1/n} \le \frac{1}{n}\sum_{i=1}^n x_i.$$

30. Let $A \ge 0$. Show that the definition $A^r = U^* \operatorname{diag}(\lambda_1^r, \lambda_2^r, \ldots, \lambda_n^r)U$ is independent of the unitary matrix U. In other words, if $A = V^* \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)V = W^* \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)W$, V, W are unitary, then $V^* \operatorname{diag}(\lambda_1^r, \lambda_2^r, \ldots, \lambda_n^r)V = W^* \operatorname{diag}(\lambda_1^r, \lambda_2^r, \ldots, \lambda_n^r)W$.

31. Let $A, B \ge 0$ and U be unitary, all $n \times n$. Show that for any $r > 0$,

    (a) $(U^*AU)^r = U^*A^rU$.

    (b) If $AB = BA$, then $AB \ge 0$ and $(AB)^r = A^rB^r$.

3.3 Annihilating Polynomials of Matrices

Given a polynomial in λ with complex coefficients $a_m, a_{m-1}, \ldots, a_0$,

$$f(\lambda) = a_m \lambda^m + a_{m-1}\lambda^{m-1} + \cdots + a_1 \lambda + a_0,$$

one can always define a matrix polynomial for $A \in \mathbb{M}_n$ by

$$f(A) = a_m A^m + a_{m-1} A^{m-1} + \cdots + a_1 A + a_0 I.$$

We consider in this section the annihilating polynomials of a matrix; that is, the polynomials $f(\lambda)$ for which $f(A) = 0$. Particular attention is paid to the characteristic and minimal polynomials.

Theorem 3.8 *Let A be an n-square complex matrix. Then there exists a nonzero polynomial $f(\lambda)$ over $\mathbb{C}$ such that $f(A) = 0$.*

Proof. $\mathbb{M}_n$ is a vector space over $\mathbb{C}$ of dimension n^2. Thus, any n^2+1 vectors in $\mathbb{M}_n$ are linearly dependent. In particular, the matrices

$$I,\ A,\ A^2,\ \ldots,\ A^{n^2}$$

are linearly dependent; namely, there exist numbers $a_0, a_1, a_2, \ldots, a_{n^2}$, not all zero, such that

$$a_0 I + a_1 A + a_2 A^2 + \cdots + a_{n^2} A^{n^2} = 0.$$

Set

$$f(\lambda) = a_0 + a_1\lambda + a_2 \lambda^2 + \cdots + a_{n^2}\lambda^{n^2}.$$

Then $f(A) = 0$, as desired. ∎

Theorem 3.9 (Cayley–Hamilton) *Let A be an n-square complex matrix and let $p_A(\lambda)$ be the characteristic polynomial of A; that is,*

$$p_A(\lambda) = \det(\lambda I - A).$$

Then

$$p_A(A) = 0.$$

Proof. Let the eigenvalues of A be $\lambda_1, \lambda_2, \ldots, \lambda_n$. We write A by triangularization as, for some invertible matrix P,

$$A = P^{-1}TP,$$

where T is an upper-triangular matrix with $\lambda_1, \lambda_2, \ldots, \lambda_n$ on the diagonal. Factor the characteristic polynomial $p(\lambda) = p_{_A}(\lambda)$ of A as

$$p(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n).$$

Then

$$p(A) = p(P^{-1}TP) = P^{-1}p(T)P.$$

Note that

$$p(T) = (T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_n I).$$

It can be shown inductively that $(T - \lambda_1 I) \cdots (T - \lambda_k I)$ has the first k columns equal to zero, $1 \leq k \leq n$. Thus, $p(T) = 0$, and $p(A) = 0$. ∎

A monic polynomial $m(\lambda)$ is called the *minimal polynomial* of a matrix A if $m(A) = 0$ and it is of the smallest degree in the set

$$\{f(\lambda) : \ f(A) = 0\}.$$

It is immediate that if $f(A) = 0$, then $m(\lambda)$ divides $f(\lambda)$, or, in symbols, $m(\lambda) | f(\lambda)$, because otherwise we may write

$$f(\lambda) = q(\lambda)m(\lambda) + r(\lambda),$$

where $r(\lambda) \neq 0$ is of smaller degree than $m(\lambda)$ and $r(A) = 0$, a contradiction. In particular, the minimal polynomial divides the characteristic polynomial. Note that both the characteristic polynomial and the minimal polynomial are uniquely determined by its matrix.

Theorem 3.10 *Similar matrices have the same minimal polynomial.*

Proof. Let A and B be similar matrices such that $A = P^{-1}BP$ for some nonsingular matrix P, and let $m_{_A}(\lambda)$ and $m_{_B}(\lambda)$ be the minimal polynomials of A and B, respectively. Then

$$m_{_B}(A) = m_{_B}(P^{-1}BP) = P^{-1}m_{_B}(B)P = 0.$$

Thus, $m_A(\lambda)$ divides $m_B(\lambda)$. Similarly, $m_B(\lambda)$ divides $m_A(\lambda)$. Hence $m_A(\lambda) = m_B(\lambda)$ as they are both of leading coefficient 1. ∎

For a polynomial with (real or) complex coefficients

$$p(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0,$$

one may construct an n-square (real or) complex matrix

$$C = \begin{pmatrix} 0 & 0 & \dots & -c_0 \\ 1 & 0 & \dots & -c_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & -c_{n-1} \end{pmatrix}.$$

Such a matrix C is known as the *companion matrix* of $p(x)$.

By expanding $\det(xI - C)$, we see $p(x) = \det(xI - C)$; that is, $p(x)$ is the characteristic polynomial of its companion matrix. $p(x)$ is also the minimal polynomial of C. For this end, let $e_1, e_2, \dots, e_n$ be the column vectors with the 1st, 2nd, $\dots$, nth component 1, respectively, and all other components 0. Then Ce_i is the ith column of C, $i = 1, 2, \dots, n$. By looking at the first $n-1$ columns of C, we have

$$Ce_1 = e_2,\ Ce_2 = e_3 = C^2 e_1,\ \dots,\ Ce_{n-1} = e_n = C^{n-1} e_1$$

and

$$Ce_n = -c_0 e_1 - c_1 e_2 - \cdots - c_{n-1} e_n.$$

If $q(x) = x^m + d_{m-1}x^{m-1} + d_{m-2}x^{m-2} + \cdots + d_1 x + d_0$ is such a polynomial that $q(C) = 0$, $m < n$, we compute $q(C)e_1 = 0$ to get

$$\begin{aligned} 0 &= C^m e_1 + d_{m-1}C^{m-1}e_1 + d_{m-2}C^{m-2}e_1 + \cdots + d_1 Ce_1 + d_0 e_1 \\ &= e_{m+1} + d_{m-1}e_m + d_{m-2}e_{m-1} + \cdots + d_1 e_2 + d_0 e_1. \end{aligned}$$

This says that $e_1, e_2, \dots, e_n$ are linearly dependent, a contradiction.

An effective method of computing minimal polynomials is given in the next section in conjunction with the discussion of Jordan canonical forms of square matrices.

Problems

1. Let $A = \left(\begin{smallmatrix} 1 & 2 \\ 3 & 1 \end{smallmatrix}\right)$ and $f(x) = 3x^2 - 5x - 2$. Find $f(A)$.

2. Find the characteristic and minimal polynomials of the matrices
$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

3. Find the characteristic and minimal polynomials of the matrices
$$\begin{pmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{pmatrix}, \quad \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}.$$

4. Find a nonzero polynomial $p(x)$ such that $p(A) = 0$, where
$$A = \begin{pmatrix} 2 & -2 & 0 \\ -2 & 1 & -2 \\ 0 & 2 & 0 \end{pmatrix}.$$

5. Compute $\det(AB - A)$ and $f(A)$, where $f(\lambda) = \det(\lambda B - A)$ and
$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

6. Let $A \in \mathbb{M}_n$. Show that there exists a polynomial $f(x)$ with real coefficients such that $f(A) = 0$.

7. Let A and B be $n \times n$ matrices, and let $f(\lambda) = \det(\lambda I - B)$. Show that $f(A)$ is invertible if and only if A and B have no common eigenvalues.

8. Let A and B be square matrices of the same size. Show that if A and B are similar, then so are $f(A)$ and $f(B)$ for any polynomial f.

9. Let A and B be $n \times n$ matrices. If A and B are similar, show that $f(A) = 0$ if and only if $f(B) = 0$. Is the converse true?

10. Show that $\operatorname{rank}(AB) = n - 2$ if A and B are n-square upper-triangular matrices of rank $n - 1$ with diagonal entries zero.

11. Let $A_1, \ldots, A_m$ be upper-triangular matrices in $\mathbb{M}_n$. If they all have diagonal entries zero, show that $A_1 \cdots A_m = 0$ when $m \geq n$.

12. Explain what is wrong with the following proof of the Cayley–Hamilton theorem. Because $p(\lambda) = \det(\lambda I - A)$, plugging A for λ directly in both sides gives $p(A) = \det(A - A) = 0$.

13. As is known, for square matrices A and B of the same size, AB and BA have the same characteristic polynomial (Section 2.4). Do they have the same minimal polynomial?

14. Let $f(x)$ be a polynomial and let λ be an eigenvalue of a square matrix A. Show that if $f(A) = 0$, then $f(\lambda) = 0$.

15. Let $v \in \mathbb{C}^n$ and $A \in \mathbb{M}_n$. If $f(\lambda)$ is the monic polynomial with the smallest degree such that $f(A)v = 0$, show that $f(\lambda)$ divides $m_A(\lambda)$.

16. Let $c_0, c_1, \ldots, c_{n-1} \in \mathbb{C}$ and let C and D be, respectively,

$$\begin{pmatrix} 0 & 0 & \ldots & -c_0 \\ 1 & 0 & \ldots & -c_1 \\ \vdots & \ddots & & \vdots \\ 0 & \ldots & 1 & -c_{n-1} \end{pmatrix}, \quad \begin{pmatrix} -c_{n-1} & -c_{n-2} & \ldots & -c_0 \\ 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 1 & 0 \end{pmatrix}.$$

Show that $SCS = D^T$, where S is the backward identity; that is,

$$S = \begin{pmatrix} 0 & & & 1 \\ & & \cdot & \\ & \cdot & & \\ 1 & & & 0 \end{pmatrix}.$$

Show that the matrices C, C^T, D, and D^T all have the polynomial

$$p(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$$

as their characteristic and minimal polynomials.

17. Let C be the companion matrix of the polynomial $p(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$. Show that C is nonsingular if and only if $c_0 \neq 0$. In case where C is nonsingular, find the inverse of C.

18. Let $A, B, C \in \mathbb{M}_n$. If $X \in \mathbb{M}_n$ satisfies $AX^2 + BX + C = 0$ and if λ is an eigenvalue of X, show that $\lambda^2 A + \lambda B + C$ is singular.

19. Let $A \in \mathbb{M}_n$, $p(x) = \det(xI - A)$, and $P(x) = \operatorname{adj}(xI - A)$. Show that every entry in the matrix $x^k P(x) - P(x)A^k$ is divisible by $p(x)$ for $k = 1, 2, \ldots$.

20. Let A and B be n-square complex matrices. Show that

$$AX - XB = 0 \quad \Rightarrow \quad f(A)X - Xf(B) = 0$$

for every polynomial f. In addition, if A and B have no common eigenvalues, then $AX - XB = 0$ has only the solution $X = 0$.

21. Let A and B be $2n \times 2n$ matrices partitioned conformally as

$$A = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

If $AB = BA$ and A_{11} and A_{22} have no common eigenvalue, show that

$$B_{12} = B_{21} = 0.$$

22. Show that for any nonsingular matrix A, matrices A^{-1} and $\operatorname{adj}(A)$ can be expressed as polynomials in A.

23. Express J^{-1} as a polynomial in J, where

$$J = \begin{pmatrix} 1 & 1 & & 0 \\ & 1 & \ddots & \\ & & \ddots & 1 \\ 0 & & & 1 \end{pmatrix}.$$

24. For a square matrix X, we denote $e^X = \sum_{k=0}^{\infty} \frac{1}{k!} X^k$. Let A be a square matrix with all its eigenvalues equal to λ. Show that

$$e^{tA} = e^{t\lambda} \sum_{k=0}^{n-1} \frac{t^k}{k!} (A - \lambda I)^k, \quad t \in \mathbb{C}.$$

In particular, if A is a 3×3 matrix having eigenvalues λ, λ, λ, then

$$e^{tA} = e^{t\lambda} \big(I + t(A - \lambda I) + \frac{1}{2} t^2 (A - \lambda I)^2 \big).$$

25. Let X and Y be positive semidefinite matrices of the same size such that $X \geq Y$, i.e., $X - Y \geq 0$. Does it necessarily follow that $e^X \geq e^Y$?

. ——————— ⊙ ——————— .

3.4 Jordan Canonical Forms

We saw in Section 3.2 that a square matrix is similar (and even unitarily similar) to an upper-triangular matrix. We now discuss the upper-triangular matrices and give simpler structures.

The main theorem of this section is the *Jordan decomposition*, which states that every square complex matrix is similar (not necessarily unitarily similar) to a direct sum of Jordan blocks, referred to as *Jordan canonical form* or simply *Jordan form*. A *Jordan block* is a square matrix in the form

$$J_x = \begin{pmatrix} x & 1 & & 0 \\ & x & \ddots & \\ & & \ddots & 1 \\ 0 & & & x \end{pmatrix}. \tag{3.2}$$

For this purpose we introduce λ-matrices as a tool and use elementary operations to bring λ-matrices to so-called standard forms. We then show that two matrices A and B in $\mathbb{M}_n$ are similar if and only if their λ-matrices $\lambda I - A$ and $\lambda I - B$ can be brought to the same standard form. Thus, a square matrix A is similar to its Jordan form that is determined by the standard form of the λ-matrix $\lambda I - A$.

To proceed, a *λ-matrix* is a matrix whose entries are complex polynomials in λ. For instance,

$$\begin{pmatrix} 1 & \lambda^2 - \sqrt{2} & (\lambda+1)^2 \\ \frac{1}{2}\lambda - 1 & 0 & 1 - 2\lambda - \lambda^2 \\ -1 & 1 & \lambda - i \end{pmatrix}$$

is a λ-matrix, for every entry is a polynomial in λ (or a constant).

We perform operations (i.e., addition and multiplication) on λ-matrices in the same way as we do for numerical matrices. Of course, here the polynomials obey their usual rules of operations. Note that division by a nonzero polynomial is not permitted.

Note that the minimal polynomial of J_x of size t, say, in (3.2) is

$$m(\lambda) = (\lambda - x)^t.$$

Elementary operations on λ-matrices are similar to those on numerical matrices. *Elementary λ-matrices* and *invertible λ-matrices* are similarly defined as those of numerical matrices.

Any square numerical matrix can be brought into a diagonal matrix with 1 and 0 on the main diagonal by elementary operations. Likewise, λ-matrices can be brought into the *standard form*

$$\begin{pmatrix} d_1(\lambda) & & & & & 0 \\ & \ddots & & & & \\ & & d_k(\lambda) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0 \end{pmatrix}, \tag{3.3}$$

where $d_i(\lambda)|d_{i+1}(\lambda)$, $i = 1, \dots, k-1$, and each $d_i(\lambda)$ is 1 or monic. Therefore, for any λ-matrix $A(\lambda)$ there exist elementary λ-matrices $P_s(\lambda)$, $\dots, P_1(\lambda)$ and $Q_1(\lambda), \dots, Q_t(\lambda)$ such that

$$P_s(\lambda) \cdots P_1(\lambda) A(\lambda) Q_1(\lambda) \cdots Q_t(\lambda) = D(\lambda)$$

is in the standard form (3.3).

If $A(\lambda)$ is an invertible λ-matrix; that is, $B(\lambda)A(\lambda) = I$ for some λ-matrix $B(\lambda)$ of the same size, then, by taking determinants, we see that $\det A(\lambda)$ is a nonzero constant. Conversely, if $\det A(\lambda)$ is a nonzero constant, then $(\det A(\lambda))^{-1} \operatorname{adj}(A(\lambda))$ is also a λ-matrix and it is the inverse of $A(\lambda)$. Moreover, a square λ-matrix is invertible if and only if its standard form (3.3) is the identity matrix and if and only if it is a product of elementary λ-matrices.

For the λ-matrix $\lambda I - A$, $A \in \mathbb{M}_n$, we have $k = n$ and $D(\lambda) = \operatorname{diag}(d_1(\lambda), \dots, d_n(\lambda))$. The $d_i(\lambda)$ are called the *invariant factors* of A, and the divisors of $d_i(\lambda)$ factored into the form $(\lambda - x)^t$ for some constant x and positive integer t are the *elementary divisors* of A.

To illustrate this, look at the example

$$A = \begin{pmatrix} -1 & 0 & 0 \\ 1 & 1 & 2 \\ 3 & 0 & 1 \end{pmatrix}.$$

We perform elementary operations on the λ-matrix

$$\lambda I - A = \begin{pmatrix} \lambda+1 & 0 & 0 \\ -1 & \lambda-1 & -2 \\ -3 & 0 & \lambda-1 \end{pmatrix}.$$

Interchange row 1 and row 2 times -1 to get a 1 for the $(1,1)$ position:

$$\begin{pmatrix} 1 & 1-\lambda & 2 \\ \lambda+1 & 0 & 0 \\ -3 & 0 & \lambda-1 \end{pmatrix}.$$

Add row 1 times $-(\lambda+1)$ and 3 to rows 2 and 3, respectively, to get 0 below 1:

$$\begin{pmatrix} 1 & 1-\lambda & 2 \\ 0 & (\lambda-1)(\lambda+1) & -2(\lambda+1) \\ 0 & -3(\lambda-1) & \lambda+5 \end{pmatrix}.$$

Add row 3 times 2 to row 2 to get a nonzero number 8:

$$\begin{pmatrix} 1 & 1-\lambda & 2 \\ 0 & \lambda^2-6\lambda+5 & 8 \\ 0 & -3(\lambda-1) & \lambda+5 \end{pmatrix}.$$

Interchange column 2 and column 3 to get a nonzero number for the $(2,2)$ position:

$$\begin{pmatrix} 1 & 2 & 1-\lambda \\ 0 & 8 & \lambda^2-6\lambda+5 \\ 0 & \lambda+5 & -3(\lambda-1) \end{pmatrix}.$$

Subtract the second row times $\frac{1}{8}(\lambda+5)$ from row 3 to get

$$\begin{pmatrix} 1 & 2 & 1-\lambda \\ 0 & 8 & \lambda^2-6\lambda+5 \\ 0 & 0 & -\frac{1}{8}(\lambda-1)^2(\lambda+1) \end{pmatrix},$$

which gives the standard form at once (by column operations)

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (\lambda+1)(\lambda-1)^2 \end{pmatrix}.$$

Thus, the invariant factors of A are

$$d_1(\lambda) = 1, \quad d_2(\lambda) = 1, \quad d_3(\lambda) = (\lambda + 1)(\lambda - 1)^2,$$

and the elementary divisors of A are

$$\lambda + 1, \quad (\lambda - 1)^2.$$

Note that the matrix, a direct sum of two Jordan blocks,

$$J = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = (-1) \oplus \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

has the same invariant factors and elementary divisors as A.

In general, each elementary divisor $(\lambda - x)^t$ corresponds to a Jordan block in the form (3.2). Consider all the elementary divisors of a matrix A, find all the corresponding Jordan blocks, and form a direct sum of them. A profound conclusion is that A is similar to this direct sum. To this end, we need to show a fundamental theorem.

Theorem 3.11 *Let A and B be n-square complex matrices. Then A and B are similar if and only if $\lambda I - A$ and $\lambda I - B$ have the same standard form. Equivalently, there exist λ-matrices $P(\lambda)$ and $Q(\lambda)$ that are products of elementary λ-matrices such that*

$$P(\lambda)(\lambda I - A)Q(\lambda) = \lambda I - B.$$

This implies our main theorem on the Jordan canonical form, which is one of the most useful results in linear algebra and matrix theory. The theorem itself is much more important than its proof. We sketch the proof of Theorem 3.11 as follows.

Proof outline. If A and B are similar, then there exists an invertible complex matrix P such that $PAP^{-1} = B$. It follows that

$$P(\lambda I - A)P^{-1} = \lambda I - B.$$

To show the other way, let $P(\lambda)$ and $Q(\lambda)$ be invertible λ-matrices such that (we put $Q(\lambda)$ for $Q(\lambda)^{-1}$ on the right for convenience)

$$P(\lambda)(\lambda I - A) = (\lambda I - B)Q(\lambda). \tag{3.4}$$

Write (Problem 4)

$$P(\lambda) = (\lambda I - B)P_1(\lambda) + P, \quad Q(\lambda) = Q_1(\lambda)(\lambda I - A) + Q,$$

where $P_1(\lambda)$, $Q_1(\lambda)$ are λ-matrices, and P, Q are numerical matrices.

Identity (3.4) implies $P_1(\lambda) - Q_1(\lambda) = 0$ by considering the degree of $P_1(\lambda) - Q_1(\lambda)$. It follows that $Q = P$, and thus $PA = BP$.

It remains to show that P is invertible. Assume $R(\lambda)$ is the inverse of $P(\lambda)$, or $P(\lambda)R(\lambda) = I$. Write $R(\lambda) = (\lambda I - A)R_1(\lambda) + R$, where R is a numerical matrix. With $PA = BP$, $I = P(\lambda)R(\lambda)$ gives

$$I = (\lambda I - B)T(\lambda) + PR, \tag{3.5}$$

where

$$T(\lambda) = P_1(\lambda)(\lambda I - A)R_1(\lambda) + P_1(\lambda)R + PR_1(\lambda).$$

By considering the degree of both sides of (3.5), $T(\lambda)$ must be zero. Therefore, $I = PR$ and hence P is nonsingular. ∎

Based on the earlier discussions, we conclude our main result.

Theorem 3.12 (Jordan Decomposition) *Let A be a square complex matrix. Then there exists an invertible matrix P such that*

$$P^{-1}AP = J_1 \oplus \cdots \oplus J_s,$$

where the J_i are the Jordan blocks of A with the eigenvalues of A on the diagonal. The Jordan blocks are uniquely determined by A.

The uniqueness of the Jordan decomposition of A up to permutations of the diagonal Jordan blocks follows from the uniqueness of the standard form (3.3) of $\lambda I - A$. Two different sets of Jordan blocks will result in two different standard forms (3.3).

To find the minimal polynomial of a given matrix $A \in \mathbb{M}_n$, reduce $\lambda I - A$ by elementary operations to a standard form with invariant factors $d_1(\lambda), \ldots, d_n(\lambda)$, $d_i(\lambda) | d_{i+1}(\lambda)$, for $i = 1, \ldots, n-1$. Note that similar matrices have the same minimal polynomial (Theorem 3.10). Thus, $d_n(\lambda)$ is the minimal polynomial of A, because it is the minimal polynomial of the Jordan canonical form of A (Problem 12).

Theorem 3.13 *Let $p(\lambda)$ and $m(\lambda)$ be, respectively, the characteristic and minimal polynomials of matrix $A \in \mathbb{M}_n$. Let $d_1(\lambda), \ldots, d_n(\lambda)$ be the invariant factors of A, where $d_i(\lambda)|d_{i+1}$, $i = 1, \ldots, n-1$. Then*

$$p(\lambda) = d_1(\lambda) \cdots d_n(\lambda), \quad m(\lambda) = d_n(\lambda).$$

In the earlier example preceding Theorem 3.11, the characteristic and minimal polynomials of A are the same, and they are equal to

$$p(\lambda) = m(\lambda) = (\lambda + 1)(\lambda - 1)^2 = \lambda^3 - \lambda^2 - \lambda + 1.$$

Problems

1. Find the invariant factors, elementary divisors, characteristic and minimal polynomials, and the Jordan canonical form of the matrix

$$A = \begin{pmatrix} 3 & 1 & -3 \\ -7 & -2 & 9 \\ -2 & -1 & 4 \end{pmatrix}.$$

2. Show that A and B are similar but not unitarily similar, where

$$A = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix}.$$

What is the Jordan canonical form J of A and B? Can one find an invertible matrix P such that

$$P^{-1}AP = P^{-1}BP = J?$$

3. Are the following matrices similar? Why?

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

4. Let $A \in \mathbb{M}_n$. If $P(\lambda)$ is an n-square λ-matrix, show that there exist a λ-matrix $S(\lambda)$ and a numerical matrix T such that

$$P(\lambda) = (\lambda I - A)S(\lambda) + T.$$

[Hint: Write $P(\lambda) = \lambda^m P_m + \lambda^{m-1} P_{m-1} + \cdots + \lambda P_1 + P_0$.]

5. Show that a λ-matrix is invertible if and only if it is a product of elementary λ-matrices.

6. Show that two matrices are similar if and only if they have the same set of Jordan blocks, counting the repeated ones.

7. Find the invariant factors, elementary divisors, and characteristic and minimal polynomials for each of the following matrices.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} \lambda & 0 & 1 & 0 \\ 0 & \lambda & 0 & 1 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \end{pmatrix}.$$

8. Find the Jordan canonical form of the matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

9. Let A be a square complex matrix with invariant factors

$$1, \quad \lambda(\lambda-2), \quad \lambda^3(\lambda-2).$$

Answer the following questions.

(a) What is the characteristic polynomial of A?

(b) What is the minimal polynomial of A?

(c) What are the elementary divisors of A?

(d) What is the size of A?

(e) What is the rank of A?

(f) What is the trace of A?

(g) What is the Jordan form of A?

10. Let J be a Jordan block. Find the Jordan forms of J^{-1} (if it exists) and J^2.

11. Show that every Jordan block J is similar to J^T via S:
$$S^{-1}JS = J^T,$$
where S is the backward identity matrix, that is, $s_{i,n-i+1} = 1$ for $i = 1, 2, \dots, n$, and 0 elsewhere.

12. Show that the last invariant factor $d_n(\lambda)$ in the standard form of $\lambda I - A$ is the minimal polynomial of $A \in \mathbb{M}_n$.

13. Let $p(x) = x^n + c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + \cdots + c_1 x + c_0$. What are the invariant factors of the companion matrix of $p(x)$? Show that the minimal polynomial of the companion matrix of $p(x)$ is $p(x)$.

14. If J is a Jordan block such that $J^2 = J$, show that $J = 1$ or 0.

15. Let A be an $n \times n$ matrix. Show that $\operatorname{rank}(A^2) = \operatorname{rank}(A)$ implies $\operatorname{rank}(A^k) = \operatorname{rank}(A)$ for any integer $k > 0$ and $\mathbb{C}^n = \operatorname{Im} A \oplus \operatorname{Ker} A$.

16. Let A be an $n \times n$ matrix. Show that $\operatorname{rank}(A^k) = \operatorname{rank}(A^{k+1})$ for some positive integer $k \le n$ and that $\operatorname{rank}(A^k) = \operatorname{rank}(A^m)$ for all positive integer $m > k$. In particular, $\operatorname{rank}(A^n) = \operatorname{rank}(A^{n+1})$.

17. Show that every matrix $A \in \mathbb{M}_n$ can be written as $A = B + C$, where $C^k = 0$ for some integer k, B is diagonalizable, and $BC = CB$.

18. Let A be a square complex matrix. If $Ax = 0$ whenever $A^2x = 0$, show that A does not have any Jordan block of order more than 1 corresponding to eigenvalue 0.

19. Show that if matrix A has all eigenvalues equal to 1, then A^k is similar to A for every positive integer k. Discuss the converse.

20. Show that the dimension of the vector space of all the polynomials in A is equal to the degree of the minimal polynomial of A.

21. Let A be an $n \times n$ matrix such that $A^k v = 0$ and $A^{k-1}v \ne 0$ for some vector v and positive integer k. Show that $v, Av, \dots, A^{k-1}v$ are linearly independent. What is the Jordan form of A?

22. Let $A \in \mathbb{M}_n$ be a Jordan block. Show that there exists a vector v such that $v, Av, \dots, A^{n-1}v$ constitute a basis for $\mathbb{C}^n$.

23. Show that the characteristic polynomial coincides with the minimal polynomial for $A \in \mathbb{M}_n$ if and only if $v, Av, \dots, A^{n-1}v$ are linearly independent for some vector $v \in \mathbb{C}^n$. What can be said about the Jordan form (or Jordan blocks) of A?

24. Let A be an n-square complex matrix. Show that for any nonzero vector $v \in \mathbb{C}^n$, there exists an eigenvector u of A that is contained in the span of $v, Av, A^2v, \ldots$. [Hint: $v, Av, A^2v, \ldots, A^kv$ are linearly dependent for some k. Find a related polynomial then factor it out.]

25. Let A be an n-square complex matrix with the characteristic polynomial factored over the complex field $\mathbb{C}$ as

$$\det(\lambda I - A) = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_s)^{r_s},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_s$ are the distinct eigenvalues of A. Show that the following statements are equivalent.

(a) A is diagonalizable; namely, A is similar to a diagonal matrix.

(b) A has n linearly independent eigenvectors.

(c) All the elementary divisors of $\lambda I - A$ are linear.

(d) The minimal polynomial of A has no repeated zeros.

(e) $\operatorname{rank}(\lambda I - A) = \operatorname{rank}(\lambda I - A)^2$ for every eigenvalue λ.

(f) $\operatorname{rank}(cI - A) = \operatorname{rank}(cI - A)^2$ for every complex number c.

(g) $(\lambda I - A)x = 0$ and $(\lambda I - A)^2 x = 0$ have the same solution space for every eigenvalue λ.

(h) $(cI - A)x = 0$ and $(cI - A)^2 x = 0$ have the same solution space for every complex number c.

(i) $\dim V_{\lambda_i} = r_i$ for each eigenspace V_{λ_i} of eigenvalue λ_i.

(j) $\operatorname{rank}(\lambda_i I - A) = n - r_i$ for every eigenvalue λ_i.

(k) $\operatorname{Im}(\lambda I - A) \cap \operatorname{Ker}(\lambda I - A) = \{0\}$ for every eigenvalue λ.

(l) $\operatorname{Im}(cI - A) \cap \operatorname{Ker}(cI - A) = \{0\}$ for every complex number c.

26. Let $\mathcal{A}$ be a linear transformation on a finite-dimensional vector space. Let λ be an eigenvalue of $\mathcal{A}$. Show that each subspace $\operatorname{Ker}(\lambda\mathcal{I} - \mathcal{A})^k$, where k is a positive integer, is invariant under $\mathcal{A}$, and that

$$\operatorname{Ker}(\lambda\mathcal{I} - \mathcal{A}) \subseteq \operatorname{Ker}(\lambda\mathcal{I} - \mathcal{A})^2 \subseteq \operatorname{Ker}(\lambda\mathcal{I} - \mathcal{A})^3 \subseteq \cdots.$$

Conclude that for some positive integer m,

$$\operatorname{Ker}(\lambda\mathcal{I} - \mathcal{A})^m = \operatorname{Ker}(\lambda\mathcal{I} - \mathcal{A})^{m+1} = \cdots$$

and that

$$\cup_{k=1}^{\infty} \operatorname{Ker}(\lambda\mathcal{I} - \mathcal{A})^k = \operatorname{Ker}(\lambda\mathcal{I} - \mathcal{A})^m.$$

3.5 The Matrices A^T, $\overline{A}$, A^*, A^TA, A^*A, and $\overline{A}A$

The matrices associated with a matrix A and often encountered are

$$A^T, \quad \overline{A}, \quad A^*, \quad A^TA, \quad A^*A, \quad \overline{A}A,$$

where T, $-$, and $*$ mean transpose, conjugate, and transpose conjugate, respectively. All these matrices, except A^TA, have the same rank as A:

$$\operatorname{rank}(A) = \operatorname{rank}(A^T) = \operatorname{rank}(\overline{A}) = \operatorname{rank}(A^*) = \operatorname{rank}(A^*A).$$

The last identity is due to the fact that the equation systems

$$(A^*A)x = 0 \quad \text{and} \quad Ax = 0$$

have the same solution space.

Theorem 3.14 *Let A be an n-square complex matrix. Then*

1. *A is similar to its transpose A^T.*
2. *A is similar to A^* (equivalently $\overline{A}$) if and only if the Jordan blocks of the nonreal eigenvalues of A occur in conjugate pairs.*
3. *A^*A is similar to AA^*.*
4. *$\overline{A}A$ is similar to $A\overline{A}$.*

Proof. For (1), recall from Theorem 3.11 that two matrices X and Y are similar if and only if $\lambda I - X$ and $\lambda I - Y$ have the same standard form. It is obvious that the matrices $\lambda I - A$ and $\lambda I - A^T$ have the same standard form. Thus, A and A^T are similar. An alternative way to show (1) is to verify that for every Jordan block J,

$$SJS^{-1} = J^T,$$

where S is the backward identity matrix (Problem 11, Section 3.4).

For (2), let $J_1, \ldots, J_k$ be the Jordan blocks of A and let

$$P^{-1}AP = J_1 \oplus \cdots \oplus J_k$$

for some invertible matrix P. Taking the transpose conjugate gives

$$P^*A^*(P^*)^{-1} = J_1^* \oplus \cdots \oplus J_k^*.$$

The right-hand side, by (1), is similar to

$$\overline{J_1} \oplus \cdots \oplus \overline{J_k}.$$

Thus, if A and A^* are similar, then

$$J_1 \oplus \cdots \oplus J_k \quad \text{and} \quad \overline{J_1} \oplus \cdots \oplus \overline{J_k}$$

are similar. It follows by the uniqueness of Jordan decomposition that the Jordan blocks of nonreal complex eigenvalues of A must occur in conjugate pairs (Problem 6, Section 3.4).

For sufficiency, we may consider the special case

$$A = J \oplus \overline{J} \oplus R, \tag{3.6}$$

where J and R are Jordan blocks, J is complex, and R is real. Then

$$A^* = \overline{J}^T \oplus J^T \oplus R^T,$$

which is, by permutation, similar to

$$J^T \oplus \overline{J}^T \oplus R^T. \tag{3.7}$$

Using (1), (3.6) and (3.7) give the similarity of A^* and A.

(3) is by a singular value decomposition of A.

We have left to show (4) that $\overline{A}A$ is similar to $A\overline{A}$. It suffices to show that $A\overline{A}$ and $\overline{A}A$ have the same Jordan decomposition (blocks).

The matrix identity

$$\begin{pmatrix} I & -A \\ 0 & I \end{pmatrix} \begin{pmatrix} A\overline{A} & 0 \\ \overline{A} & 0 \end{pmatrix} \begin{pmatrix} I & A \\ 0 & I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \overline{A} & \overline{A}A \end{pmatrix}$$

gives the similarity of the block matrices

$$\begin{pmatrix} A\overline{A} & 0 \\ \overline{A} & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 \\ \overline{A} & \overline{A}A \end{pmatrix}.$$

Thus, the nonsingular Jordan blocks of $A\overline{A}$ and $\overline{A}A$ are identical. (In general, this is true for AB and BA. See Problem 12.) On the other hand, the singular Jordan blocks of $A\overline{A}$ and $\overline{A}A = \overline{A\overline{A}}$ are obviously the same. This concludes that $A\overline{A}$ and $\overline{A}A$ are similar. ∎

Following the discussion of the case where A is similar to A^* in the proof, one may obtain a more profound result. Consider the matrix with Jordan blocks of conjugate pairs

$$\begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \overline{\lambda} & 1 \\ 0 & 0 & 0 & \overline{\lambda} \end{pmatrix},$$

which is similar via permutation to

$$\begin{pmatrix} \lambda & 0 & 1 & 0 \\ 0 & \overline{\lambda} & 0 & 1 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \overline{\lambda} \end{pmatrix} = \begin{pmatrix} C(\lambda) & I \\ 0 & C(\lambda) \end{pmatrix},$$

where

$$C(\lambda) = \begin{pmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{pmatrix}.$$

If $\lambda = a + bi$ with $a, b \in \mathbb{R}$, then we have by computation

$$\begin{pmatrix} -i & -i \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{pmatrix} \begin{pmatrix} -i & -i \\ 1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Thus, matrices

$$\begin{pmatrix} \lambda & 0 & 1 & 0 \\ 0 & \overline{\lambda} & 0 & 1 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \overline{\lambda} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a & b & 1 & 0 \\ -b & a & 0 & 1 \\ 0 & 0 & a & b \\ 0 & 0 & -b & a \end{pmatrix}$$

are similar. These observations lead to the following theorem.

Theorem 3.15 *A square matrix A is similar to A^* (equivalently $\overline{A}$) if and only if A is similar to a real matrix.*

As a result, $\overline{A}A$ is similar to a real matrix. The ideas of pairing Jordan blocks are often used in the similarity theory of matrices.

Problems

1. Let A be an $m \times n$ matrix. Prove or disprove
   (a) $\operatorname{rank}(A) = \operatorname{rank}(A^TA)$.
   (b) $\operatorname{rank}(A^*A) = \operatorname{rank}(AA^*)$.
   (c) A^TA is similar to AA^T.
2. Is it possible that $\overline{A}A = 0$ or $A^*A = 0$ for a nonzero $A \in \mathbb{M}_n$?
3. Let
$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}.$$
   Compute AB and BA. Find $\operatorname{rank}(AB)$ and $\operatorname{rank}(BA)$. What are the Jordan forms of AB and BA? Are AB and BA similar?
4. If the nonreal eigenvalues of a square matrix A occur in conjugate pairs, does it follow that A is similar to A^*?
5. Show that the characteristic polynomial of $A\overline{A}$ has only real coefficients. Conclude that the nonreal eigenvalues of $A\overline{A}$ must occur in conjugate pairs.
6. Let $A = B + iC \in \mathbb{M}_n$ be nonsingular, where B and C are real square matrices. Show that $\overline{A} = A^{-1}$ if and only if $BC = CB$ and $B^2 + C^2 = I$. Find the conditions on B and C if $A^T = \overline{A} = A^{-1}$.
7. Let $A \in \mathbb{M}_n$. Show that the following statements are equivalent.
   (a) A is similar to A^* (equivalently $\overline{A}$).
   (b) The elementary divisors occur in conjugate pairs.
   (c) The invariant factors of A are all real coefficients.

   Are they equivalent to the statement "$\det(\lambda I - A)$ is real coefficient"?
8. If $A \in \mathbb{M}_n$ has only real eigenvalues, show that A is similar to A^*.
9. Let A be a nonsingular matrix. When is A^{-1} similar to A?
10. Let $A \in \mathbb{M}_n$ and B be $m \times n$. Let M be the $(m+n)$-square matrix
$$M = \begin{pmatrix} A & 0 \\ B & 0 \end{pmatrix}.$$
   Show that the nonsingular Jordan blocks of A and M are identical.

11. Let A, B, C, and D be n-square complex matrices. If the matrices

$$\begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} C & D \\ 0 & 0 \end{pmatrix}$$

are similar, does it follow that A and C are similar?

12. Let A and B^T be $m \times n$ complex matrices. Show that the nonsingular Jordan blocks of AB and BA are identical. Conclude that AB and BA have the same nonzero eigenvalues, including multiplicity.

13. If A and $B \in \mathbb{M}_n$ have no common eigenvalues, show that the following two block matrices are similar for any $X \in \mathbb{M}_n$:

$$\begin{pmatrix} A & X \\ 0 & B \end{pmatrix}, \qquad \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

14. Let A, B, and C be matrices of appropriate sizes. Show that

$$AX - YB = C$$

for some matrices X and Y if and only if the block matrices

$$\begin{pmatrix} A & C \\ 0 & B \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

have the same rank.

15. Let A and B be n-square complex matrices and let

$$M = \begin{pmatrix} A & B \\ -\overline{B} & \overline{A} \end{pmatrix}.$$

Show that

(a) The characteristic polynomial of M is of real coefficients.

(b) The eigenvalues of M occur in conjugate pairs with eigenvectors in forms $\binom{x}{y} \in \mathbb{C}^{2n}$ and $\binom{-\bar{y}}{\bar{x}} \in \mathbb{C}^{2n}$.

(c) The eigenvectors in (b) are linearly independent.

(d) $\det M \geq 0$. In particular, $\det(I + A\overline{A}) \geq 0$ for any $A \in \mathbb{M}_n$.

CHAPTER 4

Numerical Ranges, Matrix Norms, and Special Operations

Introduction: This chapter is devoted to a few basic topics on matrices. We first study the numerical range and radius of a square matrix and matrix norms. We then introduce three important special matrix operations: the Kronecker product, the Hadamard product, and compound matrices.

4.1 Numerical Range and Radius

Let A be an $n \times n$ complex matrix. For $x = (x_1, \ldots, x_n)^T \in \mathbb{C}^n$, as usual, $\|x\| = (\sum_{i=1}^n |x_i|^2)^{1/2}$ is the norm of x (Section 1.4). The *numerical range*, also known as the *field of values*, of A is defined by

$$W(A) = \{x^* A x : \|x\| = 1, \ x \in \mathbb{C}^n\}.$$

For example, if

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

then $W(A)$ is the closed interval $[0, 1]$, and if

$$A = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix},$$

then $W(A)$ is the closed elliptical disc with foci at $(0, 0)$ and $(1, 0)$, minor axis 1, and major axis $\sqrt{2}$ (Problem 9).

F. Zhang, *Matrix Theory: Basic Results and Techniques*, Universitext,
DOI 10.1007/978-1-4614-1099-7_4,

One of the celebrated and fundamental results on numerical range is the Toeplitz–Hausdorff convexity theorem.

Theorem 4.1 (Toeplitz–Hausdorff) *The numerical range of a square matrix is a convex compact subset of the complex plane.*

Proof. For convexity, if $W(A)$ is a singleton, there is nothing to show. Suppose $W(A)$ has more than one point. We prove that the line segment joining any two distinct points in $W(A)$ lies in $W(A)$; that is, if $u, v \in W(A)$, then $tu + (1-t)v \in W(A)$ for all $t \in [0,1]$.

For any complex numbers α and β, it is easy to verify that

$$W(\alpha I + \beta A) = \{\alpha + \beta z : \, z \in W(A)\}.$$

Intuitively the convexity of $W(A)$ does not change under shifting, scaling, and rotation. Thus, we may assume that the two points to be considered are 0 and 1, and show that $[0,1] \subseteq W(A)$. Write

$$A = H + iK,$$

where

$$H = \frac{1}{2}(A + A^*) \quad \text{and} \quad K = \frac{1}{2i}(A - A^*)$$

are Hermitian matrices. Let x and y be unit vectors in $\mathbb{C}^n$ such that

$$x^*Ax = 0, \quad y^*Ay = 1.$$

It follows that x and y are linearly independent and that

$$x^*Hx = x^*Kx = y^*Ky = 0, \quad y^*Hy = 1.$$

We may further assume that x^*Ky has real part zero; otherwise, one may replace x with cx, $c \in \mathbb{C}$, and $|c| = 1$, so that cx^*Ky is 0 or a pure complex number without changing the value of x^*Ax.

Note that $tx + (1-t)y \neq 0$, $t \in [0,1]$. Define for $t \in [0,1]$

$$z(t) = \frac{1}{\|tx + (1-t)y\|^2}(tx + (1-t)y).$$

Then $z(t)$ is a unit vector. It is easy to compute that for all $t \in [0,1]$

$$z(t)^*Kz(t) = 0.$$

The convexity of $W(A)$ then follows, for

$$\{z(t)^*Az(t) : 0 \le t \le 1\} = [0, 1].$$

The compactness of $W(A)$, meaning the boundary is contained in $W(A)$, is seen by noting that $W(A)$ is the range of the continuous function $x \mapsto x^*Ax$ on the compact set $\{x \in \mathbb{C}^n : \|x\| = 1\}$. (A continuous function maps a compact set to a compact set.) ∎

When considering the smallest disc centered at the origin that covers the numerical range, we associate with $W(A)$ a number

$$w(A) = \sup\{|z| : z \in W(A)\} = \sup_{\|x\|=1} |x^*Ax|$$

and call it the *numerical radius* of $A \in \mathbb{M}_n$. Note that the "sup" can be attained by some $z \in W(A)$. It is immediate that for any $x \in \mathbb{C}^n$

$$|x^*Ax| \le w(A)\|x\|^2. \tag{4.1}$$

We now make comparisons of the numerical radius $w(A)$ to the largest eigenvalue $\rho(A)$ in absolute value, or the *spectral radius*, i.e.,

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

and to the largest singular value $\sigma_{\max}(A)$, also called the *spectral norm*. It is easy to see (Problem 7) that

$$\sigma_{\max}(A) = \sup_{\|x\|=1} \|Ax\| = \sup_{x \ne 0} \frac{\|Ax\|}{\|x\|}$$

and that for every $x \in \mathbb{C}^n$

$$\|Ax\| \le \sigma_{\max}(A)\|x\|.$$

Theorem 4.2 *Let A be a square complex matrix. Then*

$$\rho(A) \le w(A) \le \sigma_{\max}(A) \le 2w(A).$$

Proof. Let λ be the eigenvalue of A such that $\rho(A) = |\lambda|$, and let u be a unit eigenvector corresponding to λ. Then

$$\rho(A) = |\lambda u^*u| = |u^*Au| \le w(A).$$

The second inequality follows from the Cauchy–Schwarz inequality

$$|x^*Ax| = |(Ax, x)| \le \|Ax\|\|x\|.$$

We next show that $\sigma_{\max}(A) \le 2w(A)$. It can be verified that

$$\begin{aligned} 4(Ax, y) &= \Big(A(x+y), x+y\Big) - \Big(A(x-y), x-y\Big) \\ &\quad + i\Big(A(x+iy), x+iy\Big) - i\Big(A(x-iy), x-iy\Big). \end{aligned}$$

Using (4.1), it follows that

$$\begin{aligned} 4|(Ax, y)| &\le w(A)(\|x+y\|^2 + \|x-y\|^2 \\ &\quad + \|x+iy\|^2 + \|x-iy\|^2) \\ &= 4w(A)(\|x\|^2 + \|y\|^2). \end{aligned}$$

Thus, for any unit x and y in $\mathbb{C}^n$, we have

$$|(Ax, y)| \le 2w(A).$$

The inequality follows immediately from Problem 7. ∎

Theorem 4.3 *Let $A \in \mathbb{M}_n$. Then $\lim_{k\to\infty} A^k = 0$ if and only if $\rho(A) < 1$; that is, all the eigenvalues of A have moduli less than 1.*

Proof. Let $A = P^{-1}TP$ be a Jordan decomposition of A, where P is invertible and T is a direct sum of Jordan blocks with the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A on the main diagonal. Then $A^k = P^{-1}T^kP$ and $\rho(A^k) = (\rho(A))^k$. Thus, if A^k tends to zero, so does T^k. It follows that $\lambda^k \to 0$ as $k \to 0$ for every eigenvalue of A. Therefore $\rho(A) < 1$. Conversely, suppose $\rho(A) < 1$. We show that $A^k \to 0$ as $k \to \infty$. It suffices to show that $J^k \to 0$ as $k \to \infty$ for each Jordan block J.

Suppose J is an $m \times m$ Jordan block:

$$J = \begin{pmatrix} \lambda & 1 & \ldots & 0 \\ 0 & \lambda & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \ldots & 0 & \lambda \end{pmatrix}.$$

Upon computation, we have

$$J^k = \begin{pmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \dots & \binom{k}{m-1}\lambda^{k-m+1} \\ 0 & \lambda^k & \ddots & \vdots \\ \vdots & \ddots & \ddots & \binom{k}{1}\lambda^{k-1} \\ 0 & \dots & 0 & \lambda^k \end{pmatrix}.$$

Recall from calculus that for any constants l and $\lambda < 1$

$$\lim_{k\to\infty} \binom{k}{l} \lambda^k = 0.$$

It follows that J^k, thus A^k, converges to 0 as $k \to \infty$. ■

Problems

1. Find a nonzero matrix A so that $\rho(A) = 0$.
2. Find the eigenvalues, singular values, numerical radius, spectral radius, spectral norm, and numerical range for each of the following:
$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$
3. Let A be an n-square complex matrix. Show that the numerical radius, spectral radius, spectral norm, and numerical range are unitarily invariant. That is, for instance, $w(U^*AU) = w(A)$ for any n-square unitary matrix U.
4. Show that the diagonal entries and the eigenvalues of a square matrix are contained in the numerical range of the matrix.
5. Let $A \in \mathbb{M}_n$. Show that $\frac{1}{n}\operatorname{tr} A$ is contained in $W(A)$. Conclude that for any nonsingular $P \in \mathbb{M}_n$, $W(P^{-1}AP - PAP^{-1})$ contains 0.
6. Let A be a square complex matrix. Show that $\frac{\|Ax\|}{\|x\|}$ is constant for all $x \neq 0$ if and only if all the singular values of A are identical.
7. Let A be a complex matrix. Show that
$$\sigma_{\max}(A) = \sqrt{\rho(A^*A)} = \sup_{\|x\|=1} \|Ax\| = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=\|y\|=1} |(Ax, y)|.$$

8. Show that for any square matrices A and B of the same size,

$$\sigma_{\max}(AB) \le \sigma_{\max}(A)\sigma_{\max}(B),$$

and

$$\sigma_{\max}(A+B) \le \sigma_{\max}(A) + \sigma_{\max}(B).$$

9. Show that the numerical range of $\left(\begin{smallmatrix} 0 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ is a closed elliptical disc.

10. Take

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

and let $B = A^2$. Show that $w(A) < 1$. Find $w(B)$ and $w(AB)$.

11. Show that the numerical range of a normal matrix is the *convex hull* of its eigenvalues. That is, if $A \in \mathbb{M}_n$ is a normal matrix with eigenvalues $\lambda_1, \dots, \lambda_n$, then

$$W(A) = \{t_1\lambda_1 + \cdots + t_n\lambda_n : \; t_1 + \cdots + t_n = 1, \text{ each } t_i \ge 0\}.$$

12. Show that $W(A)$ is a polygon inscribed in the unit circle if A is unitary, and that $W(A) \subseteq \mathbb{R}$ if A is Hermitian. What can be said about $W(A)$ if A is positive semidefinite?

13. Show that $w(A) = \rho(A) = \sigma_{\max}(A)$ if A is normal. Discuss the converse by considering

$$A = \operatorname{diag}(1, i, -1, -i) \oplus \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

14. Prove or disprove that for any n-square complex matrices A and B

    (a) $\rho(AB) \le \rho(A)\rho(B)$.

    (b) $w(AB) \le w(A)w(B)$.

    (c) $\sigma_{\max}(AB) \le \sigma_{\max}(A)\sigma_{\max}(B)$.

15. Let A be a square matrix. Show that for every positive integer k

$$w(A^k) \le \big(w(A)\big)^k.$$

Is it true in general that

$$w(A^{k+m}) \le w(A^k)\, w(A^m)?$$

4.2 Matrix Norms

A matrix may be assigned numerical items in various ways. In addition to determinant, trace, eigenvalues, singular values, numerical radius, and spectral radius, matrix norm is another important one.

Recall from Section 1.4 of Chapter 1 that vectors can be measured by their norms. If V is an inner product space, then the norm of a vector v in V is $\|v\| = \sqrt{(v,v)}$. The norm $\|\cdot\|$ on V satisfies

i. $\|v\| \geq 0$ with equality if and only if $v = 0$,

ii. $\|cv\| \leq |c|\|v\|$ for all scalars c and vectors v, and

iii. $\|u+v\| \leq \|u\| + \|v\|$ for all vectors u, v.

Like the norms for vectors being introduced to measure the magnitudes of vectors, norms for matrices are used to measure the "sizes" of matrices. We call a matrix function $\|\cdot\| : \mathbb{M}_n \mapsto \mathbb{R}$ a *matrix norm* if for all $A, B \in \mathbb{M}_n$ and $c \in \mathbb{C}$, the following conditions are satisfied:

1. $\|A\| \geq 0$ with equality if and only if $A = 0$,
2. $\|cA\| \leq |c|\|A\|$,
3. $\|A+B\| \leq \|A\| + \|B\|$, and
4. $\|AB\| \leq \|A\|\|B\|$.

We call $\|\cdot\|$ for matrices satisfying (1)–(3) a *matrix-vector norm*. In this book by a matrix norm we mean that all conditions (1)–(4) are met. Such a matrix norm is sometimes referred to as a *multiplicative matrix norm*. We use the notation $\|\cdot\|$ for both vector norm and matrix norm. Generally speaking, this won't cause confusion as one can easily tell from what is being studied.

If a matrix is considered as a linear operator on an inner product space V, a matrix *operator norm* $\|\cdot\|_{\mathrm{op}}$ can be *induced* as follows:

$$\|A\|_{\mathrm{op}} = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1} \|Ax\|.$$

From the previous section (Problems 7 and 8, Section 4.1), we see that the *spectral norm* is a matrix (operator) norm on $\mathbb{M}_n$ induced by the ordinary inner product on $\mathbb{C}^n$.

Matrices can be viewed as vectors in the matrix space $\mathbb{M}_n$ equipped with the inner product $(A, B) = \operatorname{tr}(B^*A)$. Matrices as vectors under the inner product have vector norms. One may check that this vector norm for matrices is also a (multiplicative) matrix norm.

Two observations on matrix norms follow: first, a matrix norm $\|\cdot\| : A \mapsto \|A\|$ is a continuous function on the matrix space $\mathbb{M}_n$ (Problem 2), and second, $\rho(\cdot) \le \|\cdot\|$ for any matrix norm $\|\cdot\|$. Reason: If $Ax = \lambda x$, where $x \neq 0$ and $\rho(A) = |\lambda|$, then, by letting X be the $n \times n$ matrix with all columns equal to the eigenvector x,

$$\rho(A)\|X\| = \|\lambda X\| = \|AX\| \le \|A\|\|X\|.$$

Nevertheless, the numerical radius $\rho(\cdot)$ is not a matrix norm. (Why?) The following result reveals a relation between the two.

Theorem 4.4 *Let $\|\cdot\|$ be a matrix norm. Then for every $A \in \mathbb{M}_n$*

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

Proof. The eigenvalues of A^k are the kth powers of those of A. Since spectral radius is dominated by norm, for every positive integer k,

$$(\rho(A))^k = \rho(A^k) \le \|A^k\| \quad \text{or} \quad \rho(A) \le \|A^k\|^{1/k}.$$

On the other hand, for any $\epsilon > 0$, let

$$A_\epsilon = \frac{1}{\rho(A) + \epsilon} A.$$

Then $\rho(A_\epsilon) < 1$. By Theorem 4.3, A_ϵ^k tends to 0 as $k \to \infty$. Thus, Because the norm is a continuous function, for k large enough,

$$\|A_\epsilon^k\| < 1 \quad \text{or} \quad \|A^k\| \le (\rho(A) + \epsilon)^k.$$

Therefore

$$\|A^k\|^{1/k} \le \rho(A) + \epsilon.$$

In summary, for any $\epsilon > 0$ and k large enough

$$\rho(A) \le \|A^k\|^{1/k} \le \rho(A) + \epsilon.$$

The conclusion follows immediately by letting ϵ approach 0. ∎

Now we turn our attention to an important class of matrix norms: *unitarily invariant norms*. We say a matrix (vector) norm $\|\cdot\|$ on $\mathbb{M}_n$ is *unitarily invariant* if for any $A \in \mathbb{M}_n$ and for all unitary $U, V \in \mathbb{M}_n$

$$\|UAV\| = \|A\|.$$

The spectral norm $\sigma_{\max} : A \mapsto \sigma_{\max}(A)$ is a matrix norm and it is unitarily invariant because $\sigma_{\max}(UAV) = \sigma_{\max}(A)$. The *Frobenius norm* (also known as the *Euclidean norm* or *Hilbert–Schmidt norm*) is the matrix norm induced by the inner product $(A, B) = \operatorname{tr}(B^*A)$ on the matrix space $\mathbb{M}_n$

$$\|A\|_F = \Big(\operatorname{tr}(A^*A)\Big)^{1/2} = \Big(\sum_{i,j=1}^{n} |a_{ij}|^2\Big)^{1/2}. \tag{4.2}$$

With $\sigma_i(A)$ denoting the singular values of A, we see that

$$\|A\|_F = \Big(\sum_{i=1}^{n} \sigma_i^2(A)\Big)^{1/2}.$$

Thus $\|A\|_F$ is uniquely determined by the singular values of A. Consequently, the Frobenius norm is a unitarily invariant matrix norm.

Actually, the spectral norm and the Frobenius norm belong to two larger families of unitarily invariant norms: the *Ky Fan k-norms* and the *Schatten p-norms*, which we study more in Chapter 10.

Ky Fan k-norm: Let $k \le n$ be a positive integer. Define

$$\|A\|_{(k)} = \sum_{i=1}^{k} \sigma_i(A), \quad A \in \mathbb{M}_n.$$

Schatten p-norm: Let $p \ge 1$ be a real number. Define

$$\|A\|_p = \Big(\sum_{i=1}^{n} \sigma_i^p(A)\Big)^{1/p}, \quad A \in \mathbb{M}_n.$$

It is readily seen that the spectral norm is the Ky Fan norm when $k = 1$, it also equals the limit of $\|A\|_p$ as $p \to \infty$, i.e., $\|A\|_\infty$, whereas the Frobenius norm is the Schatten 2-norm, i.e., $\|A\|_F = \|A\|_2$.

Problems

1. Let $\|\cdot\|$ be a vector norm on $\mathbb{C}^n$ (or $\mathbb{R}^n$). Define
$$\|x\|^D = \max\{|(x, y)| : y \in \mathbb{C}^n,\ \|y\| = 1\}.$$
Show that $\|\cdot\|^D$ is a vector norm on $\mathbb{C}^n$ (known as the *dual norm*).

2. Show that for any matrix norm $\|\cdot\|$ on $\mathbb{M}_n$ and $A = (a_{ij}), B \in \mathbb{M}_n$
$$|\,\|A\| - \|B\|\,| \le \|A - B\| \quad \text{and} \quad \|A\| \le \sum_{i,j} |a_{ij}|\,\|E_{ij}\|,$$
where E_{ij} is the matrix with (i, j)-entry 1 and elsewhere 0 for all i, j.

3. Let $A, B \in \mathbb{M}_n$. Show that
$$\|A + B\|_F \le \|A\|_F + \|B\|_F \quad \text{and} \quad \|AB\|_F \le \|A\|_F\|B\|_F.$$

4. Let $A \in \mathbb{M}_n$. Show that for any matrix norm $\|\cdot\|$ and integer $k \ge 1$,
$$\|A^k\| \le \|A\|^k \quad \text{and} \quad \|A^k\|^{-1}\|I\| \le \|A^{-1}\|^k \ \text{ if } A \text{ is invertible.}$$

5. Let $A \in \mathbb{M}_n$ be given. If there exists a matrix norm $\|\cdot\|$ such that $\|A\| < 1$, show that $A^k \to 0$ as $k \to 0$.

6. Let $\|\cdot\|$ be a matrix norm on $\mathbb{M}_n$. Show that for any invertible matrix $P \in \mathbb{M}_n$, $\|\cdot\|_P : \mathbb{M}_n \mapsto \mathbb{R}$ defined by $\|A\|_P = \|P^{-1}AP\|$ for all matrices $A \in \mathbb{M}_n$ is also a matrix norm.

7. Let $A = (a_{ij}) \in \mathbb{M}_n$ and define $\|A\|_\infty = \max_{1\le i,\, j\le n} |a_{ij}|$. Show that $\|\cdot\|_\infty$ is a matrix-vector norm, but not a multiplicative matrix norm.

8. Show that $\|\cdot\|_\S$, $\|\cdot\|_1$, and $|\!|\!|\cdot|\!|\!|_\infty$ are matrix norms on $\mathbb{M}_n$, where
$$\|A\|_\S = n\|A\|_\infty, \quad \|A\|_1 = \sum_{1\le i,j\le n} |a_{ij}|, \quad |\!|\!|A|\!|\!|_\infty = \max_{1\le i\le n} \sum_{j=1}^n |a_{ij}|.$$
Are these (multiplicative) matrix norms unitarily invariant?

4.3 The Kronecker and Hadamard Products

Matrices can be multiplied in different ways. The Kronecker product and Hadamard product, defined below, used in many fields, are almost as important as the ordinary product. Another basic matrix operation is "compounding" matrices, which is evidently a useful tool in deriving matrix inequalities. This section introduces the three concepts and presents their properties.

The *Kronecker product*, also known as *tensor product* or *direct product*, of two matrices A and B of sizes $m \times n$ and $s \times t$, respectively, is defined to be the $(ms) \times (nt)$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix}.$$

In other words, the Kronecker product $A \otimes B$ is an $(ms) \times (nt)$ matrix, partitioned into mn blocks with the (i,j) block the $s \times t$ matrix $a_{ij}B$. Note that A and B can have any different sizes.

The *Hadamard product*, or the *Schur product*, of two matrices A and B of the same size is defined to be the entrywise product

$$A \circ B = (a_{ij}b_{ij}).$$

In particular, for $u = (u_1, u_2, \dots, u_n)$, $v = (v_1, v_2, \dots, v_n) \in \mathbb{C}^n$,

$$u \otimes v = (u_1v_1, \dots, u_1v_n, \dots, u_nv_1, \dots, u_nv_n)$$

and

$$u \circ v = (u_1v_1, u_2v_2, \dots, u_nv_n).$$

Note that $A \otimes B \neq B \otimes A$ in general and $A \circ B = B \circ A$.

We take, for example, $A = \left(\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Then

$$A \otimes B = \begin{pmatrix} a & b & 2a & 2b \\ c & d & 2c & 2d \\ 3a & 3b & 4a & 4b \\ 3c & 3d & 4c & 4d \end{pmatrix}, \quad A \circ B = \begin{pmatrix} a & 2b \\ 3c & 4d \end{pmatrix}.$$

The Kronecker product has the following basic properties, each of which can be verified by definition and direct computations.

Theorem 4.5 *Let A, B, C be matrices of appropriate sizes. Then*

1. $(kA) \otimes B = A \otimes (kB) = k(A \otimes B)$, *where k is a scalar.*
2. $(A + B) \otimes C = A \otimes C + B \otimes C$.
3. $A \otimes (B + C) = A \otimes B + A \otimes C$.
4. $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.
5. $A \otimes B = 0$ *if and only if* $A = 0$ *or* $B = 0$.
6. $(A \otimes B)^T = A^T \otimes B^T$. *If A and B are symmetric, so is $A \otimes B$.*
7. $(A \otimes B)^* = A^* \otimes B^*$. *If A and B are Hermitian, so is $A \otimes B$.*

Theorem 4.6 *Let A, B, C be matrices of appropriate sizes. Then*

1. $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$.
2. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ *if A and B are invertible.*
3. $A \otimes B$ *is unitary if A and B are unitary.*
4. $A \otimes B$ *is normal if A and B are normal.*

Proof. For (1), let A have n columns. Then C has n rows as indicated in the product AC on the right-hand side of (1). We write $A \otimes B = (a_{ij}B)$, $C \otimes D = (c_{ij}D)$. Then the (i, j) block of $(A \otimes B)(C \otimes D)$ is

$$\sum_{t=1}^{n} a_{it} B c_{tj} D = \sum_{t=1}^{n} a_{it} c_{tj} \, BD.$$

But this is the (i, j)-entry of AC times BD, which is the (i, j) block of $(AC) \otimes (BD)$. (1) follows. The rest are immediate from (1). ∎

To perform the Hadamard product, matrices need to have the same size. In the case of square matrices, an interesting and important observation is that the Hadamard product $A \circ B$ is contained in the Kronecker product $A \otimes B$ as a principal submatrix.

Theorem 4.7 *Let $A, B \in \mathbb{M}_n$. Then the Hadamard product $A \circ B$ is a principal submatrix of the Kronecker product $A \otimes B$ lying on the intersections of rows and columns* $1,\ n+2,\ 2n+3, \ldots, n^2$.

Proof. Let e_i be, as usual, the column vector of n components with the ith position 1 and 0 elsewhere, $i = 1, 2, \ldots, n$, and let

$$E = (e_1 \otimes e_1, \ldots, e_n \otimes e_n).$$

Then for every pair of i and j, we have by computation

$$a_{ij} b_{ij} = (e_i^T A e_j) \otimes (e_i^T B e_j) = (e_i \otimes e_i)^T (A \otimes B)(e_j \otimes e_j),$$

which equals the (i, j)-entry of the matrix $E^T(A \otimes B)E$. Thus,

$$E^T(A \otimes B)E = A \circ B.$$

This says that $A \circ B$ is the principal submatrix of $A \otimes B$ lying on the intersections of rows and columns $1,\ n+2,\ 2n+3, \ldots, n^2$. ∎

The following theorem, relating the eigenvalues of the Kronecker product to those of individual matrices, presents in its proof a common method of decomposing a Kronecker product.

Theorem 4.8 *Let A and B be m-square and n-square complex matrices with eigenvalues λ_i and μ_j, $i = 1, \ldots, m$, $j = 1, \ldots, n$, respectively. Then the eigenvalues of $A \otimes B$ are*

$$\lambda_i \mu_j, \quad i = 1, \ldots, m, \ \ j = 1, \ldots, n,$$

and the eigenvalues of $A \otimes I_n + I_m \otimes B$ are

$$\lambda_i + \mu_j, \quad i = 1, \ldots, m, \ \ j = 1, \ldots, n.$$

Proof. By the Schur decomposition (Theorem 3.3), let U and V be unitary matrices of sizes m and n, respectively, such that

$$U^*AU = T_1 \quad \text{and} \quad V^*BV = T_2,$$

where T_1 and T_2 are upper-triangular matrices with diagonal entries λ_i and μ_j, $i = 1, \ldots, m$, $j = 1, \ldots, n$, respectively. Then

$$T_1 \otimes T_2 = (U^*AU) \otimes (V^*BV) = (U^* \otimes V^*)(A \otimes B)(U \otimes V).$$

Note that $U \otimes V$ is unitary. Thus $A \otimes B$ is unitarily similar to $T_1 \otimes T_2$. The eigenvalues of the latter matrix are $\lambda_i \mu_j$.

For the second part, let $W = U \otimes V$. Then

$$W^*(A \otimes I_n)W = T_1 \otimes I_n = \begin{pmatrix} \lambda_1 I_n & & * \\ & \ddots & \\ 0 & & \lambda_m I_n \end{pmatrix}$$

and

$$W^*(I_m \otimes B)W = I_m \otimes T_2 = \begin{pmatrix} T_2 & & 0 \\ & \ddots & \\ 0 & & T_2 \end{pmatrix}.$$

Thus

$$W^*(A \otimes I_n + I_m \otimes B)W = T_1 \otimes I_n + I_m \otimes T_2$$

is an upper-triangular matrix with eigenvalues $\lambda_i + \mu_j$. ∎

Problems

1. Compute $A \otimes B$ and $B \otimes A$ for

$$A = \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & \sqrt{2} \\ \pi & 2 \\ -1 & 7 \end{pmatrix}.$$

2. Let $J_2 = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. Compute $I_n \otimes J_2$ and $J_2 \otimes I_n$.

3. Let A, B, C, and D be complex matrices. Show that

   (a) $(A \otimes B)^k = A^k \otimes B^k$.

   (b) $\operatorname{tr}(A \otimes B) = \operatorname{tr} A \operatorname{tr} B$.

   (c) $\operatorname{rank}(A \otimes B) = \operatorname{rank}(A) \operatorname{rank}(B)$.

   (d) $\det(A \otimes B) = (\det A)^n (\det B)^m$, if $A \in \mathbb{M}_m$ and $B \in \mathbb{M}_n$.

   (e) If $A \otimes B = C \otimes D \neq 0$, where A and C are of the same size, then $A = aC$ and $B = bD$ with $ab = 1$, and vice versa.

4. Let A and B be m- and n-square matrices, respectively. Show that

$$(A \otimes I_n)(I_m \otimes B) = A \otimes B = (I_m \otimes B)(A \otimes I_n).$$

5. Let $A \in \mathbb{M}_n$ have characteristic polynomial p. Show that
$$\det(A \otimes I + I \otimes A) = (-1)^n \det p(-A).$$

6. Let $A, B \in \mathbb{M}_n$. Show that for some permutation matrix $P \in \mathbb{M}_{n^2}$
$$P^{-1}(A \otimes B)P = B \otimes A.$$

7. Let $x,\ y,\ u,\ v \in \mathbb{C}^n$. With $(x,\ y) = y^*x$ and $\|x\|^2 = x^*x$, show that
$$(x,\ y)(u,\ v) = (x \otimes u,\ y \otimes v).$$
Derive
$$\|x \otimes y\| = \|x\|\,\|y\|.$$

8. Let $A,\ B \in \mathbb{M}_n$. Show that $A \circ I_n = \operatorname{diag}(a_{11}, \dots, a_{nn})$ and that
$$D_1(A \circ B)D_2 = (D_1AD_2) \circ B = A \circ (D_1BD_2)$$
for any n-square diagonal matrices D_1 and D_2.

9. Let A, B, and C be square matrices. Show that
$$(A \oplus B) \otimes C = (A \otimes C) \oplus (B \otimes C).$$
But it need not be true that
$$(A \otimes B) \oplus C = (A \oplus C) \otimes (B \oplus C).$$

10. Consider the vector space $\mathbb{M}_2$, 2×2 complex matrices, over $\mathbb{C}$.
    (a) What is the dimension of $\mathbb{M}_2$?
    (b) Find a basis for $\mathbb{M}_2$.
    (c) For $A,\ B \in \mathbb{M}_2$, define
    $$\mathcal{L}(X) = AXB, \quad X \in \mathbb{M}_2.$$
    Show that $\mathcal{L}$ is a linear transformation on $\mathbb{M}_2$.
    (d) Show that if λ and μ are eigenvalues of A and B, respectively, then $\lambda\mu$ is an eigenvalue of $\mathcal{L}$.

11. Let A and B be square matrices (of possibly different sizes). Show that
$$e^{A \otimes I} = e^A \otimes I, \quad e^{I \otimes B} = I \otimes e^B, \quad e^{A \oplus B} = e^A \otimes e^B.$$

4.4 Compound Matrices

We now turn our attention to *compound matrices*. Roughly speaking, the Kronecker product and the Hadamard product are operations on two (or more) matrices. Unlike the Kronecker and Hadamard products, "compounding" matrix is a matrix operation on a single matrix that arranges in certain order all minors of a given size from the given matrix. A rigorous definition is given as follows.

Let A be an $m \times n$ matrix, $\alpha = \{i_1, \dots, i_s\}$, and $\beta = \{j_1, \dots, j_t\}$, $1 \le i_1 < \cdots < i_s \le m$, $1 \le j_1 < \cdots < j_t \le n$. Denote by $A[i_1, \dots, i_s, j_1, \dots, j_t]$, or simply $A[\alpha, \beta]$, the submatrix of A consisting of the entries in rows $i_1, \dots, i_s$ and columns $j_1, \dots, j_t$.

Given a positive integer $k \le \min\{m, n\}$, there are $\binom{m}{k} \times \binom{n}{k}$ possible minors (numbers) that we can get from the $m \times n$ matrix A. We now form a matrix, denoted by $A^{(k)}$ and called the *kth compound matrix* of A, of size $\binom{m}{k} \times \binom{n}{k}$ by ordering these numbers lexicographically; that is, the $(1,1)$ position of $A^{(k)}$ is $\det A[1, \dots, k|1, \dots, k]$, the $(1,2)$ position of $A^{(k)}$ is $\det A[1, \dots, k|1, \dots, k-1, k+1], \dots$, whereas the $(2,1)$ position is $\det A[1, \dots, k-1, k+1|1, \dots, k]$, …, and so on. For convenience, we say that the minor $\det A[\alpha, \beta]$ is in the (α, β) position of the compound matrix and denote it by $A^{(k)}_{\alpha,\beta}$. Clearly, $A^{(1)} = A$ and $A^{(n)} = \det A$ if A is an $n \times n$ matrix.

As an example, let $m = n = 3$, $k = 2$, and take

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Then

$$\begin{aligned} A^{(2)} &= \begin{pmatrix} \det A[1,2|1,2] & \det A[1,2|1,3] & \det A[1,2|2,3] \\ \det A[1,3|1,2] & \det A[1,3|1,3] & \det A[1,3|2,3] \\ \det A[2,3|1,2] & \det A[2,3|1,3] & \det A[2,3|2,3] \end{pmatrix} \\ &= \begin{pmatrix} \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix} & \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} \\ \begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 7 & 9 \end{vmatrix} & \begin{vmatrix} 2 & 3 \\ 8 & 9 \end{vmatrix} \\ \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} & \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} & \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} \end{pmatrix} = \begin{pmatrix} -3 & -6 & -3 \\ -6 & -12 & -6 \\ -3 & -6 & -3 \end{pmatrix}. \end{aligned}$$

If A is an n-square matrix, then the main diagonal entries of $A^{(k)}$ are $\det A[\alpha|\alpha]$, i.e., the principal minors of A. For an n-square upper-triangular matrix A, $\det A[\alpha|\beta] = 0$ if α is after β in lexicographic order. This leads to the result that if A is upper (lower)-triangular, then so is $A^{(k)}$. As a consequence, if A is diagonal, then so is $A^{(k)}$.

The goal of this section is to show that the compound matrix of the product of matrices is the product of their compound matrices. For this purpose, we need to borrow a well-known result on determinant expansion, the *Binet–Cauchy formula.* (A good reference on this formula and its proof is Lancaster and Tismenetsky's book, *The Theory of Matrices*, 1985, pp. 36–42.)

Theorem 4.9 (Binet–Cauchy formula) *Let $C = AB$, where A is $m \times n$ and B is $n \times m$, $m \leq n$, and let $\alpha = \{1, 2, \dots, m\}$. Then*

$$\det C = \sum_{\beta} \det A[\alpha|\beta] \det B[\beta|\alpha],$$

where β runs over all sequences $\{j_1, \dots, j_m\}$, $1 \leq j_1 < \cdots < j_m \leq n$.

The following theorem is of fundamental importance for compound matrices, whose corollary plays a pivotal role in deriving matrix inequalities involving eigenvalue and singular value products.

Theorem 4.10 *Let A be an $m \times p$ matrix and B be a $p \times n$ matrix. If k is a positive integer, $k \leq \min\{m, p, n\}$, then $(AB)^{(k)} = A^{(k)} B^{(k)}$.*

Proof. For $\alpha = \{i_1, \dots, i_k\}$ and $\beta = \{j_1, \dots, j_k\}$, $1 \leq i_1 < \cdots < i_k \leq m$, $1 \leq j_1 < \cdots < j_k \leq m$, we compute the entry in the (α, β) position (in lexicographic order) of $(AB)^{(k)}$ by the Binet–Cauchy determinant expansion formula and get

$$\begin{aligned} (AB)^{(k)}_{\alpha,\beta} &= \det\left((AB)[\alpha|\beta]\right) \\ &= \sum_{\gamma} \det A[\alpha|\gamma] \det B[\gamma|\beta] = (A^{(k)} B^{(k)})_{\alpha,\beta}, \end{aligned}$$

where γ runs over all possible sequences $1 \leq \gamma_1 < \cdots < \gamma_k \leq p$. ∎

If A is Hermitian, it is readily seen that $A^{(k)}$ is Hermitian too.

Corollary 4.1 *Let $A \in \mathbb{M}_n$ be a positive semidefinite matrix with eigenvalues $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$. Then the largest eigenvalue of $A^{(k)}$ is the product of the first k largest eigenvalues of A; that is,*

$$\lambda_{\max}(A^{(k)}) = \prod_{i=1}^{k} \lambda_i(A).$$

Problems

1. Find $A^{(2)}$, where
$$A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 4 & 5 & 6 & 7 \\ 8 & 9 & 10 & -1 \end{pmatrix}.$$
2. Show that $A^*[\alpha|\beta] = (A[\beta|\alpha])^*$.
3. Show that $I_n^{(k)} = I_{\binom{n}{k}}$, where I_l is the $l \times l$ identity matrix.
4. Show that $(A^{(k)})^* = (A^*)^{(k)}$; $(A^{(k)})^T = (A^T)^{(k)}$.
5. Show that $(A^{(k)})^{-1} = (A^{-1})^{(k)}$ if A is nonsingular.
6. Show that $\det A^{(k)} = (\det A)^{\binom{n-1}{k-1}}$ when A is n-square.
7. If $\operatorname{rank}(A) = r$, show that $\operatorname{rank}(A^{(k)}) = \binom{r}{k}$ or 0 if $r < k$.
8. Show that if A is unitary, symmetric, positive (semi-)definite, Hermitian, or normal, then so is $A^{(k)}$, respectively.
9. If $A = \operatorname{diag}(a_1, \ldots, a_n)$, show that $A^{(k)}$ is an $\binom{n}{k} \times \binom{n}{k}$ diagonal matrix with diagonal entries $a_{i_1} \cdots a_{i_k}$, $1 \leq i_1 < \cdots < i_k \leq n$.
10. If $A \in \mathbb{M}_n$ has eigenvalues $\lambda_1, \ldots, \lambda_n$, show that $A^{(k)}$ has eigenvalues $\lambda_{i_1} \cdots \lambda_{i_k}$, $1 \leq i_1 < \cdots < i_k \leq n$.
11. If $A \in \mathbb{M}_n$ has singular values $\sigma_1, \ldots, \sigma_n$, show that $A^{(k)}$ has singular values $\sigma_{i_1} \cdots \sigma_{i_k}$, $1 \leq i_1 < \cdots < i_k \leq n$.
12. If $A \in \mathbb{M}_n$ has eigenvalues $\lambda_1, \ldots, \lambda_n$, show that $\operatorname{tr}(A^{(k)})$ equals $\sum_\gamma \lambda_{i_1} \cdots \lambda_{i_k}$, denoted by $s_k(\lambda_1, \ldots, \lambda_n)$ and called kth *elementary symmetric function*, where γ is any sequence $1 \leq i_1 < \cdots < i_k \leq n$.
13. If A is a positive semidefinite matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$, show that the smallest eigenvalue of $A^{(k)}$ is $\lambda_{\min}(A^{(k)}) = \prod_{i=1}^{k} \lambda_{n-i+1}$.

CHAPTER 5

Special Types of Matrices

Introduction: This chapter studies special types of matrices. They are: idempotent matrices, nilpotent matrices, involutary matrices, projection matrices, tridiagonal matrices, circulant matrices, Vandermonde matrices, Hadamard matrices, permutation matrices, doubly stochastic matrices, and nonnegative matrices. These matrices are often used in many subjects of mathematics and in other fields.

5.1 Idempotence, Nilpotence, Involution, and Projections

We first present three types of matrices that have simple structures under similarity: idempotent matrices, nilpotent matrices, and involutions. We then turn attention to orthogonal projection matrices.

A square matrix A is said to be *idempotent*, or a *projection*, if

$$A^2 = A,$$

nilpotent if for some positive integer k

$$A^k = 0,$$

and *involutary* if

$$A^2 = I.$$

F. Zhang, *Matrix Theory: Basic Results and Techniques*, Universitext,
DOI 10.1007/978-1-4614-1099-7_5, © Springer Science+Business Media, LLC 2011

Theorem 5.1 *Let A be an n-square complex matrix. Then*

1. *A is idempotent if and only if A is similar to a diagonal matrix of the form* $\operatorname{diag}(1, \dots, 1, 0, \dots, 0)$.
2. *A is nilpotent if and only if all the eigenvalues of A are zero.*
3. *A is involutary if and only if A is similar to a diagonal matrix of the form* $\operatorname{diag}(1, \dots, 1, -1, \dots, -1)$.

Proof. The sufficiency in (1) is obvious. To see the necessity, let

$$A = P^{-1}(J_1 \oplus \cdots \oplus J_k)P$$

be a Jordan decomposition of A. Then for each i, $i = 1, \dots, k$,

$$A^2 = A \quad \Rightarrow \quad J_i^2 = J_i.$$

Observe that if J is a Jordan block and if $J^2 = J$, then J must be of size 1; that is, J is a number. The assertion then follows.

For (2), consider the Schur (or Jordan) decomposition of A,

$$A = U^{-1} \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} U,$$

where U is an n-square unitary matrix.

If $A^k = 0$, then each $\lambda_i^k = 0$, and A has only zero eigenvalues. Conversely, it is easy to verify by computation that $A^n = 0$ if all the eigenvalues of A are equal to zero (see also Problem 11, Section 3.3).

The proof of (3) is similar to that of (1). ∎

Theorem 5.2 *Let A and B be nilpotent matrices of the same size. If A and B commute, then $A + B$ is nilpotent.*

Proof. Let $A^m = 0$ and $B^n = 0$. Upon computation, we have

$$(A + B)^{m+n} = 0,$$

for each term in the expansion of $(A + B)^{m+n}$ is A^{m+n}, is B^{m+n}, or contains $A^s B^t$, $s \geq m$ or $t \geq n$. In any case, every term vanishes. ∎

By choosing a suitable basis for $\mathbb{C}^n$, we can interpret Theorem 5.1(1) as follows. A matrix A is a projection if and only if $\mathbb{C}^n$ can be decomposed as

$$\mathbb{C}^n = W_1 \oplus W_2, \tag{5.1}$$

where W_1 and W_2 are subspaces such that for all $w_1 \in W_1$, $w_2 \in W_2$,

$$Aw_1 = w_1, \quad Aw_2 = 0.$$

Thus, if $w = w_1 + w_2 \in \mathbb{C}^n$, where $w_1 \in W_1$ and $w_2 \in W_2$, then

$$Aw = Aw_1 + Aw_2 = w_1.$$

Such a w_1 is called the *projection* of w on W_1. Note that

$$W_1 = \operatorname{Im} A, \quad W_2 = \operatorname{Ker} A = \operatorname{Im}(I - A).$$

Using this and Theorem 5.1(1), one may prove the next result.

Theorem 5.3 *For any $A \in \mathbb{M}_n$ the following are equivalent.*

1. *A is a projection matrix; that is, $A^2 = A$.*
2. *$\mathbb{C}^n = \operatorname{Im} A + \operatorname{Ker} A$ with $Ax = x$ for every $x \in \operatorname{Im} A$.*
3. *$\operatorname{Ker} A = \operatorname{Im}(I - A)$.*
4. *$\operatorname{rank}(A) + \operatorname{rank}(I - A) = n$.*
5. *$\operatorname{Im} A \cap \operatorname{Im}(I - A) = \{0\}$.*

We now turn our attention to orthogonal projection matrices. A square complex matrix A is called an *orthogonal projection* if

$$A^2 = A = A^*.$$

For orthogonal projection matrices, the subspaces

$$W_1 = \operatorname{Im} A \quad \text{and} \quad W_2 = \operatorname{Im}(I - A)$$

in (5.1) are orthogonal; that is, for all $w_1 \in W_1$ and $w_2 \in W_2$,

$$(w_1, w_2) = 0. \tag{5.2}$$

In other words, $(Ax, (I - A)x) = 0$ for all $x \in \mathbb{C}^n$; this is because

$$(w_1, w_2) = (Aw_1, w_2) = (w_1, A^*w_2) = (w_1, Aw_2) = 0.$$

Theorem 5.4 *For any $A \in \mathbb{M}_n$ the following are equivalent.*

1. *A is an orthogonal projection matrix; that is, $A^2 = A = A^*$.*
2. *$A = U^* \operatorname{diag}(1, \ldots, 1, 0, \ldots, 0)U$ for some unitary matrix U.*
3. *$\|x - Ax\| \le \|x - Ay\|$ for every x and y in $\mathbb{C}^n$.*
4. *$A^2 = A$ and $\|Ax\| \le \|x\|$ for every $x \in \mathbb{C}^n$.*
5. *$A = A^*A$.*

Proof. (1)$\Leftrightarrow$(2): We show (1)$\Rightarrow$(2). The other direction is obvious.

Because A is Hermitian, by the spectral decomposition theorem (Theorem 3.4), we have $A = V^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n)V$ for some unitary matrix V, where the λ_i are the eigenvalues of A. However, A is idempotent and thus has only eigenvalues 1 and 0 according to the previous theorem. It follows that

$$A = U^* \operatorname{diag}(\overbrace{1, \ldots, 1}^{r}, 0, \ldots, 0)U,$$

where r is the rank of A and U is some unitary matrix.

(1)$\Leftrightarrow$(3): For (1)$\Rightarrow$(3), let A be an orthogonal projection. We have the decomposition (5.1) with the orthogonality condition (5.2).

Let $x = x_1 + x_2$, where $x_1 \in W_1$, $x_2 \in W_2$, and $(x_1, x_2) = 0$. Similarly, write $y = y_1 + y_2$. Note that $x_1 - y_1 \in W_1$ and $W_1 \perp W_2$.

Since $(u, v) = 0$ implies $\|u\|^2 + \|v\|^2 = \|u + v\|^2$, we have

$$\|x-Ax\|^2 = \|x_2\|^2 \le \|x_2\|^2 + \|x_1-y_1\|^2 = \|x_2+(x_1-y_1)\|^2 = \|x-Ay\|^2.$$

We now show (3)$\Rightarrow$(1). It is sufficient to show that the decomposition (5.1) with the orthogonality condition (5.2) holds, where $\operatorname{Im} A$ serves as W_1 and $\operatorname{Im}(I - A)$ as W_2.

As $x = Ax + (I - A)x$ for every $x \in \mathbb{C}^n$, it is obvious that

$$\mathbb{C}^n = \operatorname{Im} A + \operatorname{Im}(I - A).$$

We have left to show that $(x, y) = 0$ if $x \in \operatorname{Im} A$ and $y \in \operatorname{Im}(I - A)$. Suppose instead that $((I - A)x, Ay) \ne 0$ for some x and $y \in \mathbb{C}^n$. We show that there exists a vector $z \in \mathbb{C}^n$ such that

$$\|x - Az\| < \|x - Ax\|,$$

which is a contradiction to the given condition (3).

Let $((I-A)x, Ay) = \alpha \neq 0$. We may assume that $\alpha < 0$. Otherwise, replace x with $e^{i\theta}x$, where $\theta \in \mathbb{R}$ is such that $e^{i\theta}\alpha < 0$.

Let $z_\epsilon = x - \epsilon\, y$, where $\epsilon > 0$. Then

$$\begin{aligned} \|x - Az_\epsilon\|^2 &= \|(x - Ax) + (Ax - Az_\epsilon)\|^2 \\ &= \|x - Ax\|^2 + \|Ax - Az_\epsilon\|^2 \\ &\quad + 2\,\mathrm{Re}((I-A)x, A(x - z_\epsilon)) \\ &= \|x - Ax\|^2 + \|Ax - Az_\epsilon\|^2 \\ &\quad + 2\epsilon((I-A)x, Ay) \\ &= \|x - Ax\|^2 + \epsilon^2\|Ay\|^2 + 2\epsilon\alpha. \end{aligned}$$

Because $\alpha < 0$, we have $\epsilon^2\|Ay\|^2 + 2\epsilon\alpha < 0$ for some ϵ small enough, which results in a contradiction to the assumption in (3):

$$\|x - Az_\epsilon\| < \|x - Ax\|.$$

(1)⇒(4): If A is an orthogonal projection matrix, then the orthogonality condition (5.2) holds. Thus, $(Ax, (I-A)x) = 0$ and

$$\|Ax\|^2 \leq \|Ax\|^2 + \|(I-A)x\|^2 = \|Ax + (I-A)x\|^2 = \|x\|^2.$$

(4)⇒(5): If $A \neq A^*A$; that is, $(A^* - I)A \neq 0$ or $A^*(I-A) \neq 0$, then $\operatorname{rank}(I-A) < n$ and $\dim \operatorname{Im}(I-A) < n$ by Theorem 5.1(1).

We show that there exists a nonzero x such that

$$(x, (I-A)x) = 0, \quad \text{but} \quad (I-A)x \neq 0.$$

Thus, for this x,

$$\|Ax\|^2 = \|x - (I-A)x\|^2 = \|x\|^2 + \|(I-A)x\|^2 > \|x\|^2,$$

which contradicts the condition $\|Ax\| \leq \|x\|$ for every $x \in \mathbb{C}^n$.

To show the existence of such a vector x, it is sufficient to show that there exists a nonzero x in $(\operatorname{Im}(I-A))^\perp$ but not in $\operatorname{Ker}(I-A)$; that is, $(\operatorname{Im}(I-A))^\perp$ is not contained in $\operatorname{Ker}(I-A)$.

Notice that (Theorem 1.5)

$$\dim \operatorname{Im}(I-A) + \dim \operatorname{Ker}(I-A) = n$$

and that

$$\mathbb{C}^n = \operatorname{Im}(I - A) \oplus (\operatorname{Im}(I - A))^{\perp}.$$

Now if $(\operatorname{Im}(I - A))^{\perp}$ is contained in $\operatorname{Ker}(I - A)$, then they must be equal, for they have the same dimension:

$$\dim(\operatorname{Im}(I - A))^{\perp} = n - \dim \operatorname{Im}(I - A) = \dim \operatorname{Ker}(I - A).$$

It follows, by (1.11) in Section 1.4 of Chapter 1, that

$$\operatorname{Im}(I - A) = \operatorname{Im}(I - A^*).$$

Thus, $I - A = I - A^*$ and A is Hermitian. Then (5) follows easily.

(5)⇒(1): If $A = A^*A$, then A is obviously Hermitian. Thus,

$$A = A^*A = AA = A^2. \blacksquare$$

Problems

1. Characterize all 2×2 idempotent, nilpotent, and involutary matrices up to similarity.

2. Can a nonzero matrix be both idempotent and nilpotent? Why?

3. What is the characteristic polynomial of a nilpotent matrix?

4. What idempotent matrices are nonsingular?

5. Show that if $A \in \mathbb{M}_n$ is idempotent, then so is $P^{-1}AP$ for any invertible $P \in \mathbb{M}_n$.

6. Show that the rank of an idempotent matrix is equal to the number of nonzero eigenvalues of the matrix.

7. Let A and B be idempotent matrices of the same size. Find the necessary and sufficient conditions for $A + B$ to be idempotent. Discuss the analogue for $A - B$.

8. Show that $\frac{1}{2}(I + A)$ is idempotent if and only if A is an involution.

9. Let A be a square complex matrix. Show that

$$A^2 = A \quad \Leftrightarrow \quad \operatorname{rank}(A) = \operatorname{tr}(A) \text{ and } \operatorname{rank}(I - A) = \operatorname{tr}(I - A).$$

   If $A^2 = -A$, what rank conditions on A does one get?

10. Let A be an idempotent matrix. Show that

$$A = A^* \quad \Leftrightarrow \quad \operatorname{Im} A = \operatorname{Im} A^*.$$

11. Show that a Hermitian idempotent matrix is positive semidefinite and that the matrix M is positive semidefinite, where

$$M = I - \frac{1}{y^*y} yy^*, \quad y \in \mathbb{C}^n.$$

12. Show that $\mathcal{T}^3 = 0$, where $\mathcal{T}$ is a transformation defined on $\mathbb{M}_2$ by

$$\mathcal{T}(X) = TX - XT, \quad X \in \mathbb{M}_2,$$

with

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

Find the Jordan form of a matrix representation of $\mathcal{T}$.

13. Let $A \in \mathbb{M}_n$. Define a linear transformation on $\mathbb{M}_n$ by

$$\mathcal{T}(X) = AX - XA, \quad X \in \mathbb{M}_n.$$

Show that if A is nilpotent then $\mathcal{T}$ is nilpotent and that if A is diagonalizable then the matrix representation of $\mathcal{T}$ is diagonalizable.

14. Let A and B be square matrices of the same size. If A is nilpotent and $AB = BA$, show that AB is nilpotent. Is the converse true?

15. Let A be an n-square nonsingular matrix. If X is a matrix such that

$$AXA^{-1} = \lambda X, \quad \lambda \in \mathbb{C},$$

show that $|\lambda| = 1$ or X is nilpotent.

16. Give a 2×2 matrix such that $A^2 = I$ but $A^*A \neq I$.

17. Let $A^2 = A$. Show that $(A + I)^k = I + (2^k - 1)A$ for $k = 1, 2, \ldots$.

18. Find a matrix that is a projection but not an orthogonal projection.

19. Let A and B be square matrices of the same size. If $AB = A$ and $BA = B$, show that A and B are projection matrices.

20. Let A be a projection matrix. Show that A is Hermitian if and only if $\operatorname{Im} A$ and $\operatorname{Ker} A$ are orthogonal; that is, $\operatorname{Im} A \perp \operatorname{Ker} A$.

21. Prove Theorem 5.3 along the line: (1)$\Leftrightarrow$(2)$\Rightarrow$(3)$\Rightarrow$(4)$\Rightarrow$(5)$\Rightarrow$(2).

22. Show that A is an orthogonal projection matrix if and only if $A = B^*B$ for some matrix B with $BB^* = I$.

23. Let $A_1, \ldots, A_m$ be $n \times n$ idempotent matrices. If

$$A_1 + \cdots + A_m = I_n,$$

show that

$$A_i A_j = 0, \quad i \neq j.$$

[Hint: Show that $\mathbb{C}^n = \operatorname{Im} A_1 \oplus \cdots \oplus \operatorname{Im} A_m$ by using trace.]

24. If W is a subspace of an inner product space V, one may write

$$V = W \oplus W^\perp$$

and define a transformation $\mathcal{A}$ on V by

$$\mathcal{A}(v) = w, \quad \text{if } v = w + w^\perp,\ w \in W,\ w^\perp \in W^\perp,$$

where w is called the *projection* of v on W. Show that

(a) $\mathcal{A}$ is a linear transformation.

(b) $\mathcal{A}^2 = \mathcal{A}$.

(c) $\operatorname{Im}(\mathcal{A}) = W$ and $\operatorname{Ker}(\mathcal{A}) = W^\perp$.

(d) $\|v - \mathcal{A}(v)\| \leq \|v - \mathcal{A}(u)\|$ for any $u \in V$.

(e) Every $v \in V$ has a unique projection $w \in W$.

(f) $\|v\|^2 = \|w\|^2 + \|w^\perp\|^2$.

25. When does equality in Theorem 5.3(3) hold?

26. Let A and B be $m \times n$ complex matrices of rank n, $n \leq m$. Show that the matrix $A(A^*A)^{-1}A^*$ is idempotent and that

$$A(A^*A)^{-1}A^* = B(B^*B)^{-1}B^* \quad \Leftrightarrow \quad A = BX$$

for some nonsingular matrix X. [Hint: Multiply by A.]

27. Let A and B be orthogonal projections of the same size. Show that

(a) $A + B$ is an orthogonal projection if and only if $AB = BA = 0$.

(b) $A - B$ is an orthogonal projection if and only if $AB = BA = B$.

(c) AB is an orthogonal projection if and only if $AB = BA$.

5.2 Tridiagonal Matrices

One of the frequently used techniques in determinant computation is recursion. We illustrate this method by computing the determinant of a tridiagonal matrix and go on studying the eigenvalues of matrices of this kind.

An n-square *tridiagonal matrix* is a matrix with entries $t_{ij} = 0$ whenever $|i-j| > 1$. The determinant of a tridiagonal matrix can be calculated inductively. For simplicity, we consider the special tridiagonal matrix

$$T_n = \begin{pmatrix} a & b & & & & 0 \\ c & a & b & & & \\ & c & a & b & & \\ & & \ddots & \ddots & \ddots & \\ & & & c & a & b \\ 0 & & & & c & a \end{pmatrix}. \tag{5.3}$$

Theorem 5.5 *Let T_n be defined as in (5.3). Then*

$$\det T_n = \begin{cases} a^n & \text{if } bc = 0, \\ (n+1)(a/2)^n & \text{if } a^2 = 4bc, \\ (\alpha^{n+1} - \beta^{n+1})/(\alpha - \beta) & \text{if } a^2 \neq 4bc, \end{cases}$$

where

$$\alpha = \frac{a + \sqrt{a^2 - 4bc}}{2}, \quad \beta = \frac{a - \sqrt{a^2 - 4bc}}{2}.$$

Proof. Expand the determinant along the first row of the matrix in (5.3) to obtain the recursive formula

$$\det T_n = a \det T_{n-1} - bc \det T_{n-2}. \tag{5.4}$$

If $bc = 0$, then $b = 0$ or $c = 0$, and from (5.3) obviously $\det T_n = a^n$. If $bc \neq 0$, let α and β be the solutions to $x^2 - ax + bc = 0$. Then

$$\alpha + \beta = a, \quad \alpha\beta = bc.$$

Note that

$$a^2 - 4bc = (\alpha - \beta)^2.$$

From the recursive formula (5.4), we have

$$\det T_n - \alpha \det T_{n-1} = \beta(\det T_{n-1} - \alpha \det T_{n-2})$$

and

$$\det T_n - \beta \det T_{n-1} = \alpha(\det T_{n-1} - \beta \det T_{n-2}).$$

Denote

$$f_n = \det T_n - \alpha \det T_{n-1}, \quad g_n = \det T_n - \beta \det T_{n-1}.$$

Then

$$f_n = \beta f_{n-1}, \quad g_n = \alpha g_{n-1},$$

with (by a simple computation)

$$f_2 = \beta^2, \quad g_2 = \alpha^2.$$

Thus,

$$f_n = \beta^n, \quad g_n = \alpha^n;$$

that is,

$$\det T_n - \alpha \det T_{n-1} = \beta^n, \quad \det T_n - \beta \det T_{n-1} = \alpha^n. \tag{5.5}$$

It follows, using T_{n+1} in (5.5) and subtracting the equations, that

$$\det T_n = \frac{\alpha^{n+1} - \beta^{n+1}}{\alpha - \beta}, \quad \text{if } \alpha \neq \beta.$$

If $\alpha = \beta$, one can easily prove by induction that

$$\det T_n = (n+1)\left(\frac{a}{2}\right)^n. \ \blacksquare$$

Note that the recursive formula (5.4) in the proof depends not on the single values of b and c but on the product bc. Thus, if $a \in \mathbb{R}$, $bc > 0$, we may replace b and c by d and $\bar{d}$, respectively, where $d\bar{d} = bc$, to get a tridiagonal Hermitian matrix H_n, for which

$$\det(\lambda I - T_n) = \det(\lambda I - H_n).$$

It follows that T_n has only real eigenvalues because H_n does. In fact, when a, b, $c \in \mathbb{R}$, and $bc > 0$, matrix DT_nD^{-1} is real symmetric, where D is the diagonal matrix $\operatorname{diag}(1, e, \ldots, e^{n-1})$ with $e = \sqrt{b/c}$.

Theorem 5.6 *If T_n is a tridiagonal matrix defined as in (5.3) with $a, b, c \in \mathbb{R}$ and $bc > 0$, then the eigenvalues of T_n are all real and have eigenspaces of dimension one.*

Proof. The first half follows from the argument prior to the theorem. For the second part, it is sufficient to prove that each eigenvalue has only one eigenvector up to a factor.

Let $x = (x_1, \ldots, x_n)^T$ be an eigenvector of T_n corresponding to the eigenvalue λ. Then

$$(\lambda I - T_n)x = 0, \quad x \neq 0,$$

or equivalently

$$\begin{aligned}
(\lambda - a)x_1 - bx_2 &= 0, \\
-cx_1 + (\lambda - a)x_2 - bx_3 &= 0, \\
&\vdots \\
-cx_{n-2} + (\lambda - a)x_{n-1} - bx_n &= 0, \\
-cx_{n-1} + (\lambda - a)x_n &= 0.
\end{aligned}$$

Because $b \neq 0$, x_2 is determined by x_1 in the first equation, so are $x_3, \ldots, x_n$ successively by x_2, x_3, and so on in the equations 2, 3, $\ldots$, $n-1$. If x_1 is replaced by kx_1, then $x_2, x_3, \ldots, x_n$ become $kx_2, kx_3, \ldots, kx_n$, and the eigenvector is unique up to a factor. ∎

Note that the theorem is in fact true for a general tridiagonal matrix when a_i is real and $b_i c_i > 0$ for each i.

Problems

1. Compute the determinant

$$\begin{vmatrix} a & b & 0 \\ c & a & b \\ 0 & c & a \end{vmatrix}.$$

2. Carry out in detail the proof that T_n is similar to a real symmetric matrix if a, b, $c \in \mathbb{R}$ and $bc > 0$.

3. Compute the $n \times n$ determinant

$$\begin{vmatrix} 0 & 1 & & & & 0 \\ 1 & 0 & 1 & & & \\ & 1 & 0 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & 0 & 1 \\ 0 & & & & 1 & 0 \end{vmatrix}.$$

4. Compute the $n \times n$ determinant

$$\begin{vmatrix} 1 & 1 & & & & 0 \\ -1 & 1 & 1 & & & \\ & -1 & 1 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 1 & 1 \\ 0 & & & & -1 & 1 \end{vmatrix}.$$

5. Compute the $n \times n$ determinant

$$\begin{vmatrix} 2 & 1 & & & & 0 \\ 1 & 2 & 1 & & & \\ & 1 & 2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & 2 & 1 \\ 0 & & & & 1 & 2 \end{vmatrix}.$$

6. Compute the $n \times n$ determinant

$$\begin{vmatrix} 3 & 2 & & & & & 0 \\ 1 & 3 & 1 & & & & \\ & 2 & 3 & 2 & & & \\ & & 1 & 3 & 1 & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & 1 & 3 & 1 \\ 0 & & & & & 2 & 3 \end{vmatrix}.$$

7. Find the inverse of the $n \times n$ matrix

$$\begin{pmatrix} 2 & -1 & & & & & 0 \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & -1 & 2 & -1 & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & -1 & 2 & -1 \\ 0 & & & & & -1 & 2 \end{pmatrix}.$$

8. Show that the value of the following determinant is independent of x:

$$\begin{vmatrix} a & x & & & & 0 \\ \frac{1}{x} & a & x & & & \\ & \frac{1}{x} & a & x & & \\ & & \ddots & \ddots & \ddots & \\ & & & \frac{1}{x} & a & x \\ 0 & & & & \frac{1}{x} & a \end{vmatrix}.$$

9. (**Cauchy matrix**) Let $\lambda_1, \ldots, \lambda_n$ be positive numbers and let

$$\Lambda = \left(\frac{1}{\lambda_i + \lambda_j} \right).$$

Show that

$$\det \Lambda = \frac{\prod_{i>j}(\lambda_i - \lambda_j)^2}{\prod_{i,j}(\lambda_i + \lambda_j)}.$$

[Hint: Subtract the last column from each of the other columns, then factor; do the same thing for rows; use induction.]

10. Show that

$$\begin{vmatrix} \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{n+2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{n+1} & \frac{1}{n+2} & \cdots & \frac{1}{2n} \end{vmatrix} \geq 0.$$

5.3 Circulant Matrices

An n-square *circulant matrix* is a matrix of the form

$$\begin{pmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \cdots & c_{n-2} \\ c_{n-2} & c_{n-1} & c_0 & \cdots & c_{n-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_1 & c_2 & c_3 & \cdots & c_0 \end{pmatrix}, \tag{5.6}$$

where $c_0, c_1, \dots, c_{n-1}$ are complex numbers. For instance,

$$N = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ n & 1 & 2 & \cdots & n-1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 3 & 4 & 5 & \cdots & 2 \\ 2 & 3 & 4 & \cdots & 1 \end{pmatrix}$$

and

$$P = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix} \tag{5.7}$$

are circulant matrices. Note that P is also a permutation matrix. We refer to this P as the $n \times n$ *primary permutation matrix*.

This section deals with the basic properties of circulant matrices. The following theorem may be shown by a direct verification.

Theorem 5.7 *An n-square matrix C is circulant if and only if*

$$C = PCP^T,$$

where P is the $n \times n$ primary permutation matrix.

We call a complex number ω an nth *primitive root* of unity if $\omega^n - 1 = 0$ and $\omega^k - 1 \neq 0$ for every positive integer $k < n$.

Note that if ω is an nth primitive root of unity, then ω^k is a solution to $x^n - 1 = 0$, $0 < k < n$. It follows by factoring $x^n - 1$ that

$$\sum_{i=0}^{n-1} \omega^{ik} = \frac{(\omega^k)^n - 1}{\omega^k - 1} = 0.$$

Theorem 5.8 *Let C be a circulant matrix in the form (5.6), and let $f(\lambda) = c_0 + c_1\lambda + \cdots + c_{n-1}\lambda^{n-1}$. Then*

1. *$C = f(P)$, where P is the $n \times n$ primary permutation matrix.*
2. *C is a normal matrix; that is, $C^*C = CC^*$.*
3. *The eigenvalues of C are $f(\omega^k)$, $k = 0, 1, \ldots, n-1$.*
4. *$\det C = f(\omega^0)f(\omega^1)\cdots f(\omega^{n-1})$.*
5. *F^*CF is a diagonal matrix, where F is the unitary matrix with the (i, j)-entry equal to $\frac{1}{\sqrt{n}}\omega^{(i-1)(j-1)}$, i, $j = 1, \ldots, n$.*

Proof. (1) is easy to see by a direct computation. (2) is due to the fact that if matrices A and B commute, so do $p(A)$ and $q(B)$, where p and q are any polynomials (Problem 4). Note that $PP^* = P^*P$.

For (3) and (4), the characteristic polynomial of P is

$$\det(\lambda I - P) = \lambda^n - 1 = \prod_{k=0}^{n-1}(\lambda - \omega^k).$$

Thus, the eigenvalues of P and P^i are, respectively, ω^k and ω^{ik}, $k = 0, 1, \ldots, n-1$. It follows that the eigenvalues of $C = f(P)$ are $f(\omega^k)$, $k = 0, 1, \ldots, n-1$ (Problem 7, Section 3.2), and that

$$\det C = \prod_{k=0}^{n-1} f(\omega^k).$$

To show (5), for each $k = 0, 1, \ldots, n-1$, let

$$x_k = (1,\ \omega^k,\ \omega^{2k},\ \ldots,\ \omega^{(n-1)k})^T.$$

Then

$$Px_k = (\omega^k,\ \omega^{2k},\ \dots,\ \omega^{(n-1)k},\ 1)^T = \omega^k x_k$$

and

$$Cx_k = f(P)x_k = f(\omega^k)x_k.$$

In other words, x_k are the eigenvectors of P and C corresponding to the eigenvalues ω^k and $f(\omega^k)$, respectively, $k = 0, 1, \dots, n-1$.

However, because

$$(x_i, x_j) = \sum_{k=0}^{n-1} \overline{\omega^{jk}}\,\omega^{ik} = \sum_{k=0}^{n-1} \omega^{(i-j)k} = \begin{cases} 0, & i \neq j, \\ n, & i = j, \end{cases}$$

we have that

$$\left\{ \frac{1}{\sqrt{n}} x_0,\ \frac{1}{\sqrt{n}} x_1,\ \dots,\ \frac{1}{\sqrt{n}} x_{n-1} \right\}$$

is an orthonormal basis for $\mathbb{C}^n$. Thus, we get a unitary matrix

$$F = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \dots & \omega^{(n-1)(n-1)} \end{pmatrix}$$

such that

$$F^* C F = \operatorname{diag}(f(\omega^0), f(\omega^1), \dots, f(\omega^{n-1})).$$

That F is a unitary matrix is verified by a direct computation. ∎

Note that F, called a *Fourier matrix*, is independent of C.

Problems

1. Let ω be an nth primitive root of unity. Show that

   (a) $\omega\overline{\omega} = 1$.

   (b) $\overline{\omega}^k = \omega^{-k} = \omega^{n-k}$.

   (c) $1 + \omega + \dots + \omega^{n-1} = 0$.

2. Let ω be an nth primitive root of unity. Show that ω^k is also an nth primitive root of unity if and only if $(n,k)=1$; that is, n and k have no common positive divisors other than 1.

3. Show that if A is a circulant matrix, then so are A^*, A^k, and A^{-1} if the inverse exists.

4. Let A and B be square matrices of the same size. If $AB = BA$, show that $p(A)q(B) = q(B)p(A)$ for any polynomials p and q.

5. Let A and B be circulant matrices of the same size. Show that A and B commute and that AB is a circulant matrix.

6. Let A be a circulant matrix. Show that for every positive integer k
$$\operatorname{rank}(A^k) = \operatorname{rank}(A).$$

7. Find the eigenvalues of the circulant matrices:
$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & \omega & \omega^2 \\ \omega^2 & 1 & \omega \\ \omega & \omega^2 & 1 \end{pmatrix}, \quad \omega^3 = 1.$$

8. Find the eigenvalues and the eigenvectors of the circulant matrix
$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$
Find the matrix F that diagonalizes the above matrix.

9. Let P be the $n \times n$ primary permutation matrix. Show that
$$P^n = I, \quad P^T = P^{-1} = P^{n-1}.$$

10. Find a matrix X such that
$$\begin{pmatrix} c_0 & c_1 & c_2 \\ c_1 & c_2 & c_0 \\ c_2 & c_0 & c_1 \end{pmatrix} = X \begin{pmatrix} c_0 & c_1 & c_2 \\ c_2 & c_0 & c_1 \\ c_1 & c_2 & c_0 \end{pmatrix}.$$

11. Find an invertible matrix Q such that
$$Q^* \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} Q = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

12. Let F be the $n \times n$ Fourier matrix. Show that
    (a) F is symmetric; namely, $F^T = F$.
    (b) $(F^*)^2 = F^2$ is a permutation matrix.
    (c) $(F^*)^3 = F$ and $F^4 = I$.
    (d) The eigenvalues of F are ± 1 and $\pm i$ with appropriate multiplicity (which is the number of times the eigenvalue repeats).
    (e) $F^* = n^{-1/2} V(1, \overline{\omega}, \overline{\omega}^2, \ldots, \overline{\omega}^{n-1})$, where V stands for the *Vandermonde matrix* (see the next section).
    (f) If $F = R + iS$, where R and S are real, then $R^2 + S^2 = I$, $RS = SR$, and R and S are symmetric.

13. Let e_i be the column vectors of n components with the ith component 1 and 0 elsewhere, $i = 1, 2, \ldots, n$, $c_1, c_2, \ldots, c_n \in \mathbb{C}$, and let

$$A = \operatorname{diag}(c_1, c_2, \ldots, c_n)P,$$

where P is the $n \times n$ primary permutation matrix. Show that
    (a) $Ae_i = c_{i-1}e_{i-1}$ for each $i = 1, 2, \ldots, n$, where $c_0 = c_n,\ e_0 = e_n$.
    (b) $A^n = cI$, where $c = c_1 c_2 \cdots c_n$.
    (c) $\det(I + A + \cdots + A^{n-1}) = (1 - c)^{n-1}$.

14. A matrix A is called a *Toeplitz matrix* if all entries of A are constant down the diagonals parallel to the main diagonal. In symbols,

$$A = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_n \\ a_{-1} & a_0 & a_1 & \cdots & a_{n-1} \\ a_{-2} & a_{-1} & a_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & a_1 \\ a_{-n} & a_{-n+1} & \cdots & a_{-1} & a_0 \end{pmatrix}.$$

For example, matrix $F = (f_{ij})$ with $f_{i,i+1} = 1$, $i = 1, 2, \ldots, n-1$, and 0 elsewhere, is a Toeplitz matrix. Show that (i) a matrix A is a Toeplitz matrix if and only if A can be written in the form

$$A = \sum_{k=1}^{n} a_{-k}(F^T)^k + \sum_{k=0}^{n} a_k F^k,$$

(ii) the sum of two Toeplitz matrices is a Toeplitz matrix, (iii) a circulant matrix is a Toeplitz matrix, and (iv) BA is a symmetric matrix, known as a *Hankel matrix*, where B is the backward identity.

. ———— ⊙ ———— .

5.4 Vandermonde Matrices

An n-square *Vandermonde matrix* is a matrix of the form

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ a_1 & a_2 & a_3 & \cdots & a_n \\ a_1^2 & a_2^2 & a_3^2 & \cdots & a_n^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_1^{n-1} & a_2^{n-1} & a_3^{n-1} & \cdots & a_n^{n-1} \end{pmatrix},$$

denoted by $V_n(a_1, a_2, \ldots, a_n)$ or simply V.

Vandermonde matrices play a role in many places such as interpolation problems and solving systems of linear equations. We consider the determinant and the inverse of a Vandermonde matrix in this section.

Theorem 5.9 *Let $V_n(a_1, a_2, \ldots, a_n)$ be a Vandermonde matrix. Then*

$$\det V_n(a_1, a_2, \ldots, a_n) = \prod_{1 \le i < j \le n} (a_j - a_i),$$

and $V_n(a_1, a_2, \ldots, a_n)$ is invertible if and only if all the a_i are distinct.

Proof. We proceed with the proof by induction. There is nothing to show if $n = 1$ or 2. Let $n \ge 3$.

Suppose the assertion is true when the size of the matrix is $n-1$. For the case of n, subtracting row i multiplied by a_1 from row $i+1$, for i going down from $n-1$ to 1, we have

$$\begin{aligned} \det V &= \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & a_2 - a_1 & a_3 - a_1 & \cdots & a_n - a_1 \\ 0 & a_2(a_2 - a_1) & a_3(a_3 - a_1) & \cdots & a_n(a_n - a_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & a_2^{n-2}(a_2 - a_1) & a_3^{n-2}(a_3 - a_1) & \cdots & a_n^{n-2}(a_n - a_1) \end{vmatrix} \\ &= \begin{vmatrix} a_2 - a_1 & a_3 - a_1 & \cdots & a_n - a_1 \\ a_2(a_2 - a_1) & a_3(a_3 - a_1) & \cdots & a_n(a_n - a_1) \\ \vdots & \vdots & \vdots & \vdots \\ a_2^{n-2}(a_2 - a_1) & a_3^{n-2}(a_3 - a_1) & \cdots & a_n^{n-2}(a_n - a_1) \end{vmatrix} \end{aligned}$$

$$\begin{aligned}
&= \prod_{j=2}^{n}(a_j - a_1)\det V_{n-1}(a_2, a_3, \dots, a_n) \\
&= \prod_{j=2}^{n}(a_j - a_1) \prod_{2\le i<j\le n}(a_j - a_i) \ \text{ (by the hypothesis)} \\
&= \prod_{1\le i<j\le n}(a_j - a_i).
\end{aligned}$$

It is readily seen that the Vandermonde matrix is singular if and only if at least two of the a_i are equal. ∎

An interesting application follows: Let $A \in \mathbb{M}_n$. Then

$$A^n = 0 \quad \Leftrightarrow \quad \operatorname{tr} A^k = 0, \ \ k = 1, 2, \dots, n.$$

Because $A^n = 0$, A is nilpotent, thus, A has only zero eigenvalues; so does A^k for each k. For the other way around, let the eigenvalues of A be $\lambda_1, \lambda_2, \dots, \lambda_n$. Then the trace identities imply

$$\begin{gathered}
\lambda_1 + \lambda_2 + \cdots + \lambda_n = 0, \\
\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_n^2 = 0, \\
\vdots \\
\lambda_1^n + \lambda_2^n + \cdots + \lambda_n^n = 0,
\end{gathered}$$

rewritten as

$$V_n(\lambda_1, \lambda_2, \dots, \lambda_n)(\lambda_1, \lambda_2, \dots, \lambda_n)^T = 0.$$

If all of the λ_i are distinct, then by the preceding theorem the Vandermonde matrix is nonsingular and the system of equations in $\lambda_1, \lambda_2, \dots, \lambda_n$ has only the trivial solution $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$. If some of the λ_i are identical, for instance, $\lambda_1 = \lambda_2$ and $\lambda_2, \lambda_3, \dots, \lambda_n$ are distinct, we then write the system as

$$V_{n-1}(\lambda_2, \dots, \lambda_n)(2\lambda_2, \dots, \lambda_n)^T = 0.$$

A similar argument will result in $\lambda_2 = \cdots = \lambda_n = 0$.

This idea applies to the *interpolation* problem of finding a polynomial $f(x)$ of degree at most $n - 1$ satisfying

$$f(x_i) = y_i, \quad i = 1, 2, \dots, n,$$

where the x_i and y_i are given constants (Problem 4).

Theorem 5.10 *For any integers $k_1 < k_2 < \cdots < k_n$, the quotient*

$$\frac{\det V_n(k_1, k_2, \ldots, k_n)}{\det V_n(1, 2, \ldots, n)}$$

is an integer.

Proof. Let f_i be any monic polynomial of degree i for $i = 1, 2, \ldots, n-1$. The additive property of determinants (see Section 1.2) shows that

$$\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ f_1(k_1) & f_1(k_2) & f_1(k_3) & \cdots & f_1(k_n) \\ f_2(k_1) & f_2(k_2) & f_2(k_3) & \cdots & f_2(k_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_{n-1}(k_1) & f_{n-1}(k_2) & f_{n-1}(k_3) & \cdots & f_{n-1}(k_n) \end{vmatrix} \tag{5.8}$$

is the same as $\det V_n(k_1, k_2, \ldots, k_n)$. By taking, for any integer a,

$$f_i(a) = a(a-1)(a-2)\cdots(a-i+1) = i!\binom{a}{i},$$

we see that $f_i(a)$ is divisible by $(i-1)!$.

Factoring out $(i-1)!$ from row i, $i = 2, 3, \ldots, n$, we see that the determinant in (5.8), thus $\det V_n(k_1, k_2, \ldots, k_n)$, is divisible by the product $\prod_{i=1}^{n}(i-1)!$.

The proof is complete, for $\prod_{i=1}^{n}(i-1)! = \det V_n(1, 2, \ldots, n)$. ∎

We now turn our attention to the inverse of a Vandermonde matrix. Consider the polynomial in x given by the product

$$p(x) = (x + a_1)(x + a_2)\cdots(x + a_n),$$

where $a_1, a_2, \ldots, a_n$ are constants. Expand $p(x)$ as a polynomial

$$p(x) = s_0 x^n + s_1 x^{n-1} + \cdots + s_{n-1}x + s_n,$$

where $s_0 = 1$ and for each $k = 1, 2, \ldots, n$,

$$s_k = s_k(a_1, a_2, \ldots, a_n) = \sum_{1 \le p_1 < \cdots < p_k \le n} \prod_{q=1}^{k} a_{p_q}.$$

We refer to s_k, depending on $a_1, a_2, \ldots, a_n$, as the kth *elementary symmetric function* of $a_1, a_2, \ldots, a_n$. (See also Section 4.4.)

One may expand $p(x) = (x + a_1)(x + a_2)(x + a_3)$ as an example.

Theorem 5.11 *Suppose that $a_1, a_2, \ldots, a_n$ are distinct. Then*

$$(V_n(a_1, a_2, \ldots, a_n))^{-1} = (\alpha_{ij}),$$

where for each pair of i and j

$$\alpha_{ij} = \frac{(-1)^{1+j} \sum_{p_1 < \cdots < p_{n-j}} \prod_{q=1,\, p_q \neq i}^{n-j} a_{p_q}}{\prod_{k=1,\, k\neq i}^{n} (a_k - a_i)}.$$

Proof. Recall from elementary linear algebra (Section 1.2) that the entries of the inverse of the matrix V are the cofactors of order $n-1$ divided by $\det V$; that is,

$$V^{-1} = \left(\frac{1}{\det V} c_{ij} \right)^T,$$

where c_{ij} is the cofactor of the (i, j)-entry of V.

In what follows we compute the cofactors c_{ij}. Let V_k be the matrix obtained from V by deleting row $k+1$ (the kth powers) and adjoining as a new nth row the nth powers of the a_i. We show

$$\det V_k = s_{n-k} \det V. \tag{5.9}$$

Augment V with the nth powers of the a_i as the $(n+1)$th row and with $(1,\ -x,\ (-x)^2,\ \ldots,\ (-x)^n)^T$ as the first column. Denote the resulting matrix by W. Then W is a Vandermonde matrix and

$$\begin{aligned} \det W &= (x + a_1) \cdots (x + a_n) \det V \\ &= (x^n + s_1 x^{n-1} + \cdots + s_{n-1} x + s_n) \det V. \end{aligned} \tag{5.10}$$

Expanding $\det W$ along the first column, we have

$$\det W = \det V_0 + x \det V_1 + \cdots + x^n \det V. \tag{5.11}$$

Identity (5.9) follows by comparing (5.10) and (5.11).

Now notice that each cofactor c_{ij} is a determinant of order $n-1$ in the same form as $\det V_k$. Let $V(\widehat{a_j})$ and $s_k(\widehat{a_j})$ denote, respectively, the $(n-1)$-square Vandermonde matrix and the kth elementary symmetric function of $a_1, a_2, \ldots, a_n$ without a_j. Using (5.9) we have

$$\begin{aligned} c_{ij} &= (-1)^{i+j} \det V(i|j) \\ &= (-1)^{i+j} s_{(n-1)-(i-1)}(\widehat{a_j}) \det V(\widehat{a_j}) \\ &= (-1)^{i+j} s_{n-i}(\widehat{a_j}) \det V(\widehat{a_j}). \end{aligned}$$

Thus,

$$\begin{aligned} \frac{1}{\det V} c_{ij} &= \frac{(-1)^{i+j} s_{n-i}(\widehat{a_j}) \det V(\widehat{a_j})}{\prod_{t>s}(a_t - a_s)} \\ &= \frac{(-1)^{i+j} s_{n-i}(\widehat{a_j})}{\prod_{s<j}(a_j - a_s) \prod_{j<t}(a_t - a_j)} \\ &= \frac{(-1)^{i+1} s_{n-i}(\widehat{a_j})}{\prod_{k=1,\, k\neq j}^{n}(a_k - a_j)} \\ &= \frac{(-1)^{i+1} \sum_{p_1<\cdots<p_{n-i}} \prod_{q=1,\, p_q\neq j}^{n-i} a_{p_q}}{\prod_{k=1,\, k\neq j}^{n}(a_k - a_j)} \end{aligned}$$

or

$$\alpha_{ij} = \frac{(-1)^{1+j} \sum_{p_1<\cdots<p_{n-j}} \prod_{q=1,\, p_q\neq i}^{n-j} a_{p_q}}{\prod_{k=1,\, k\neq i}^{n}(a_k - a_i)}. \quad \blacksquare$$

Problems

1. Find the solution to the equation in x:

$$\begin{vmatrix} x^2 & 4 & 9 \\ x & 2 & 3 \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

2. Evaluate the determinant

$$\begin{vmatrix} 1 & ax & a^2 + x^2 \\ 1 & ay & a^2 + y^2 \\ 1 & az & a^2 + z^2 \end{vmatrix}.$$

3. Find all solutions to the system of equations in $x_1, \dots, x_n$,

$$\begin{cases} x_1 + \cdots + x_n = a \\ x_1^2 + \cdots + x_n^2 = a^2 \\ \vdots \\ x_1^n + \cdots + x_n^n = a^n. \end{cases}$$

4. Let $x_1, x_2, \dots, x_n$ be different numbers. Show that for any set of n numbers $y_1, y_2, \dots, y_n$, there exists a polynomial $f(x)$ of degree at most $n-1$ such that $f(x_i) = y_i,\ i = 1, 2, \dots, n$. In particular, for any numbers $\lambda_1, \lambda_2, \dots, \lambda_n$, there exist polynomials $g(x)$, and $h(x)$ if each $\lambda_i \geq 0$, of degree at most $n-1$ such that

$$g(\lambda_i) = \overline{\lambda_i}, \quad h(\lambda_i) = \sqrt{\lambda_i}, \quad i = 1, 2, \dots, n.$$

5. Let $A = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, where $\lambda_i \neq \lambda_j$ for $i \neq j$. Show that for every normal matrix $B \in \mathbb{M}_n$, there exist a unitary matrix U and a polynomial f such that $B = U^* f(A) U$.

6. Let $U = (u_{ij})$ be the p-square unitary matrix with

$$u_{ij} = \omega^{(i-1)(j-1)}, \quad i, j = 1, 2, \dots, p,$$

where p is a prime integer and ω is a pth primitive root of unity. Show that all square submatrices of U are nonsingular.

7. Let $S_1 = \{\alpha_i\}_{i=1}^r$ and $S_2 = \{\beta_i\}_{i=1}^s$ be nonzero complex number multisets (i.e., repetition of elements is allowed. Say, $\{1, 2, 2, 3\}$). If

$$\sum_{i=1}^{r} \alpha_i^k = \sum_{i=1}^{s} \beta_i^k, \quad \text{for every positive integer } k \leq r+s,$$

show that $r = s$ and $S_1 = S_2$; that is, the two sets are the same.

8. Show that two n-square complex matrices A and B have the same set of eigenvalues if and only if $\operatorname{tr} A^k = \operatorname{tr} B^k,\ k = 1, 2, \dots, n$.

9. Find the inverse, if it exists, of the Vandermonde matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & x & y \\ 1 & x^2 & y^2 \end{pmatrix}.$$

10. Expand and find elementary symmetric functions for

$$p(x) = (x + a_1)(x + a_2)(x + a_3)(x + a_4).$$

11. Let W be the matrix obtained from $V = V_n(a_1, \dots, a_n)$ by replacing the last row $(a_1^{n-1}, \dots, a_n^{n-1})$ with $(a_1^n, \dots, a_n^n)$. Show directly without using (5.9) that $\det W = (a_1 + \cdots + a_n) \det V$.

12. Let $f_1(x)$, $f_2(x)$, $\dots$, $f_n(x)$ be polynomials with degree at most $n-2$. Show that for any numbers a_1, a_2, $\dots$, a_n

$$\begin{vmatrix} f_1(a_1) & f_1(a_2) & \cdots & f_1(a_n) \\ f_2(a_1) & f_2(a_2) & \cdots & f_2(a_n) \\ \vdots & \vdots & \vdots & \vdots \\ f_n(a_1) & f_n(a_2) & \cdots & f_n(a_n) \end{vmatrix} = 0.$$

13. Let $A = (a_{ij}) \in \mathbb{M}_n$ and $f_i(x) = a_{1i} + a_{2i}x + \cdots + a_{ni}x^{n-1}$ for $i = 1, 2, \dots, n$. Show that

$$\begin{vmatrix} f_1(x_1) & f_1(x_2) & \cdots & f_1(x_n) \\ f_2(x_1) & f_2(x_2) & \cdots & f_2(x_n) \\ \vdots & \vdots & \vdots & \vdots \\ f_n(x_1) & f_n(x_2) & \cdots & f_n(x_n) \end{vmatrix} = \det A \prod_{1 \le i < j \le n} (x_j - x_i).$$

14. Let $a_1, a_2, \dots, a_n$ be complex numbers. Show that

$$\begin{vmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_n \\ \vdots & \vdots & \vdots \\ a_1^{n-2} & \cdots & a_n^{n-2} \\ a_2 a_3 \cdots a_n & \cdots & a_1 a_2 \cdots a_{n-1} \end{vmatrix} = (-1)^{n-1} \det V_n(a_1, \dots, a_n).$$

15. Let $V = V_n(x_1, \dots, x_n)$ be the $n \times n$ Vandermonde matrix of $x_1, \dots, x_n$, where $x_i \neq x_j$ whenever $i \neq j$. Define $F(x) = \prod_{i=1}^n (x - x_i)$ and $f_k(x) = F(x)/(x_k - x)$. Show that $f_k(x_j) = 0$ if $j \neq k$ and $f_k(x_k) = -F'(x_k)$, where F' is the 1st derivative of F. Expand $-f_k(x)/F'(x_k)$ and form a matrix M by its coefficients as the kth row of M, $k = 1, \dots, n$. Show that M is the inverse of V.

16. Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \dots, \lambda_n$. Show that

$$\det(\lambda I - A) = \lambda^n - \sigma_1 \lambda^{n-1} + \sigma_2 \lambda^{n-2} - \cdots + (-1)^n \sigma_n,$$

where $\sigma_k = s_k(\lambda_1, \lambda_2, \dots, \lambda_n)$, $k = 1, 2, \dots, n$. Describe σ_k in terms of principal minors (see Problem 19, Section 1.3.)

5.5 Hadamard Matrices

An n-square matrix A is called a *Hadamard matrix* if each entry of A is 1 or -1 and if the rows or columns of A are orthogonal; that is,

$$AA^T = nI \quad \text{or} \quad A^T A = nI.$$

Note that $AA^T = nI$ and $A^T A = nI$ are equivalent (Problem 5).

The following are two examples of Hadamard matrices:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}.$$

Notice that if A is a Hadamard matrix, then so is AP for any matrix P with entries ± 1 satisfying $PP^T = I$. Thus, one may change the -1 in the first row of A to $+1$ by multiplying an appropriate matrix P with diagonal entries ± 1. There is only one 2×2 Hadamard matrix of this kind. Can one construct a 3×3 Hadamard matrix?

Theorem 5.12 *Let $n > 2$. A necessary condition for an n-square matrix A to be a Hadamard matrix is that n is a multiple of* 4.

Proof 1. Let $A = (a_{ij})$ be an n-square Hadamard matrix. Noticing that the entries of A are ± 1, the equation $AA^T = nI$ yields

$$\sum_{k=1}^n a_{ik} a_{jk} = \begin{cases} 0, & \text{if } i \neq j, \\ n, & \text{if } i = j. \end{cases}$$

Upon computation, we have

$$\begin{aligned} \sum_{k=1}^n (a_{1k} + a_{2k})(a_{1k} + a_{3k}) &= \sum_{k=1}^n a_{1k}^2 + \sum_{k=1}^n a_{1k} a_{2k} \\ &\quad + \sum_{k=1}^n a_{1k} a_{3k} + \sum_{k=1}^n a_{2k} a_{3k} \\ &= \sum_{k=1}^n a_{1k}^2 \\ &= n. \end{aligned}$$

Observe that the possible values for $a_{1k}+a_{2k}$ and $a_{1k}+a_{3k}$ are $+2$, 0, and -2. Thus, each term in the summation

$$\sum_{k=1}^{n}(a_{1k}+a_{2k})(a_{1k}+a_{3k})$$

must be $+4$, 0, or -4. It follows that n is divisible by 4.

Proof 2. Let P be an n-square matrix with main diagonal entries 1 or -1 such that the first row of AP consists entirely of $+1$. Note that AP is also a Hadamard matrix. Since the second and third rows of AP are orthogonal to the first row, they must each have the same number, say r, of $+1$s and -1s. Thus $n = 2r$ is an even number.

Let n_-^+ be the number of columns of AP that contain a $+1$ of row 2 and a -1 of row 3. Similarly, define n_+^-, n_+^+, and n_-^-. Then

$$n_+^+ + n_-^+ = n_+^+ + n_+^- = n_-^- + n_-^+ = r.$$

Thus,

$$n_+^+ = n_-^-, \quad n_-^+ = n_+^-.$$

The orthogonality of rows 2 and 3 implies that

$$n_+^+ + n_-^- = n_+^- + n_-^+.$$

This gives $n_+^+ = n_-^+$. Therefore, $n = 2r = 4n_+^+$ is a multiple of 4. ∎

It has been conjectured that a Hadamard matrix of size $4k \times 4k$ exists for every positive integer k. The conjecture is not resolved yet.

The following theorem, verified by a direct computation, gives a way to construct Hadamard matrices of larger dimensions.

Theorem 5.13 *If A is a Hadamard matrix, then so is*

$$\begin{pmatrix} A & A \\ A & -A \end{pmatrix}. \tag{5.12}$$

By this theorem, Hadamard matrices H_n of order 2^n can be generated recursively by defining

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad H_n = \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix}, \quad n \geq 2. \tag{5.13}$$

Let

$$x_1 = \begin{pmatrix} 1 \\ c \end{pmatrix} \text{ and } x_n = \begin{pmatrix} x_{n-1} \\ cx_{n-1} \end{pmatrix}, \text{ where } c = -1 + \sqrt{2}.$$

By a simple computation, H_1 has two eigenvalues $\pm\sqrt{2}$, and x_1 is an eigenvector corresponding to $\sqrt{2}$. This generalizes as follows.

Theorem 5.14 *Let H_n be defined as in (5.13). Then H_n has eigenvalues $+2^{n/2}$ and $-2^{n/2}$ each of multiplicity 2^{n-1}, and an eigenvector x_n corresponding to the positive eigenvalue $2^{n/2}$.*

Proof. The proof is done by induction on n. The case of $n = 1$ was discussed just prior to the theorem. Now for $n \geq 2$, we have

$$\begin{aligned} \det(\lambda I - H_n) &= \begin{vmatrix} \lambda I - H_{n-1} & -H_{n-1} \\ -H_{n-1} & \lambda I + H_{n-1} \end{vmatrix} \\ &= \det\Big((\lambda I - H_{n-1})(\lambda I + H_{n-1}) - H_{n-1}^2\Big) \\ &= \det(\lambda^2 I - 2H_{n-1}^2) \\ &= \det(\lambda I - \sqrt{2}H_{n-1}) \det(\lambda I + \sqrt{2}H_{n-1}). \end{aligned}$$

Thus each eigenvalue μ of H_{n-1} generates two eigenvalues $\pm\sqrt{2}\mu$ of H_n. The assertion then follows by the induction hypothesis, for H_{n-1} has eigenvalues $+2^{(n-1)/2}$ and $-2^{(n-1)/2}$ each of multiplicity 2^{n-2}.

To see the eigenvector part, we observe that, by induction again,

$$\begin{aligned} H_n x_n &= \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix} \begin{pmatrix} x_{n-1} \\ (-1 + \sqrt{2}\,)x_{n-1} \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{2}H_{n-1}x_{n-1} \\ (2 - \sqrt{2}\,)H_{n-1}x_{n-1} \end{pmatrix} \\ &= 2^{n/2} \begin{pmatrix} x_{n-1} \\ (-1 + \sqrt{2}\,)x_{n-1} \end{pmatrix} \\ &= 2^{n/2} x_n. \quad \blacksquare \end{aligned}$$

Let J_n denote the n-square matrix whose entries are all equal to 1. We give a lower bound for the size of a Hadamard matrix that contains a J_n as a submatrix.

Theorem 5.15 *If A is an m-square Hadamard matrix that contains a J_n as a submatrix, then $m \geq n^2$.*

Proof. We may assume by permutation that A is partitioned as

$$A = \begin{pmatrix} J_n & X \\ Y & Z_s \end{pmatrix}, \tag{5.14}$$

where Z_s is an s-square matrix of entries ± 1, and $s = m - n$.

Since A is a Hadamard matrix of size $m = n + s$, we have

$$AA^T = (n+s)I_m,$$

which implies, by using the block form (5.14) of A, that

$$J_n^2 + XX^T = (n+s)I_n.$$

Thus,

$$XX^T = (n+s)I_n - nJ_n. \tag{5.15}$$

The eigenvalues of the right-hand matrix in (5.15) are

$$n+s-n^2, \quad n+s, \quad \ldots, \quad n+s.$$

However, XX^T is positive semidefinite, and thus has nonnegative eigenvalues. Therefore, $n + s - n^2 \geq 0$ or $m \geq n^2$. ∎

Problems

1. Show that A is a Hadamard matrix, and then find A^4, where

$$A = \begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{pmatrix}.$$

2. Does there exist a 3×3 Hadamard matrix? How about a 6×6? Construct an 8×8 Hadamard matrix.

3. What is the determinant of an n-square Hadamard matrix?

4. Let $A = (a_{ij})$ be a 3×3 matrix. Show that if each $a_{ij} = 1$ or -1, then $\det A$ is an even number. What is the maximum for $\det A$?

5. Let A be an n-square matrix with entries ± 1. Show that

$$AA^T = nI \quad \Leftrightarrow \quad A^T A = nI.$$

Conclude that if A is Hadamard, then $\frac{1}{\sqrt{n}}A$ is orthogonal.

6. Find the eigenvalues and eigenvectors of the Hadamard matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

7. Show that the Kronecker product of two Hadamard matrices is also a Hadamard matrix.

8. Let $n \geq 2$ and define recursively, as in (5.13),

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad H_n = \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix},$$

and let

$$F_1 = \frac{1}{2}(I + 2^{-n/2}H_n), \quad F_2 = -\frac{1}{2}(I - 2^{-n/2}H_n).$$

Show that F_1 and F_2 are idempotent matrices and that

$$F_1 + F_2 = 2^{-n/2}H_n.$$

9. Let $n \geq 2$ and define recursively

$$E_1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad E_{n+1} = \begin{pmatrix} 0 & -E_n \\ E_n & 0 \end{pmatrix}.$$

Show that

(a) $E_n^2 = (-1)^n I_{2^n}$.

(b) E_n is symmetric if n is even, and skew-symmetric if n is odd.

(c) $E_n H_n = (-1)^n H_n E_n$, where H_n is defined as in (5.13).

10. Let A be a square matrix with entries 1, -1, or 0. If each row and column of A contains only one nonzero entry 1 or -1, show that $A^k = I$ for some positive integer k.

11. How many $n \times n$ matrices of 0 and 1 entries are there for which the number of 1s in each row and column is even? [Answer: $2^{(n-1)(n-1)}$.]

5.6 Permutation and Doubly Stochastic Matrices

A square matrix is called a *permutation matrix* if each row and column of the matrix has exactly one 1 and all other entries are 0.

It is easy to see that there are $n!$ permutation matrices of size n. Furthermore, the product of two permutation matrices of the same size is a permutation matrix, and if P is a permutation matrix, then P is invertible, and $P^{-1} = P^T$ (Problem 1).

Our goal in this section is to show that every permutation matrix is a direct sum of primary permutation matrices under permutation similarity and that every doubly stochastic matrix is a convex combination of permutation matrices.

A matrix A of order n is said to be *reducible* if there exists a permutation matrix P such that

$$P^T AP = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}, \tag{5.16}$$

where B and D are square matrices of order at least 1.

A matrix is said to be *irreducible* if it is not reducible. Note that a matrix of order 1 is considered to be irreducible. The matrix $P^T AP = P^{-1}AP$ in (5.16) is similar to A through the permutation matrix P. We say that they are *permutation similar*.

It is obvious that the diagonal entries of irreducible permutation matrices are all equal to 0, but not vice versa. For example,

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Theorem 5.16 *Every reducible permutation matrix is permutation similar to a direct sum of irreducible permutation matrices.*

Proof. Let A be an n-square reducible permutation matrix, as in (5.16). The matrix C in this case must be zero, for otherwise, let B

be $r \times r$ and D be $s \times s$, where $r + s = n$. Then B contains r 1s (in columns) and D contains s 1s (in rows). If C contained a 1, then A would have at least $r + s + 1 = n + 1$ 1s, a contradiction. The assertion then follows by the induction on B and D. ∎

We now show that every n-square irreducible permutation matrix is permutation similar to the $n \times n$ *primary permutation matrix*

$$P = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{5.17}$$

Theorem 5.17 *A primary permutation matrix is irreducible.*

Proof. Suppose the $n \times n$ primary permutation matrix P is reducible. Let $S^T PS = J_1 \oplus \cdots \oplus J_k$, $k \geq 2$, where S is some permutation matrix and the J_i are irreducible matrices with order $< n$.

The rank of $P - I$ is $n - 1$, for $\det(P - I) = 0$ and the submatrix of size $n - 1$ by deleting the last row and the last column from $P - I$ is nonsingular. It follows that

$$\operatorname{rank}(S^T PS - I) = \operatorname{rank}(S^T (P - I)S) = n - 1.$$

By using the above decomposition, we have

$$\operatorname{rank}(S^T PS - I) = \sum_{i=1}^{k} \operatorname{rank}(J_i - I) \leq n - k < n - 1.$$

This is a contradiction. The proof is complete. ∎

The eigenvalues of the $n \times n$ primary permutation matrix P are exactly all the roots of the equation $\lambda^n = 1$; that is, $1, \omega, \dots, \omega^{n-1}$, where ω is an nth primitive root of unity, because

$$\det(\lambda I - P) = \lambda^n - 1$$

by a direct computation. In addition, for any positive integer $k < n$,

$$P^{n-1} = P^T, \quad P^n = I_n, \quad P^k = \begin{pmatrix} 0 & I_{n-k} \\ I_k & 0 \end{pmatrix}.$$

Theorem 5.18 *A permutation matrix is irreducible if and only if it is permutation similar to a primary permutation matrix.*

Proof. Let Q be an $n \times n$ permutation matrix and P the $n \times n$ primary permutation matrix in (5.17). If Q is permutation similar to P, then Q is irreducible by the previous theorem. Conversely, suppose that Q is irreducible. We show that Q can be brought to P through simultaneous row and column permutations.

Let the 1 of the first row be in the position $(1, i_1)$. Then $i_1 \neq 1$ since Q is irreducible. If $i_1 = 2$, we proceed to the next step, considering the 1 in the second row. Otherwise, $i_1 > 2$. Permute columns 2 and i_1 so that the 1 is placed in the $(1, 2)$ position.

Permute rows 2 and i_1 to get a matrix Q_1. This matrix is permutation similar to Q and also irreducible. If the $(2, 3)$-entry of Q_1 is 1, we go on to the next step. Otherwise, let the $(2, i_2)$-entry be 1, $i_2 \neq 3$. If $i_2 = 1$, then Q_1 would be reducible, for all entries in the first two columns but not in the first two rows equal 0. Thus, $i_2 \geq 3$. Permute columns 3 and i_2 so that the 1 is in the $(2, 3)$ position.

Note that the 1 in the $(1, 2)$ position was not affected by the permutations in the second step. Continuing in this way, one obtains the permutation matrix P in the form of (5.17). The product of a sequence of permutation matrices is also a permutation matrix, therefore we have a permutation matrix S such that

$$S^T QS = S^{-1} QS = P. \quad \blacksquare$$

Combining the above theorems, we see that every reducible permutation matrix is permutation similar to a direct sum of primary permutation matrices. Moreover, the rank of an n-square irreducible permutation matrix minus I is $n - 1$ (Problem 4).

Theorem 5.19 *Let Q be an n-square permutation matrix. Then Q is irreducible if and only if the eigenvalues of Q are 1, ω, $\ldots$, ω^{n-1}, where ω is an nth primitive root of unity.*

Proof. If Q is irreducible, then Q is similar to the $n \times n$ primary permutation matrix, according to Theorem 5.18, which has the eigenvalues 1, ω, $\ldots$, ω^{n-1}; so does matrix Q.

Conversely, suppose that 1, ω, ..., ω^{n-1} are the eigenvalues of Q. Note that $\omega^k \neq 1$ for any $1 \leq k < n$ since ω is an nth primitive root of unity. If Q is reducible, then we may write

$$S^T QS = J_1 \oplus \cdots \oplus J_k,$$

where S is a permutation matrix, and the J_i are primary permutation matrices with order less than n.

The eigenvalues of those J_i are the eigenvalues of Q, none of which is an nth primitive root of unity, for the order of every J_i is less than n. This is a contradiction. Thus, Q is irreducible. ∎

We next present a beautiful relation between permutation matrices and doubly stochastic matrices, a type of matrices that plays an important role in statistics and in some other subjects.

A square matrix is said to be *doubly stochastic* if all entries of the matrix are nonnegative and the sum of the entries in each row and each column equals 1. Equivalently, a matrix A with nonnegative entries is doubly stochastic if

$$e^T A = e^T \quad \text{and} \quad Ae = e, \quad \text{where } e = (1, 1, \ldots, 1)^T. \tag{5.18}$$

It is readily seen that permutation matrices are doubly stochastic and so is the product of two doubly stochastic matrices.

We show that a matrix is a doubly stochastic matrix if and only if it is a convex combination of finite permutation matrices. To prove this, we need a result, which is of interest in its own right.

Theorem 5.20 (Frobenius–König) *Let A be an n-square complex matrix. Then every product of n entries of A taken from distinct rows and columns equals 0; in symbols,*

$$a_{1i_1} a_{2i_2} \cdots a_{ni_n} = 0, \quad \{i_1, i_2, \ldots, i_n\} = \{1, 2, \ldots, n\}, \tag{5.19}$$

if and only if A contains an $r \times s$ zero submatrix, where $r + s = n + 1$.

Proof. First notice that property (5.19) of A will remain true when row or column permutations are applied to A. In other words, an n-square matrix A has property (5.19) if and only if PAQ has the property, where P and Q are any n-square permutation matrices.

Sufficiency: We may assume by permutation that the $r \times s$ zero submatrix is in the lower-left corner, and write

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}.$$

Because $n - r = s - 1$, B is of size $(s-1) \times s$. Thus, there must be a zero among any s entries taken from the first s columns and any s different rows. Therefore, every product $a_{1i_1}a_{2i_2}\cdots a_{ni_n}$ has to contain a zero factor, hence equals zero.

Necessity: If all the entries of A are zero, there is nothing to prove. Suppose A has a nonzero entry and consider the submatrix obtained from A by deleting the row and the column that contain the nonzero entry. An application of induction on the $(n-1)\times(n-1)$ submatrix results in a zero submatrix of size $p \times q$, where $p+q = (n-1)+1 = n$. We thus may write A, by permutation, as

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix},$$

where B is $q \times q$ and D is $p \times p$. Since every product of the entries of A from different rows and columns is 0, this property must be inherited by B or D, say B. Applying the induction to B, we see that B has an $l \times s$ zero submatrix such that $l + s = q + 1$. Putting this zero submatrix in the lower-left corner of B, we see that A has an $r \times s$ zero submatrix, where $r = p + l$ and $r + s = n + 1$. ∎

Theorem 5.21 (Birkhoff) *A matrix A is doubly stochastic if and only if it is a convex combination of permutation matrices.*

Proof. To show sufficiency, let A be a *convex combination* of permutation matrices $P_1, P_2, \ldots, P_m$; that is,

$$A = t_1P_1 + t_2P_2 + \cdots + t_mP_m,$$

where $t_1, t_2, \ldots, t_m$ are nonnegative numbers of a sum equal to 1. Then it is easy to see that $e^TA = e^T$ and $Ae = e$, where $e = (1, \ldots, 1)^T$. By (5.18) A is doubly stochastic.

For necessity, we apply induction on the number of zero entries of the doubly stochastic matrices. If A has (at most) $n^2 - n$ zeros, then

A is a permutation matrix, and we have nothing to show. Suppose that the doubly stochastic matrices with at least k zeros are convex combinations of permutation matrices. We show that the assertion holds for the doubly stochastic matrices with $k-1$ zeros.

Let A be an n-square doubly stochastic matrix of $k-1$ zero entries. If every product of the entries of A from distinct rows and columns is zero, then A may be written as, up to permutation,

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix},$$

where the zero submatrix is of size $r \times s$ with $r + s = n + 1$.

Since the entries in each column A add up to 1, the sum of all entries of B equals s. Similarly, by considering rows, the sum of all entries of D is r. Thus, the sum of all entries of A would be at least $r + s = n + 1$. This is impossible, for the sum of all entries of A is n. Therefore, some product $a_{1i_1 1}a_{2i_2} \cdots a_{ni_n} \neq 0$.

Let P_1 be a permutation matrix with 1 in the positions (j, i_j), $j = 1, 2, \ldots, n$, and 0 elsewhere. Consider the matrix

$$E = (1 - \delta)^{-1}(A - \delta P_1),$$

where $\delta = \min\{a_{1i_1},\ a_{2i_2},\ \ldots,\ a_{ni_n}\}$.

It is readily seen by (5.18) that E is also a doubly stochastic matrix and that E has at least one more zero than A. By the induction hypothesis, there are positive numbers $t_2, \ldots, t_m$ of sum 1, and permutation matrices $P_2, \ldots, P_m$, such that

$$E = t_2 P_2 + \cdots + t_m P_m.$$

It follows that

$$A = \delta P_1 + (1 - \delta)t_2 P_2 + \cdots + (1 - \delta)t_m P_m,$$

where P_i are permutation matrices, and their coefficients are nonnegative and sum up to 1. ∎

Problems

1. Show that the determinant of a permutation matrix is ± 1 and that permutation matrices are unitary and hence normal.

2. Find permutations that bring the reducible permutation matrix

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

to a direct sum of irreducible matrices. Show that $P^2 = I$.

3. Let $A = B = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Show that A and B are both irreducible but AB, A^2, and $A \otimes B$ are reducible.

4. Let P be an $n \times n$ irreducible permutation matrix. Show that

$$\operatorname{rank}(P - I) = n - 1.$$

5. Let P be an $n \times n$ irreducible permutation matrix. Show that P^k is irreducible if and only if $(n, k) = 1$; that is, n and k have no common positive divisors other than 1.

6. Let P be an n-square permutation matrix. Show that P is irreducible if and only if P has the property $P^m = I_n \iff n|m$.

7. Let P be an $n \times n$ irreducible permutation matrix. Show that P is diagonalizable over $\mathbb{C}$ but not over $\mathbb{R}$ when $n \geq 3$.

8. Show that if two permutation matrices A and B are similar, i.e., $S^{-1}AS = B$ for some nonsingular S, then they are permutation similar; that is, S can be chosen to be a permutation matrix.

9. Show that for any $n \times n$ permutation matrix P, $P^{n!} = I$, and further that if P is irreducible, then $P^n = I$. Is the converse true?

10. Prove or disprove that a symmetric permutation matrix (of odd or even order greater than 1) is reducible.

11. Find the rank of the partitioned permutation matrix Q, where

$$Q = \begin{pmatrix} 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \\ I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \end{pmatrix}.$$

12. Show that

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

is not permutation similar to its Jordan canonical form

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

13. Show that an n-square matrix A with nonnegative entries is reducible if and only if there exists a proper subset $\{i_1, \ldots, i_k\}$ of $\{1, \ldots, n\}$ such that

$$\operatorname{Span}\{Ae_{i_1}, \ldots, Ae_{i_k}\} \subseteq \operatorname{Span}\{e_{i_1}, \ldots, e_{i_k}\},$$

where the e_i are the standard basis vectors for $\mathbb{C}^n$.

14. Show that the product of two doubly stochastic matrices is a doubly stochastic matrix. How about the sum?

15. Show that if $U = (u_{ij})$ is a unitary matrix, then $A = (|u_{ij}|^2)$ is a doubly stochastic matrix. How about $B = (|u_{ij}||v_{ij}|)$, where $V = (v_{ij})$ is a unitary matrix of the same size as U?

16. Let P and Q be permutation matrices of the same size. Show that

$$\alpha P + (1-\alpha)Q$$

is a doubly stochastic matrix for any $\alpha \in [0,1]$.

17. Show that the following determinant is zero:

$$\begin{vmatrix} a & b & 0 & 0 & 0 \\ c & d & 0 & 0 & 0 \\ e & f & 0 & 0 & 0 \\ g & h & i & j & k \\ l & m & n & o & p \end{vmatrix}.$$

18. Show that every $n \times n$ doubly stochastic matrix is a convex combination of at most $n^2 - 2n + 2$ permutation matrices.

19. If A is an $n \times n$ nonsingular matrix, how many zero entries can A have at most?

20. Show that every nilpotent matrix with nonnegative entries is permutation similar to a strictly upper-triangular matrix. [Hint: Let $e_1, \ldots, e_n$ be the standard basis vectors of $\mathbb{C}^n$. Show $Ae_i = 0$ for some i; that is, A contains a zero column, by induction.]

21. A *permutation of an n-element set* $\{1, 2, \ldots, n\}$ is a mapping

$$p : 1 \to i_1,\ 2 \to i_2,\ \ldots,\ n \to i_n,$$

written as

$$p = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}.$$

Assign p a permutation matrix $P = f(p)$, which has 1 in the (k, i_k) position and 0 elsewhere, $k = 1, 2, \ldots, n$. Define the product of

$$p = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix} \quad \text{and} \quad q = \begin{pmatrix} i_1 & i_2 & \cdots & i_n \\ j_1 & j_2 & \cdots & j_n \end{pmatrix}$$

by

$$pq = \begin{pmatrix} 1 & 2 & \cdots & n \\ j_1 & j_2 & \cdots & j_n \end{pmatrix}.$$

Show that

$$f(pq) = f(p)f(q).$$

22. A permutation is called an *interchange* if only two elements are permuted. For instance, the following p is an interchange, where

$$p = \begin{pmatrix} 1 & 2 & 3 & 4 & \cdots & n \\ 3 & 2 & 1 & 4 & \cdots & n \end{pmatrix}.$$

Show that every permutation on $\{1, 2, \ldots, n\}$ can be obtained by a sequence of interchanges (permutation of two numbers each time).

23. Show that any $n \times n$ permutation matrix can be expressed as the product of at most $n - 1$ symmetric permutation matrices.

5.7 Nonnegative Matrices

A *nonnegative matrix* is a matrix all of whose entries are nonnegative. Permutation matrices and doubly stochastic matrices are nonnegative matrices. If $A = (a_{ij})$ is a nonnegative matrix, i.e., $a_{ij} \geq 0$ for all i and j, we write $A \geq 0$ (or $0 \leq A$). If all the entries of A are positive, we call A a *positive* matrix and denote $A > 0$ (or $0 < A$).

Any matrix A (including row and column vectors) is associated with a nonnegative matrix (vector), written as $|A|$, whose entries are the absolute values of the entries of A; that is, $|A| = (|a_{ij}|)$.

Important note: Let A be a matrix (or a vector). In this section, and only in this section, of the book, $A \geq (>)0$ means that A is a nonnegative (positive) matrix and $|A|$ stands for the resulting matrix by taking the absolute values of the entries of A. In all other sections of the book, $A \geq 0$ means that A is positive semidefinite and $|A|$ is the modulus of A, i.e., $|A| = (A^*A)^{1/2}$. Each of these notations is about equally and widely used in the literature for both meanings. One may use $A \geq_e 0$ with a subscript e for an entrywise nonnegative matrix and $|A|_e = (|a_{ij}|)$ for an entrywise absolute value matrix.

For matrices A and B of the same size, we write

$$A \geq (>) B \quad \text{if} \quad A - B \geq (>) 0.$$

Obviously, when A and B are matrices of the same size

$$|A + B| \leq |A| + |B|$$

and for matrices A and B of sizes $m \times n$ and $n \times m$, respectively,

$$|AB| \leq |A||B|.$$

In particular, for any n-square complex matrix A and vector x in $\mathbb{C}^n$,

$$|Ax| \leq |A||x|$$

and for any square complex matrix A and positive integer k,

$$|A^k| \leq |A|^k.$$

Furthermore, if $A > 0$ and $x \neq 0$, then $A|x| > 0$.

We are now ready to present a theorem on comparison of the spectral radii of nonnegative matrices. Recall that the *spectral radius* $\rho(A)$ of an n-square complex matrix A is defined to be

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

Note that $\rho(A)$ is in general not an eigenvalue of A. Our main goal of this section is to show that if $A \geq 0$, then $\rho(A)$ is an eigenvalue of A and that a positive eigenvector belonging to this eigenvalue exists.

Theorem 5.22 *Let $A \geq 0$ and $B \geq 0$ be n-square matrices. Then*

$$A \geq B \quad \Rightarrow \quad \rho(A) \geq \rho(B).$$

Proof. Since $A \geq B \geq 0$, i.e., $a_{ij} \geq b_{ij} \geq 0$ for all i and j, we have

$$\sum_{i,j} a_{ij}^2 \geq \sum_{i,j} b_{ij}^2.$$

It follows that

$$\Big(\sum_{i,j} a_{ij}^2\Big)^{1/2} \geq \Big(\sum_{i,j} b_{ij}^2\Big)^{1/2}$$

or in Frobenius norm,

$$\|A\|_F \geq \|B\|_F.$$

On the other hand (Problem 4), for all positive integers k, we have

$$A \geq B \geq 0 \quad \Rightarrow \quad A^k \geq B^k.$$

This yields

$$\|A^k\|_F \geq \|B^k\|_F$$

or

$$\|A^k\|_F^{1/k} \geq \|B^k\|_F^{1/k}.$$

Taking limits and using Theorem 4.4, we have $\rho(A) \geq \rho(B)$. ∎

Consider a pair of n-square nonnegative matrices A and B. Since the Hadamard product $A \circ B$ is a principal submatrix of the Kronecker product $A \otimes B$ (Theorem 4.7), by Problem 11, we have

$$\rho(A \circ B) \leq \rho(A \otimes B) = \rho(A)\rho(B).$$

The example below shows that $\rho(AB) \le \rho(A)\rho(B)$ is not true:

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

The following result ensures that $\rho(AB)$ dominates $\rho(A \circ B)$.

Theorem 5.23 *Let A and B be $n \times n$ nonnegative matrices. Then*

$$\rho(A \circ B) \le \rho(AB).$$

Proof. The proof is based on two facts: If $0 \le X \le Y$ then $\rho(X) \le \rho(Y)$ (Theorem 5.22); and if P is a principal submatrix of a nonnegative square matrix Q then $\rho(P) \le \rho(Q)$ (Problem 11).

We first show that

$$(A \circ B)(B \circ A) \le (AB) \circ (BA). \tag{5.20}$$

Computing the (i, j)-entry of the right-hand side of (5.20), we have

$$\Big(\sum_s a_{is} b_{sj}\Big)\Big(\sum_t b_{it} a_{tj}\Big) = \sum_{s,\,t} a_{is} b_{it} a_{tj} b_{sj}.$$

Setting $s = t = p$ yields $\sum_p a_{ip} b_{ip} a_{pj} b_{pj}$, the left-hand side of (5.20).

Recalling that the Hadamard product is a principal submatrix of the Kronecker product (Theorem 4.7), we have

$$(\rho(A \circ B))^2 \le \rho(AB \circ BA) \le \rho(AB \otimes BA) = (\rho(AB))^2.$$

By taking the square roots, we obtain the desired inequality. ∎

The next result reveals lower and upper bounds for the spectral radius of a nonnegative matrix in terms of the entries, precisely, the row and column sums of the matrix.

Theorem 5.24 *Let A be an n-square nonnegative matrix. Then*

$$\min_{1 \le i \le n} \sum_{j=1}^n a_{ij} \le \rho(A) \le \max_{1 \le i \le n} \sum_{j=1}^n a_{ij}.$$

In other words, the spectral radius of a nonnegative square matrix is between the smallest row sum and the largest row sum.

Proof. We may assume $A \neq 0$. Denote the ith row sum by r_i; namely, $r_i = a_{i1} + a_{i2} + \cdots + a_{in}$. Let r be the smallest row sum of A; that is, $r = \min r_i$. We construct a new matrix B such that

$$0 \leq B \leq A \quad \text{and} \quad r = \rho(B).$$

If $r = 0$, set $B = 0$. Otherwise, let

$$b_{ij} = r(r_i^{-1} a_{ij}).$$

It is immediate that $0 \leq B \leq A$. The preceding theorem ensures $\rho(B) \leq \rho(A)$. To see $\rho(B) = r$, first observe that $Be_0 = re_0$, where $e_0 = (1, \ldots, 1)^T$. This says that r is an eigenvalue of B. So $\rho(B) \geq r$. However, considering the maximum row sum matrix norm $\| \cdot \|_\infty$ (Problem 8, Section 4.2), we have $\rho(B) \leq \|B\|_\infty = r$. Thus, $\rho(B) = r$. The upper bound is similarly shown. ∎

Similar results hold for the columns with A^T in place of A.

We now show a fundamental theorem on nonnegative matrices. We present the theorem for positive matrices. The results can be generalized and amplified to nonnegative matrices (due to Frobenius).

Theorem 5.25 (Perron) *Let A be an $n \times n$ positive matrix. Then*

1. $\rho(A) > 0$.
2. *$\rho(A)$ is an eigenvalue of A.*
3. *$Ax = \rho(A)x$ for some vector $x > 0$.*
4. *If $Au = \rho(A)u$ and $Av = \rho(A)v$, then $u = \alpha v$ for some $\alpha \in \mathbb{C}$.*
5. *If λ is an eigenvalue of A and $\lambda \neq \rho(A)$, then $|\lambda| < \rho(A)$.*

Proof. If $A > 0$, then each row sum (and column) sum is greater than zero. By the preceding theorem, (1) is true.

To show (2) and (3), let $Ax = \lambda x$, where $x \neq 0$ and $|\lambda| = \rho(A)$. We show that $A|x| = \rho(A)|x|$. Denote $\rho = \rho(A)$ for simplicity. Then

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

Note that $A|u| > 0$ for all $u \neq 0$ since every entry of A is positive. So $A|x| > 0$. Now set $y = A|x| - \rho|x|$. Then $y \geq 0$. If $y \neq 0$, then

$$0 < Ay = A(A|x|) - \rho A|x|$$

or

$$Az > \rho z, \quad \text{where } z = A|x| > 0.$$

If we write $z = A|x| = (z_1, \dots, z_n)$, then all z_i are positive and $Az > \rho z$ implies that $\sum_j a_{ij} z_j > \rho z_i$ for all i. Let Z be the diagonal matrix $\operatorname{diag}(z_1, \dots, z_n)$. Using Theorem 5.24, we have

$$\rho(A) = \rho(Z^{-1}AZ) \geq \min_i \frac{1}{z_i} \sum_j a_{ij} z_j > \rho = \rho(A),$$

a contradiction. So $y = 0$, i.e., $A|x| = \rho|x|$, and $|x| > 0$ as $A|x| > 0$.

We have shown that for any square positive matrix A,

$$Ax = \lambda x,\ x \neq 0,\ |\lambda| = \rho(A) \quad \Rightarrow \quad A|x| = \rho(A)|x|,\ |x| > 0. \tag{5.21}$$

To show (4), following the above argument, we have $A|x| = \rho|x|$ and $|x| > 0$ whenever $Ax = \rho x$ and $x \neq 0$. Thus for $t = 1, 2, \dots, n$,

$$\rho|x_t| = \sum_{j=1}^{n} a_{tj} |x_j|.$$

However, $\rho|x| = |\rho x| = |Ax|$, thus we have

$$\rho|x_t| = \Big| \sum_{j=1}^{n} a_{tj} x_j \Big| \quad \text{and} \quad \Big| \sum_{j=1}^{n} a_{tj} x_j \Big| = \sum_{j=1}^{n} a_{tj} |x_j|$$

which holds if and only if all $a_{t1}x_1, a_{t2}x_2, \dots, a_{tn}x_n$, thus $x_1, x_2, \dots,$ x_n (as $a_{t1}, a_{t2}, \dots, a_{tn}$ are positive), have the same argument (Problem 6), namely, there exists $\theta \in \mathbb{R}$ such that for all $j = 1, 2, \dots, n$,

$$e^{\theta i} x_j > 0; \quad \text{that is,} \quad |x| = e^{\theta i} x > 0. \tag{5.22}$$

Now suppose u and v are two eigenvectors belonging to the eigenvalue ρ. By the above discussion, there exist real numbers θ_1 and θ_2 such that $|u| = e^{\theta_1 i} u > 0$ and $|v| = e^{\theta_2 i} v > 0$. Set $\beta = \min_t |u_t|/|v_t|$ and define $w = |u| - \beta|v|$. Then w is a nonnegative vector having at least one component 0. On the other hand,

$$Aw = A(|u| - \beta|v|) = A|u| - \beta A|v| = \rho|u| - \beta\rho|v| = \rho w.$$

If $w \neq 0$, then $w > 0$ by (5.21), a contraction. Thus $w = 0$ and $|u| = \beta|v|$ which implies $u = \alpha v$, where $\alpha = \beta e^{(\theta_2 - \theta_1)i}$.

For (5), if $|\lambda| = \rho$ and $Ax = \lambda x$, $x \neq 0$, then $e^{\theta i}x > 0$ for some $\theta \in \mathbb{R}$ by (5.22). Set $y = e^{\theta i}x$. Then $Ay = \lambda y$, which is positive. So $\lambda > 0$ and $\lambda = \rho$. In other words, if $\lambda \neq \rho$, then $|\lambda| < \rho$. ∎

The Perron theorem simply states that for any square positive matrix A, the spectral radius $\rho(A)$ is an eigenvalue of A, known as the *Perron root*. The Perron eigenvalue of a positive matrix is the only eigenvalue that attains the spectral radius. Moreover, the positive eigenvectors, known as *Perron vectors*, of the Perron eigenvalue are unique up to magnitude. These statements hold for irreducible nonnegative matrices. However, it is possible for a nonnegative A to have several eigenvalues that have the maximum modulus $\rho(A)$.

Problems

1. Does there exist a positive matrix that is similar to $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$?

2. Compute explicitly the eigenvalues of the matrix

$$A = \begin{pmatrix} 1-\alpha & \alpha \\ 1-\alpha & \alpha \end{pmatrix}, \quad 0 < \alpha < 1.$$

   Show that the spectral radius $\rho(A)$ can be strictly less than the numerical radius $w(A) = \max_{\|x\|=1} x^*Ax$ for some α. What if $\alpha = \frac{1}{2}$?

3. Can the inverse of a positive matrix be also positive?

4. Let A, B, C, D be square matrices of the same size. Show that

$$0 \leq B \leq A, \;\; 0 \leq D \leq C \quad \Rightarrow \quad 0 \leq BD \leq AC.$$

5. Let A be an n-square positive matrix. Show that there exists a unique vector $x > 0$ such that $Ax = \rho(A)x$ and $x_1 + x_2 + \cdots + x_n = 1$.

6. Let $p_1, p_2, \ldots, p_n$ be positive numbers and let $c_1, c_2, \ldots, c_n \in \mathbb{C}$. If

$$|p_1c_1 + p_2c_2 + \cdots + p_nc_n| = p_1|c_1| + p_2|c_2| + \cdots + p_n|c_n|,$$

   show that $e^{\theta i}c_k \geq 0$ for some $\theta \in \mathbb{R}$ and all $k = 1, 2, \ldots, n$.

7. Let $A \geq 0$ be a square matrix. If $\alpha x \leq Ax \leq \beta x$ for some scalars α, β, and nonnegative vector $x \neq 0$, show that $\alpha \leq \rho(A) \leq \beta$.

8. Let $A \geq 0$ be a square matrix. If $Au = \lambda u$ for some positive vector u and a scalar λ, show that $\lambda = \rho(A)$. Is it true that $Av \leq \rho(A)v$ for all positive vectors v?

9. Show that $\rho(A) \leq \rho(|A|)$ for any complex square matrix A.

10. Compute $\rho(AB)$, $\rho(A)\rho(B)$, and $\rho(A \circ B)$ for each pair of the following matrices and conclude that any two of these quantities can be the same whereas the third one may be different:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}; \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}; \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

11. Let $A \geq 0$ be a square matrix and B be a (proper) principal submatrix of A. Show that $\rho(B) \leq \rho(A)$. If $A > 0$, show that $\rho(B) < \rho(A)$.

12. Let A and B be $n \times n$ positive matrices. Show that $\rho(A \circ B) < \rho(AB)$.

13. Let A and B be $n \times n$ nonnegative matrices. Show that $\|A \circ B\|_{\mathrm{op}} \leq \rho(A^T B)$, where $\|A \circ B\|_{\mathrm{op}}$ is the operator (spectral) norm of $A \circ B$. Show by example that $\|A \circ B\|_{\mathrm{op}} \leq \rho(AB)$ does not hold in general.

14. Let $A > 0$ and let $M = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}$. What is $\rho(M)$? How many eigenvalues of M are there that have the maximum modulus $\rho(M)$?

15. Show that every square positive matrix is similar to a positive matrix all of whose row sums are constant.

16. Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. Show that $Au \neq Bv$ for any positive vectors u and v. Use this to show that no two multiplicative elements of A and B written in different ways are the same. For example, $A^2B^3ABA^5B \neq B^2A^3BA^3B^2$.

17. An *M-matrix* A is a matrix that can be written as

$$A = \alpha I - P, \quad \text{where } P \geq 0, \ \alpha \geq \rho(P).$$

Let A be an M-matrix. Show that

(a) All principal submatrices of A are M-matrices.

(b) All real eigenvalues of A are nonnegative.

(c) The determinant of A is nonnegative.

(d) If A is nonsingular, then $A^{-1} \geq 0$.

CHAPTER 6

Unitary Matrices and Contractions

Introduction: This chapter studies unitary matrices and contractions. Section 6.1 gives basic properties of unitary matrices, Section 6.2 discusses the structure of real orthogonal matrices under similarity, and Section 6.3 develops metric spaces and the fixed-point theorem of strict contractions. Section 6.4 deals with the connections of contractions with unitary matrices, Section 6.5 concerns the unitary similarity of real matrices, and Section 6.6 presents a trace inequality for unitary matrices, relating the average of the eigenvalues of each of two unitary matrices to that of their product.

6.1 Properties of Unitary Matrices

A *unitary matrix* is a square complex matrix satisfying

$$U^*U = UU^* = I.$$

Notice that $U^* = U^{-1}$ and $|\det U| = 1$ for any unitary matrix U. A complex (real) matrix A is called *complex* (*real*) *orthogonal* if

$$A^T A = AA^T = I.$$

Unitary matrices and complex orthogonal matrices are different in general, but real unitary and real orthogonal matrices are the same.

F. Zhang, *Matrix Theory: Basic Results and Techniques*, Universitext,
DOI 10.1007/978-1-4614-1099-7_6,

Theorem 6.1 *Let $U \in \mathbb{M}_n$ be a unitary matrix. Then*

1. $\|Ux\| = \|x\|$ *for every* $x \in \mathbb{C}^n$.
2. $|\lambda| = 1$ *for every eigenvalue* λ *of* U.
3. $U = V \operatorname{diag}(\lambda_1, \ldots, \lambda_n)V^*$, *where* V *is unitary and each* $|\lambda_i| = 1$.
4. *The column vectors of* U *form an orthonormal basis for* $\mathbb{C}^n$.

Proof. (1) is obtained by rewriting the norm as an inner product:

$$\|Ux\| = \sqrt{(Ux, Ux)} = \sqrt{(x, U^*Ux)} = \sqrt{(x, x)} = \|x\|.$$

To show (2), let x be a unit eigenvector of U corresponding to eigenvalue λ. Then, by using (1),

$$|\lambda| = |\lambda|\|x\| = \|\lambda x\| = \|Ux\| = \|x\| = 1.$$

(3) is by the spectral decomposition theorem (Theorem 3.4).

For (4), suppose that u_i is the ith column of U, $i = 1, \ldots, n$. Then the matrix identity $U^*U = I$ is equivalent to

$$\begin{pmatrix} u_1^* \\ \vdots \\ u_n^* \end{pmatrix} (u_1, \ldots, u_n) = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}.$$

This says $(u_i, u_j) = u_j^* u_i = 1$ if $i = j$, and 0 otherwise. ∎

An interesting observation follows. Note that for any unitary U,

$$\operatorname{adj}(U) = (\det U)U^{-1} = (\det U)U^*.$$

If we partition U as

$$U = \begin{pmatrix} u & \alpha \\ \beta & U_1 \end{pmatrix}, \quad u \in \mathbb{C},$$

then, by comparing the $(1,1)$-entries in $\operatorname{adj}(U) = (\det U)U^*$,

$$\det U_1 = \det U \; \bar{u},$$

which is also a consequence of Theorem 2.3.

As we saw in Theorem 6.1, the eigenvalues of a unitary matrix are necessarily equal to 1 in absolute value. The converse is not true in general. We have, however, the following result.

Theorem 6.2 *Let $A \in \mathbb{M}_n$ have all the eigenvalues equal to 1 in absolute value. Then A is unitary if, for all $x \in \mathbb{C}^n$,*

$$\|Ax\| \leq \|x\|.$$

Proof 1. The given inequality is equivalent to

$$\|Ax\| \leq 1, \quad \text{for all unit } x \in \mathbb{C}^n.$$

This specifies that $\sigma_{\max}(A) \leq 1$ (Problem 7, Section 4.1).

On the other hand, the identity $|\det A|^2 = \det(A^*A)$ implies that the product of eigenvalues in absolute value equals the product of singular values. If A has only eigenvalues 1 in absolute value, then the smallest singular value of A has to be 1; thus, all the singular values of A are equal to 1. Therefore, $A^*A = I$, and A is unitary.

Proof 2. Let $A = U^*DU$ be a Schur decomposition of A, where U is a unitary matrix, and D is an upper-triangular matrix:

$$D = \begin{pmatrix} \lambda_1 & t_{12} & \dots & t_{1n} \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{n-1,n} \\ 0 & \dots & 0 & \lambda_n \end{pmatrix},$$

where $|\lambda_i| = 1$, $i = 1, 2, \dots, n$, and t_{ij} are complex numbers.

Take $x = U^*e_n = U^*(0, \dots, 0, 1)^T$. Then $\|x\| = 1$. We have

$$\|Ax\| = \|De_n\| = (|t_{1n}|^2 + \dots + |t_{n-1,n}|^2 + |\lambda_n|^2)^{1/2}.$$

The conditions $\|Ax\| \leq 1$ for unit x and $|\lambda_n| = 1$ force each $t_{in} = 0$ for $i = 1, 2, \dots, n-1$. By induction, one sees that D is a diagonal matrix with the λ_i on the diagonal. Thus, A is unitary, for

$$A^*A = U^*D^*UU^*DU = U^*D^*DU = U^*U = I. \quad \blacksquare$$

We now show a result on the singular values of the principal submatrices of a unitary matrix.

Theorem 6.3 *Let U be a unitary matrix partitioned as*

$$U = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where A is $m \times m$ and D is $n \times n$. If $m = n$, then A and D have the same singular values. If $m < n$ and if the singular values of A are $\sigma_1, \ldots, \sigma_m$, then the singular values of D are $\sigma_1, \ldots, \sigma_m, \overbrace{1, \ldots, 1}^{n-m}$.

Proof. Since U is unitary, the identities $U^*U = UU^* = I$ imply

$$A^*A + C^*C = I_m, \quad CC^* + DD^* = I_n.$$

It follows that

$$A^*A = I_m - C^*C, \quad DD^* = I_n - CC^*.$$

Note that CC^* and C^*C have the same nonzero eigenvalues. Hence, $I_n - CC^*$ and $I_m - C^*C$ have the same eigenvalues except $n - m$ 1s; that is, A^*A and DD^* have the same eigenvalues except $n - m$ 1s. Therefore, if $m = n$, then A and D have the same singular values, and if $m < n$ and A has singular values $\sigma_1, \ldots, \sigma_m$, then the singular values of D are $\sigma_1, \ldots, \sigma_m$, plus $n - m$ 1s. ∎

An interesting result on the unitary matrix U in Theorem 6.3 is

$$|\det A| = |\det D|.$$

In other words, the complementary principal submatrices of a unitary matrix always have the same determinant in absolute value.

Problems

1. Which of the items in Theorem 6.1 implies that U is a unitary matrix?
2. Show that for any $\theta_1, \ldots, \theta_n \in \mathbb{R}$, $\operatorname{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n})$ is unitary.
3. What are the singular values of a unitary matrix?
4. Find an $m \times n$ matrix V, $m \neq n$, such that $V^*V = I_n$.
5. Let $A = \left(\begin{smallmatrix} 0 & x \\ 0 & 0 \end{smallmatrix}\right)$, where $x \in \mathbb{C}$. What are the eigenvalues and singular values of A in terms of x?

6. Show that for any $\theta \in \mathbb{R}$, the following two matrices are similar:

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

[Note: They are both unitary; the first is complex; the second is real.]

7. Show that any 2×2 unitary matrix with determinant equal to 1 is similar to a real orthogonal matrix.

8. Let A and C be m- and n-square matrices, respectively, and let

$$M = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}.$$

Show that M is unitary if and only if $B = 0$ and A and C are unitary.

9. Show that the 2×2 block matrix below is real orthogonal:

$$\begin{pmatrix} \sqrt{\lambda}\, I & -\sqrt{1-\lambda}\, I \\ \sqrt{1-\lambda}\, I & \sqrt{\lambda} I \end{pmatrix}, \quad \lambda \in [0, 1].$$

10. If A is a unitary matrix with all eigenvalues real, show that

$$A^2 = I \quad \text{and} \quad A^* = A.$$

11. If A is similar to a unitary matrix, show that A^* is similar to A^{-1}.

12. Let $A \in \mathbb{M}_n$ be Hermitian. Show that $(A - iI)^{-1}(A + iI)$ is unitary.

13. Let A be an $n \times n$ unitary matrix. If $A - I$ is nonsingular, show that $i(A - I)^{-1}(A + I)$ is Hermitian.

14. Let a and b be real numbers such that $a^2 - b^2 = 1$, $ab \neq 0$, and let

$$K = \begin{pmatrix} a & bi \\ -bi & a \end{pmatrix}.$$

   (a) Show that K is complex orthogonal but not unitary; that is,

$$K^T K = I \quad \text{but} \quad K^* K \neq I.$$

   (b) Let $a = \sqrt{2}$ and $b = 1$. Find the eigenvalues of K.

   (c) Let $a = \frac{1}{2}(e + e^{-1})$ and $b = \frac{1}{2}(e - e^{-1})$, where $e = 2.718\ldots$. Show that the eigenvalues of K are e and e^{-1}.

   (d) Let $a = \frac{1}{2}(e^t + e^{-t})$ and $b = \frac{1}{2}(e^t - e^{-t})$, $t \in \mathbb{R}$. What are the eigenvalues of K? Is the trace of K bounded?

15. Let A be an n-square complex matrix. Show that $A = 0$ if A satisfies

$$|(Ay, y)| \leq \|y\|, \quad \text{for all } y \in \mathbb{C}^n,$$

or

$$|(Ay, y)| \leq \|Ay\|, \quad \text{for all } y \in \mathbb{C}^n.$$

[Hint: If $A \neq 0$, then $(Ay_0, y_0) \neq 0$ for some $y_0 \in \mathbb{C}^n$.]

16. Let $A \in \mathbb{M}_n$ and $\alpha \in (0, 1)$. What can be said about A if

$$|(Ay, y)| \leq (y, y)^\alpha, \quad \text{for all } y \in \mathbb{C}^n?$$

17. Let $A \in \mathbb{M}_n$ have all eigenvalues equal to 1 in absolute value. Show that A is unitary if A satisfies

$$|(Ay, y)| \leq \|Ay\|^2, \quad \text{for all } y \in \mathbb{C}^n,$$

or

$$|(Ay, y)| \leq \|y\|^2, \quad \text{for all } y \in \mathbb{C}^n.$$

18. Let A be an $n \times n$ complex matrix having the largest and the smallest singular values $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$, respectively. Show that

$$(Ay, Ay) \leq (y, y), \quad \text{for all } y \in \mathbb{C}^n, \quad \Rightarrow \quad \sigma_{\max}(A) \leq 1$$

and

$$(y, y) \leq (Ay, Ay), \quad \text{for all } y \in \mathbb{C}^n, \quad \Rightarrow \quad \sigma_{\min}(A) \geq 1.$$

19. Let $A \in \mathbb{M}_n$ have all eigenvalues equal to 1 in absolute value. Show that A is unitary if A satisfies, for some real $\alpha \neq \frac{1}{2}$,

$$|(Ay, y)| \leq (Ay, Ay)^\alpha, \ \text{ for all unit } y \in \mathbb{C}^n.$$

[Hint: Assume that A is upper-triangular with $a_{11} = 1$, $a_{1i} > 0$ for some $i > 1$. Take $y = (\cos t, 0, \ldots, 0, \sin t, 0, \ldots, 0)$ and consider the behavior of the function $f(t) = (Ay, Ay)^\alpha - |(Ay, y)|$ near the origin.]

6.2 Real Orthogonal Matrices

This section is devoted to real orthogonal matrices, the real matrices A satisfying $AA^T = A^TA = I$. We discuss the structure of real orthogonal matrices under similarity and show that real orthogonal matrices with a commutativity condition are necessarily involutions.

We begin with 2×2 real orthogonal matrices A:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad a,\ b,\ c,\ d \in \mathbb{R}.$$

The identities $AA^T = A^TA = I$ imply several equations in a, b, c, and d, one of which is $a^2 + b^2 = 1$. Since a is real between -1 and 1, one may set $a = \cos\theta$ for some $\theta \in \mathbb{R}$, and get b, c, and d in terms of θ. Thus, there are only two types of 2×2 real orthogonal matrices:

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}, \quad \theta \in \mathbb{R}, \tag{6.1}$$

where the first type is called *rotation* and the second *reflection*.

We show that a real orthogonal matrix is similar to a direct sum of real orthogonal matrices of order 1 or 2.

Theorem 6.4 *Every real orthogonal matrix is real orthogonally similar to a direct sum of real orthogonal matrices of order at most 2.*

Proof. Let A be an $n \times n$ real orthogonal matrix. We apply induction on n. If $n = 1$ or 2, there is nothing to prove.

Suppose $n > 2$. If A has a real eigenvalue λ with a real unit eigenvector x, then

$$Ax = \lambda x, \ \ x \neq 0 \ \ \Rightarrow \ \ x^TA^TAx = \lambda^2 x^Tx.$$

Thus, $\lambda = \pm 1$, say, 1. Extend the real unit eigenvector x to a real orthogonal matrix P. Then P^TAP has $(1,1)$-entry 1 and 0 elsewhere in the first column. Write in symbols

$$P^TAP = \begin{pmatrix} 1 & u \\ 0 & A_1 \end{pmatrix}.$$

The orthogonality of A implies $u = 0$. Notice that A_1 is also real orthogonal. The conclusion then follows from an induction on A_1.

Assume that A has no real eigenvalues. Then for any nonzero $x \in \mathbb{R}^n$, vectors x and Ax cannot be linearly dependent. Recall the angle $\angle_{x,\, y}$ between two vectors x and y and note that $\angle_{x,\, y} = \angle_{Ax, Ay}$ due to the orthogonality of A. Define $f(x)$ to be the angle function

$$f(x) = \angle_{x, Ax} = \cos^{-1} \frac{(x, Ax)}{\|x\| \|Ax\|}.$$

Then $f(x)$ is continuous on the compact set $S = \{x \in \mathbb{R}^n : \ \|x\| = 1\}$.

Let $\theta_0 = \angle_{x_0, Ax_0}$ be the minimum of $f(x)$ on S. Let y_0 be the unit vector in $\operatorname{Span}\{x_0, Ax_0\}$ such that $\angle_{x_0,\, y_0} = \angle_{y_0, Ax_0}$. Then by Theorem 1.10, we have

$$\theta_0 \leq \angle_{y_0, Ay_0} \leq \angle_{y_0, Ax_0} + \angle_{Ax_0, Ay_0} = \frac{\theta_0}{2} + \frac{\theta_0}{2} = \theta_0$$

and $Ax_0 \in \operatorname{Span}\{y_0, Ay_0\}$. Thus, Ay_0 has to belong to $\operatorname{Span}\{x_0, y_0\}$. It follows, because Ax_0 is also in the subspace, that $\operatorname{Span}\{x_0, y_0\}$ is an invariant subspace under A. We thus write A, up to similarity by taking a suitable orthonormal basis (equivalently, via a real orthogonal matrix), as

$$A = \begin{pmatrix} T_0 & 0 \\ 0 & B \end{pmatrix},$$

where T_0 is a 2×2 matrix, and B is a matrix of order $n - 2$. Since A is orthogonal, so are T_0 and B. An application of the induction hypothesis to B completes the proof. ∎

Another way to attack the problem is to consider the eigenvectors of A. One may again focus on the nonreal eigenvalues. Since A is real, the characteristic polynomial of A has real coefficients, and the nonreal eigenvalues of A thus occur in conjugate pairs. Furthermore, their eigenvectors are in the forms $\alpha + \beta i$ and $\alpha - \beta i$, where α and β are real, $(\alpha, \beta) = 0$, and $A\alpha = a\alpha - b\beta$, $A\beta = b\alpha + a\beta$ for some real a and b with $a^2 + b^2 = 1$. Matrix A will have the desired form (via orthogonal similarity) by choosing a suitable real orthonormal basis.

Our next theorem shows that a matrix with a certain commuting property is necessarily an involution. For this purpose, we need a result that is of interest in its own right.

If two complex square matrices F and G of orders m and n, respectively, have no eigenvalues in common, then the matrix equation $FX - XG = 0$ has a unique solution $X = 0$.

To see this, rewrite the equation as $FX = XG$. Then for every positive integer k, $F^k X = XG^k$. It follows that

$$f(F)X = Xf(G)$$

for every polynomial f. In particular, we take f to be the characteristic polynomial $\det(\lambda I - F)$ of F; then $f(F) = 0$, and thus $Xf(G) = 0$. However, because F and G have no eigenvalues in common, $f(G)$ is nonsingular and hence $X = 0$.

Theorem 6.5 *Let A and U be real orthogonal matrices of the same size. If U has no repeated eigenvalues and if*

$$UA = AU^T,$$

then A is an involution, that is, $A^2 = I$.

Proof. By the previous theorem, let P be a real orthogonal matrix such that $P^{-1}UP$ is a direct sum of orthogonal matrices V_i of order 1 or 2. The identity $UA = AU^T$ results in

$$(P^{-1}UP)(P^{-1}AP) = (P^{-1}AP)(P^{-1}UP)^T.$$

Partition $P^{-1}AP$ conformally with $P^{-1}UP$ as $P^{-1}AP = (B_{ij})$, where the B_{ij} are matrices whose number of rows (or columns) is 1 or 2. Then $UA = AU^T$ gives

$$V_i B_{ij} = B_{ij} V_j^T, \quad i, j = 1, \ldots, k. \tag{6.2}$$

Since U, thus $P^{-1}UP$, has no repeated eigenvalues, we have

$$B_{ij} = 0, \quad i \neq j.$$

Hence $P^{-1}AP$ is a direct sum of matrices of order no more than 2:

$$P^{-1}AP = B_{11} \oplus \cdots \oplus B_{kk}.$$

The orthogonality of A, thus $P^{-1}AP$, implies that each B_{ii} is either an orthogonal matrix of order 1 or an orthogonal matrix of form (6.1). Obviously $B_{ii}^2 = I$ if B_{ii} is ± 1 or a reflection. Now suppose B_{ii} is a rotation; then V_i is not a rotation. Otherwise, V_i and B_{ii} are both rotations and hence commute (Problem 4), so that $V_i^2 B_{ii} = B_{ii}$ and $V_i^2 = I$. Using the rotation in (6.1), we have $V_i = \pm I$, contradicting the fact that V_i has two distinct eigenvalues. Thus, V_i is a reflection, hence orthogonally similar to $\operatorname{diag}(1, -1)$. It follows that B_{ii} is similar to $\operatorname{diag}(\pm 1, \pm 1)$ by (6.2). In either case $B_{ii}^2 = I$. Thus, $(P^{-1}AP)^2 = I$ and $A^2 = I$. ∎

Problems

1. Give a 2×2 matrix such that $A^2 = I$ but $A^* A \neq I$.

2. When is an upper-triangular matrix (complex or real) orthogonal?

3. If A is a 2×2 real matrix with a complex eigenvalue $\lambda = a + bi$, $a,\ b \in \mathbb{R}$, $b \neq 0$, show that A is similar to the real matrix

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

4. Verify that A and B commute; that is, $AB = BA$, where

$$A = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}, \quad B = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix}, \quad \alpha,\ \beta \in \mathbb{R}.$$

5. Show that a real matrix P is an orthogonal projection if and only if P is orthogonally similar to a matrix in the form $\operatorname{diag}(1, \ldots, 1, 0, \ldots, 0)$.

6. Show that, with a rotation (in the xy-plane) of angle θ written as

$$I_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

   (a) $(I_\theta)^T = I_{-\theta}$,

   (b) $I_\theta I_\phi = I_\phi I_\theta = I_{\theta+\phi}$,

   (c) $I_\theta^n = I_{n\theta}$, and

   (d) a reflection is expressed as $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) I_\theta$.

7. If A is a real orthogonal matrix with $\det A = -1$, show that A has an eigenvalue -1.

8. Let A be a 3×3 real orthogonal matrix with $\det A = 1$. Show that
$$(\operatorname{tr} A - 1)^2 + \sum_{i<j} (a_{ij} - a_{ji})^2 = 4.$$

9. Let A be an $n \times n$ real matrix. Denote $ss(A) = \sum_{i,j=1}^n a_{ij}^2$. Show that A is real orthogonal if and only if
$$ss(A^T X A) = ss(X), \quad \text{for all } n \times n \text{ real } X.$$

10. If $A \in \mathbb{M}_n$ is real symmetric and idempotent, show that $0 \le a_{ii} \le 1$ for each i and $|a_{ij}| \le \frac{1}{2}$ for all $i \ne j$. Moreover, if $a_{ii} = 0$ or 1, then $a_{ij} = a_{ji} = 0$ for all j with $j \ne i$.

11. Let A be an $n \times n$ real orthogonal matrix such that $\operatorname{rank}(A - I_n) = 1$. Show that A is real orthogonally similar to $\operatorname{diag}(-1, 1, \ldots, 1)$.

12. Let $A = (a_{ij})$ be an $n \times n$ real matrix and C_{ij} be the cofactor of a_{ij}, $i, j = 1, 2, \ldots, n$. Show that A is orthogonal if and only if $\det A = \pm 1$ and $a_{ij} = C_{ij}$ if $\det A = 1$, $a_{ij} = -C_{ij}$ if $\det A = -1$ for all i, j.

13. Let $A = (a_{ij}) \ne 0$ be an $n \times n$ real matrix and C_{ij} be the cofactor of a_{ij}, $i, j = 1, 2, \ldots, n$. If $n > 2$, show that A is orthogonal if $a_{ij} = C_{ij}$ for all i, j, or $a_{ij} = -C_{ij}$ for all i, j.

14. Let
$$A = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix}, \quad U = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}.$$

   (a) Show that $UA = AU^T$.

   (b) Find the eigenvalues of U.

   (c) Show that $A^2 \ne I$.

6.3 Metric Space and Contractions

A *metric space* consists of a set M and a mapping

$$d : M \times M \mapsto \mathbb{R},$$

called a *metric* of M, for which

1. $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$,
2. $d(x, y) = d(y, x)$ for all x and y in M, and
3. $d(x, y) \leq d(x, z) + d(z, y)$ for all x, y, and z in M.

Consider a sequence of points $\{x_i\}$ in a metric space M. If for every $\epsilon > 0$ there exists a positive integer N such that $d(x_i, x_j) < \epsilon$ for all $i,\ j > N$, then the sequence is called a *Cauchy sequence*. A sequence $\{x_i\}$ *converges* to a point x if for every $\epsilon > 0$ there exists a positive integer N such that $d(x, x_i) < \epsilon$ for all $i > N$. A metric space M is said to be *complete* if every Cauchy sequence converges to a point of M. For instance, $\{c^n\}$, $0 < c < 1$, is a Cauchy sequence of the complete metric space $\mathbb{R}$ with metric $d(x, y) = |x - y|$.

$\mathbb{C}^n$ is a metric space with metric

$$d(x, y) = \|x - y\|, \quad x,\ y \in \mathbb{C}^n, \tag{6.3}$$

defined by the vector norm

$$\|x\| = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}, \quad x \in \mathbb{C}^n.$$

Let $f : M \mapsto M$ be a mapping of a metric space M with metric d into itself. We call f a *contraction* if there exists a constant c with $0 < c \leq 1$ such that

$$d(f(x), f(y)) \leq c d(x, y), \quad \text{for all } x,\ y \in M. \tag{6.4}$$

If $0 < c < 1$, we say that f is a *strict contraction*.

For the metric space $\mathbb{R}$ with the usual metric

$$d(x, y) = |x - y|, \quad x,\ y \in \mathbb{R},$$

the mapping $x \mapsto \sin \frac{x}{2}$ is a strict contraction, since by the sum-to-product trigonometric identity (or by using the mean value theorem)

$$\sin \frac{x}{2} - \sin \frac{y}{2} = 2 \cos \frac{x+y}{4} \sin \frac{x-y}{4},$$

together with the inequality $|\sin x| \le |x|$, we have for all x, y in $\mathbb{R}$

$$\left| \sin \frac{x}{2} - \sin \frac{y}{2} \right| \le \frac{1}{2} |x - y|.$$

The mapping $x \mapsto \sin x$ is a contraction, but not strict, since

$$\lim_{x \to 0} \frac{\sin x}{x} = 1.$$

A point x in a metric space M is referred to as a *fixed point* of a mapping f if $f(x) = x$. The following fixed-point theorem of a contraction has applications in many fields. For example, it gives a useful method for constructing solutions of differential equations.

Theorem 6.6 *Let $f : M \mapsto M$ be a strict contraction mapping of a complete metric space M into itself. Then f has one and only one fixed point. Moreover, for any point $x \in M$, the sequence*

$$x,\ f(x),\ f^2(x),\ f^3(x),\ \ldots$$

converges to the fixed point.

Proof. Let x be a point in M. Denote $d(x, f(x)) = \delta$. By (6.4)

$$d(f^n(x), f^{n+1}(x)) \le c^n \delta, \quad n \ge 1. \tag{6.5}$$

The series $\sum_{n=1}^{\infty} c^n$ converges to $\frac{c}{1-c}$ for every fixed c, $0 < c < 1$. Hence, the sequence $f^n(x), n = 1, 2, \ldots,$ is a Cauchy sequence, since

$$\begin{aligned} d(f^m(x), f^n(x)) &\le d(f^m(x), f^{m+1}(x)) + \cdots + d(f^{n-1}(x), f^n(x)) \\ &\le (c^m + \cdots + c^{n-1})\delta. \end{aligned}$$

Thus, the limit $\lim_{n\to\infty} f^n(x)$ exists in M, for M is complete. Let the limit be X. We show that X is the fixed point. Note that a contraction mapping is a continuous function by (6.4). Therefore,

$$f(X) = f(\lim_{n\to\infty} f^n(x)) = \lim_{n\to\infty} f^{n+1}(x) = X.$$

If $Y \in M$ is also a fixed point of f, then

$$d(X, Y) = d(f(X), f(Y)) \le cd(X, Y).$$

It follows that $d(X, Y) = 0$; that is, $X = Y$ if $0 < c < 1$. ∎

Let A be an $m \times n$ complex matrix, and consider A as a mapping from $\mathbb{C}^n$ into itself defined by the ordinary matrix-vector product; namely, Ax, where $x \in \mathbb{C}^n$. Then inequality (6.4) is rewritten as

$$\|Ax - Ay\| \le c\|x - y\|.$$

We show that A is a contraction if and only if $\sigma_{\max}(A)$, the largest singular value of A, does not exceed 1.

Theorem 6.7 *Matrix A is a contraction if and only if $\sigma_{\max}(A) \le 1$.*

Proof. Let A be $m \times n$. For any x, $y \in \mathbb{C}^n$, we have (Section 4.1)

$$\|Ax - Ay\| = \|A(x - y)\| \le \sigma_{\max}(A)\|x - y\|.$$

It follows that A is a contraction if $\sigma_{\max}(A) \le 1$. Conversely, suppose that A is a contraction; then for some c, $0 < c \le 1$, and all x, $y \in \mathbb{C}^n$,

$$\|Ax - Ay\| \le c\|x - y\|.$$

In particular, $\|Ax\| \le c\|x\|$ for $x \in \mathbb{C}^n$. Thus, $\sigma_{\max}(A) \le c \le 1$. ∎

Note that unitary matrices are contractions, but not strict. One can also prove that a matrix A is a contraction if and only if

$$A^*A \le I, \quad AA^* \le I, \quad \text{or} \quad \begin{pmatrix} I & A \\ A^* & I \end{pmatrix} \ge 0.$$

Here $X \ge Y$, or $Y \le X$, means that $X - Y$ is positive semidefinite.

We conclude this section by presenting a result on partitioned positive semidefinite matrices, from which a variety of matrix inequalities can be derived.

Let A be a positive semidefinite matrix. Recall that $A^{1/2}$ is the square root of A; that is, $A^{1/2} \ge 0$ and $(A^{1/2})^2 = A$ (Section 3.2).

Theorem 6.8 *Let L and M be positive semidefinite matrices. Then*

$$\begin{pmatrix} L & X \\ X^* & M \end{pmatrix} \geq 0 \quad \Leftrightarrow \quad X = L^{1/2} C M^{1/2} \text{ for some contraction } C.$$

Proof. Sufficiency: If $X = L^{1/2}CM^{1/2}$, then we write

$$\begin{pmatrix} L & X \\ X^* & M \end{pmatrix} = \begin{pmatrix} L^{1/2} & 0 \\ 0 & M^{1/2} \end{pmatrix} \begin{pmatrix} I & C \\ C^* & I \end{pmatrix} \begin{pmatrix} L^{1/2} & 0 \\ 0 & M^{1/2} \end{pmatrix}.$$

For the positive semidefiniteness, it suffices to note that (Problem 9)

$$\sigma_{\max}(C) \leq 1 \quad \Rightarrow \quad \begin{pmatrix} I & C \\ C^* & I \end{pmatrix} \geq 0.$$

For the other direction, assume that L and M are nonsingular and let $C = L^{-1/2}XM^{-1/2}$. Here the exponent $-1/2$ means the inverse of the square root. Then $X = L^{1/2}CM^{1/2}$ has the desired form. We need to show that C is a contraction. First notice that

$$C^*C = M^{-1/2}X^*L^{-1}XM^{-1/2}.$$

Notice also that, since the partitioned matrix is positive semidefinite,

$$P^* \begin{pmatrix} L & X \\ X^* & M \end{pmatrix} P = \begin{pmatrix} L & 0 \\ 0 & M - X^*L^{-1}X \end{pmatrix} \geq 0,$$

where

$$P = \begin{pmatrix} I & -L^{-1}X \\ 0 & I \end{pmatrix}.$$

Thus, $M - X^*L^{-1}X \geq 0$ (Problem 9, Section 3.2). Therefore,

$$M^{-1/2}(M - X^*L^{-1}X)M^{-1/2} = I - M^{-1/2}X^*L^{-1}XM^{-1/2} \geq 0.$$

That is, $I - C^*C \geq 0$, and thus C is a contraction. The singular case of L and M follows from a continuity argument (Problem 17). ∎

We end this section by noting that targeting the submatrices X and X^* in the upper-right and lower-left corners in the given partitioned matrix by row and column elementary operations for block matrices is a basic technique in matrix theory. It is used repeatedly in later chapters of this book.

Problems

1. What are the differences between vector space, inner product space, normed space, and metric space?

2. Following the proof of Theorem 6.6, show that

$$d(x, X) \le \frac{\delta}{1-c}.$$

3. Show that a contraction is a continuous function.

4. If f is a strict contraction of a complete metric space, show that

$$d(f^n(x), f^n(y)) \to 0, \quad \text{as } n \to \infty,$$

for any fixed x and y in the space.

5. Show that the product of contractions is again a contraction.

6. Is the mapping $x \mapsto \sin(2x)$ a contraction on $\mathbb{R}$? How about the mappings $x \mapsto 2\sin x$, $x \mapsto \frac{1}{2}\sin x$, and $x \mapsto 2\sin\frac{x}{2}$?

7. Construct an example of a map f for a metric space such that $d(f(x), f(y)) < d(x, y)$ for all $x \neq y$, but f has no fixed point.

8. If $A \in \mathbb{M}_n$ is a contraction with eigenvalues $\lambda(A)$, show that

$$|\det A| \le 1, \quad |\lambda(A)| \le 1, \quad |x^*Ax| \le 1 \text{ for unit } x \in \mathbb{C}^n.$$

9. Show that an $m \times n$ matrix A is a contraction if and only if

   (a) $A^*A \le I_n$.

   (b) $AA^* \le I_m$.

   (c) $\begin{pmatrix} I & A \\ A^* & I \end{pmatrix} \ge 0$.

   (d) $x^*(A^*A)x \le 1$ for every unit $x \in \mathbb{C}^n$.

   (e) $\|Ax\| \le \|x\|$ for every $x \in \mathbb{C}^n$.

10. Let A be an $n \times n$ matrix and B be an $m \times n$ matrix. Show that

$$\begin{pmatrix} A & B^* \\ B & I \end{pmatrix} \ge 0 \quad \Leftrightarrow \quad A \ge B^*B.$$

11. Let $A \in \mathbb{M}_n$. If $\sigma_{\max}(A) < 1$, show that $I_n - A$ is invertible.

12. Let A and B be n-square complex matrices. Show that

$$A^*A \leq B^*B$$

if and only if $A = CB$ for some contraction matrix C.

13. Consider the complete metric space $\mathbb{R}^2$ and let

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}, \quad \lambda \in (0, 1).$$

Discuss the effect of an application of A to a nonzero vector $v \in \mathbb{R}^2$. Describe the geometric orbit of the iterates $A^n v$. What is the fixed point of A? What if the second λ in A is replaced with $\mu \in (0, 1)$?

14. Let $A \in \mathbb{M}_n$ be a projection matrix; that is, $A^2 = A$. Show that $\|Ax\| \leq \|x\|$ for all $x \in \mathbb{C}^n$ if and only if A is Hermitian.

15. Let A and B be n-square positive definite matrices. Find the conditions on the invertible n-square matrix X so that

$$\begin{pmatrix} A & X^* \\ X & B \end{pmatrix} \geq 0 \quad \text{and} \quad \begin{pmatrix} A^{-1} & X^{-1} \\ (X^*)^{-1} & B^{-1} \end{pmatrix} \geq 0.$$

16. Let A be a matrix. If there exists a Hermitian matrix X such that

$$\begin{pmatrix} I + X & A \\ A^* & I - X \end{pmatrix} \geq 0,$$

show that

$$|(Ay, y)| \leq (y, y), \quad \text{for all } y.$$

17. Prove Theorem 6.8 for the singular case.

18. Show that for any matrices X and Y of the same size,

$$X + Y = (I + XX^*)^{1/2} C (I + Y^*Y)^{1/2}$$

for some contraction C. Derive the matrix inequality

$$|\det(X + Y)|^2 \leq \det(I + XX^*) \det(I + Y^*Y).$$

6.4 Contractions and Unitary Matrices

The goal of this section is to present two theorems connecting contractions and unitary matrices. We focus on square matrices, for otherwise we can augment by zero entries to make the matrices square. We show that a matrix is a contraction if and only if it can be embedded in a unitary matrix, and if and only if it is a (finite) convex combination of unitary matrices.

Theorem 6.9 *A matrix A is a contraction if and only if*

$$U = \begin{pmatrix} A & X \\ Y & Z \end{pmatrix}$$

is unitary for some matrices X, Y, and Z of appropriate sizes.

Proof. The sufficiency is easy to see, since if U is unitary, then

$$U^*U = I \quad \Rightarrow \quad A^*A + Y^*Y = I \quad \Rightarrow \quad A^*A \leq I.$$

Thus, A is a contraction. For the necessity, we take

$$U = \begin{pmatrix} A & (I - AA^*)^{1/2} \\ (I - A^*A)^{1/2} & -A^* \end{pmatrix}$$

and show that U is a unitary matrix as follows.

Let $A = VDW$ be a singular value decomposition of A, where V and W are unitary, and D is a nonnegative diagonal matrix with diagonal entries (singular values of A) not exceeding 1. Then

$$(I - AA^*)^{1/2} = V(I - D^2)^{1/2}V^*, \quad (I - A^*A)^{1/2} = W^*(I - D^2)^{1/2}W.$$

In as much as D is diagonal, it is easy to see that

$$D(I - D^2)^{1/2} = (I - D^2)^{1/2}D.$$

Multiplying by V and W from the left and right gives

$$VD(I - D^2)^{1/2}W = V(I - D^2)^{1/2}DW$$

or equivalently

$$A(I - A^*A)^{1/2} = (I - AA^*)^{1/2}A.$$

With this, a simple computation results in $U^*U = I$. ∎

Recall from Problem 8 of Section 4.1 that for any A, $B \in \mathbb{M}_n$

$$\sigma_{\max}(A + B) \leq \sigma_{\max}(A) + \sigma_{\max}(B).$$

Thus, for unitary matrices U and V of the same size and $t \in (0, 1)$,

$$\sigma_{\max}(tU + (1 - t)V) \leq t\sigma_{\max}(U) + (1 - t)\sigma_{\max}(V) = 1.$$

In other words, the matrix $tU + (1 - t)V$, a convex combination of unitary matrices U and V, is a contraction.

Inductively, a convex combination of unitary matrices is a contraction (Problem 3). The converse is also true.

Theorem 6.10 *A matrix A is a contraction if and only if A is a finite convex combination of unitary matrices.*

Proof. As discussed earlier, a convex combination of unitary matrices is a contraction. Let A be a contraction. We show that A is a convex combination of unitary matrices. The proof goes as follows. A is a convex combination of matrices $\operatorname{diag}(1, \dots, 1, 0, \dots, 0)$; each matrix in such a form is a convex combination of diagonal (unitary) matrices with diagonal entries ± 1. We then reach the conclusion that A is a convex combination of unitary matrices.

Let A be of rank r and $A = UDV$ be a singular value decomposition of A, where U and V are unitary, $D = \operatorname{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0)$ with $1 \geq \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. We may assume $r > 0$.

If D is a convex combination of unitary matrices, say, W_i, then A is a convex combination of unitary matrices UW_iV. We may thus consider the diagonal matrix $A = \operatorname{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0)$. Write

$$\begin{aligned} A &= \operatorname{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0) \\ &= (1 - \sigma_1)0 + (\sigma_1 - \sigma_2)\operatorname{diag}(1, 0, \dots, 0) \\ &\quad + (\sigma_2 - \sigma_3)\operatorname{diag}(1, 1, 0, \dots, 0) + \cdots \\ &\quad + (\sigma_{r-1} - \sigma_r)\operatorname{diag}(\overbrace{1, \dots, 1}^{r-1}, 0, \dots, 0) \\ &\quad + \sigma_r \operatorname{diag}(\overbrace{1, \dots, 1}^{r}, 0, \dots, 0). \end{aligned}$$

That is, matrix A is a (finite) convex combination of matrices $E_i = \operatorname{diag}(1, \dots, 1, 0, \dots, 0)$ with i copies of 1, where $0 \le i \le r$. We now show that such a matrix is a convex combination of diagonal matrices with entries ± 1. Let

$$F_i = \operatorname{diag}(0, \dots, 0, \overbrace{-1, \dots, -1}^{n-i}).$$

Then

$$E_i = \frac{1}{2} I + \frac{1}{2}(E_i + F_i)$$

is a convex combination of unitary matrices I and $E_i + F_i$.

It follows that if $\sigma_1 \le 1$, then the matrix $\operatorname{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0)$, thus A, is a convex combination of diagonal matrices in the form $\operatorname{diag}(1, \dots, 1, 0, \dots, 0)$, which in turn is a convex combination of (diagonal) unitary matrices. The proof is complete by Problem 11. ∎

Problems

1. Let $t \in [0, 1]$. Write t as a convex combination of 1 and -1. Write matrix A as a convex combination of unitary matrices, where

$$A = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}.$$

2. Let λ and μ be positive numbers. Show that for any $t \in [0, 1]$,

$$\lambda\mu \le \big(t\lambda + (1-t)\mu\big)\big(t\mu + (1-t)\lambda\big).$$

3. Show by induction that a finite convex combination of unitary matrices of the same size is a contraction.

4. For any two matrices U and V of the same size, show that

$$U^*V + V^*U \le U^*U + V^*V.$$

   In particular, if U and V are unitary, then

$$U^*V + V^*U \le 2I.$$

   Also show that for any $t \in [0, 1]$ and unitary U and V,

$$\big(tU + (1-t)V\big)^*\big(tU + (1-t)V\big) \le I.$$

5. Prove or disprove that a convex combination, the sum, or the product of two unitary matrices is a unitary matrix.

6. If A is a contraction satisfying $A + A^* = 2I$, show that $A = I$.

7. Let A be a complex contraction matrix. Show that

$$\begin{pmatrix} A & (I - AA^*)^{1/2} \\ -(I - A^*A)^{1/2} & A^* \end{pmatrix}$$

and

$$\begin{pmatrix} (I - AA^*)^{1/2} & A \\ -A^* & (I - A^*A)^{1/2} \end{pmatrix}$$

are unitary matrices.

8. Let B_m be the $m \times m$ backward identity matrix. Show that

$$\frac{1}{\sqrt{2}} \begin{pmatrix} I_m & I_m \\ B_m & -B_m \end{pmatrix} \quad \text{and} \quad \frac{1}{\sqrt{2}} \begin{pmatrix} I_m & 0 & I_m \\ 0 & \sqrt{2} & 0 \\ B_m & 0 & -B_m \end{pmatrix}$$

are $2m$- and $(2m+1)$-square unitary matrices, respectively.

9. Let $\sigma_1 \geq \cdots \geq \sigma_r \geq 0$. Show that $\operatorname{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$ is a convex combination of matrices $\operatorname{diag}(\sigma_1, \ldots, \sigma_1, 0, \ldots, 0)$ with k copies of σ_1, $k = 1, 2, \ldots, r$.

10. Let σ be such that $0 < \sigma \leq 1$. Show that $\operatorname{diag}(\sigma, \ldots, \sigma, 0, \ldots, 0)$ is a convex combination of the following diagonal unitary matrices:

$$I, \quad G_i = \operatorname{diag}(\overbrace{-1, \ldots, -1}^{i}, 1, \ldots, 1), \quad -G_i, \quad -I.$$

[Hint: Consider the cases for $0 \leq \sigma < \frac{1}{2}$ and $\frac{1}{2} \leq \sigma \leq 1$.]

11. Let $P_1, \ldots, P_m$ be a set of matrices. If each P_i is a convex combination of matrices $Q_1, \ldots, Q_n$, show that a convex combination of $P_1, \ldots, P_m$ is also a convex combination of $Q_1, \ldots, Q_n$.

6.5 The Unitary Similarity of Real Matrices

We show in this section that if two real matrices are similar over the complex number field $\mathbb{C}$, then they are similar over the real $\mathbb{R}$. The statement also holds for unitary similarity. Precisely, if two real matrices are unitarily similar, then they are real orthogonally similar.

Theorem 6.11 *Let A and B be real square matrices of the same size. If P is a complex invertible matrix such that $P^{-1}AP = B$, then there exists a real invertible matrix Q such that $Q^{-1}AQ = B$.*

Proof. Write $P = P_1 + P_2 i$, where P_1 and P_2 are real square matrices. If $P_2 = 0$, we have nothing to show. Otherwise, by rewriting $P^{-1}AP = B$ as $AP = PB$, we have $AP_1 = P_1 B$ and $AP_2 = P_2 B$. It follows that for any real number t,

$$A(P_1 + tP_2) = (P_1 + tP_2)B.$$

Because $\det(P_1 + tP_2) = 0$ for a finite number of t, we can choose a real t so that the matrix $Q = P_1 + tP_2$ is invertible. Thus, A and B are similar via the real invertible matrix Q. ∎

For the unitary similarity, we begin with a result that is of interest in its own right.

Theorem 6.12 *Let U be a symmetric unitary matrix, that is, $U^T = U$ and $U^* = U^{-1}$. Then there exists a complex matrix S satisfying*

1. *$S^2 = U$.*
2. *S is unitary.*
3. *S is symmetric.*
4. *S commutes with every matrix that commutes with U.*

In other words, every symmetric unitary U has a symmetric unitary square root that commutes with any matrix commuting with U.

Proof. Since U is unitary, it is unitarily diagonalizable (Theorem 6.1). Let $U = VDV^*$, where V is unitary and $D = a_1 I_1 \oplus \cdots \oplus a_k I_k$ with all a_j distinct and I_i identity matrices of certain sizes. Because U is unitary and hence has eigenvalues of modulus 1, we write each $a_j = e^{i\theta_j}$ for some θ_j real.

Now let $S = V(b_1 I_1 \oplus \cdots \oplus b_k I_k)V^*$, where $b_j = e^{i\theta_j/2}$. Obviously S is a unitary matrix and $S^2 = U$.

If A is a matrix commuting with U, then V^*AV commutes with D. It follows that $V^*AV = A_1 \oplus \cdots \oplus A_k$, with each A_i having the same size as I_i (Problem 4). Thus, S commutes with A.

Since $U = U^T$, this implies that $V^T V$ commutes with D, so that $V^T V$ commutes with $b_1 I_1 \oplus \cdots \oplus b_k I_k$. Thus, S is symmetric. ∎

Theorem 6.13 *Let A and B be real square matrices of the same size. If $A = UBU^*$ for some unitary matrix U, then there exists a real orthogonal matrix Q such that $A = QBQ^T$.*

Proof. Since A and B are real, we have $UBU^* = A = \overline{A} = \overline{U}BU^T$. This gives $U^TUB = BU^TU$. Now that U^TU is symmetric unitary, by the preceding theorem it has a symmetric unitary square root, say S; that is, $U^TU = S^2$, which commutes with B.

Let $Q = US^{-1}$ or $U = QS$. Then Q is also unitary. Notice that

$$Q^TQ = (US^{-1})^T(US^{-1}) = S^{-1}U^TUS^{-1} = I.$$

Hence Q is orthogonal. Q is real, for $Q^T = Q^{-1} = Q^*$ yields $Q = \overline{Q}$. Putting it all together, S and B commute, S is unitary, and Q is real orthogonal. We thus have

$$\begin{aligned} A &= UBU^* = (US^{-1})(SB)U^* = Q(BS)U^* \\ &= QB(S^{-1})^*U^* = QBQ^* = QBQ^T. \ \blacksquare \end{aligned}$$

Problems

1. If A^2 is a unitary matrix, is A necessarily a unitary matrix?
2. If A is an invertible matrix with complex, real, rational, or integer entries, is the inverse of A also a matrix with complex, real, rational, or integer entries, respectively?

3. Let A be a normal matrix. Show that there exists a normal matrix B such that $B^2 = A$. Is such a B unique?

4. If matrix A commutes with $B = b_1 I_1 \oplus \cdots \oplus b_k I_k$, where the I_i are identity matrices and all b_i are distinct, show that A is of the form $A = A_1 \oplus \cdots \oplus A_k$, where each A_i has the same size as the corresponding I_i.

5. Show that A and B are similar via a real invertible matrix Q, where
$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$
Find Q. Are they real orthogonally (or unitarily) similar?

6. If two matrices A and B with rational entries are similar over the complex $\mathbb{C}$, are they similar over the real $\mathbb{R}$? The rational $\mathbb{Q}$?

7. Let Q be a real orthogonal matrix. If λ is an imaginary eigenvalue of Q and $u = x + yi$ is a corresponding eigenvector, where x and y are real, show that x and y are orthogonal and have the same length.

8. Let b and c be complex numbers such that $|b| \neq |c|$. Show that for any complex numbers a and d, matrix A cannot be normal, where
$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

9. Show that if A is a real symmetric matrix, then there exists a real orthogonal Q such that $Q^T AQ$ is real diagonal. Give an example of a real normal matrix that is unitarily similar to a (complex) diagonal matrix but is not real orthogonally similar to a diagonal matrix.

10. Let
$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & \pm\sqrt{2} \\ 0 & 0 & 2 \end{pmatrix}.$$
   (a) What are the eigenvalues and eigenvectors of A and B?
   (b) Why are A and B similar?
   (c) Show that 1 is a singular value of B but not of A.
   (d) Show that A cannot be unitarily similar to a direct sum of upper-triangular matrices of order 1 or 2.
   (e) Can A and B be unitarily similar?

6.6 A Trace Inequality of Unitary Matrices

The set of all complex matrices of the same size forms an inner product space over $\mathbb{C}$ with respect to the inner product defined as

$$(A, B)_{\mathbb{M}} = \operatorname{tr}(B^* A).$$

In what follows we consider the inner product space $\mathbb{M}_n$ over $\mathbb{C}$ and present a trace inequality for complex unitary matrices, relating the average of the eigenvalues of each of two unitary matrices to that of their product. For this purpose, we first show an inequality for an inner product space V, which is of interest in its own right.

Theorem 6.14 *Let u, v, and w be unit vectors in V over $\mathbb{C}$. Then*

$$\sqrt{1 - |(u, v)|^2} \leq \sqrt{1 - |(u, w)|^2} + \sqrt{1 - |(w, v)|^2}\,. \tag{6.6}$$

Equality holds if and only if w is a multiple of u or v.

Proof. To prove this, we first notice that any component of w that is orthogonal to the span of u and v plays no role in (6.6); namely, we really have a problem in which u and v are arbitrary unit vectors, w is in the span of u and v, and $(w, w) \leq 1$. The case $w = 0$ is trivial. If $w \neq 0$, scaling up w to have length 1 diminishes the right-hand side of (6.6), so we are done if we can prove inequality (6.6) for arbitrary unit vectors u, v, and w with w in the span of u and v. The case in which u and v are dependent is trivial. Suppose u and v are linearly independent, and let $\{u, z\}$ be an orthonormal basis of $\operatorname{Span}\{u, v\}$, so that $v = \mu u + \lambda z$ and $w = \alpha u + \beta z$ for some complex numbers μ, λ, α, and β. Then we have

$$|\lambda|^2 + |\mu|^2 = 1 \quad \text{and} \quad |\alpha|^2 + |\beta|^2 = 1.$$

Use these relations and the arithmetic-geometric mean inequality, together with $|c| \geq \operatorname{Re}(c)$ for any complex number c, to compute

$$\begin{aligned}
|\lambda\beta| &= \frac{1}{2}|\lambda\beta|(|\mu|^2+|\lambda|^2+|\alpha|^2+|\beta|^2) \\
&\geq |\lambda\beta|(|\lambda\beta|+|\alpha\mu|) \\
&= |\lambda\beta|^2+|\lambda\beta\alpha\mu| \\
&= |\lambda\beta|^2+|\lambda\bar{\beta}\alpha\bar{\mu}| \\
&\geq |\lambda\beta|^2+\mathrm{Re}(\lambda\bar{\beta}\alpha\bar{\mu}),
\end{aligned}$$

so that $-2|\lambda\beta| \leq -2|\lambda\beta|^2 - 2\,\mathrm{Re}(\lambda\bar{\beta}\alpha\bar{\mu})$. Thus, we have

$$\begin{aligned}
(|\lambda|-|\beta|)^2 &= |\lambda|^2-2|\lambda\beta|+|\beta|^2 \\
&\leq |\lambda|^2+|\beta|^2-2|\lambda\beta|^2-2\,\mathrm{Re}(\lambda\bar{\beta}\alpha\bar{\mu}) \\
&= |\lambda|^2+|\beta|^2(1-|\lambda|^2)-|\lambda\beta|^2-2\,\mathrm{Re}(\lambda\bar{\beta}\alpha\bar{\mu}) \\
&= (1-|\mu|^2)+|\beta|^2|\mu|^2-|\lambda\beta|^2-2\,\mathrm{Re}(\lambda\bar{\beta}\alpha\bar{\mu}) \\
&= 1-|\mu|^2(1-|\beta|^2)-|\lambda\beta|^2-2\,\mathrm{Re}(\lambda\bar{\beta}\alpha\bar{\mu}) \\
&= 1-|\mu\alpha|^2-|\lambda\beta|^2-2\,\mathrm{Re}(\lambda\bar{\beta}\alpha\bar{\mu}) \\
&= 1-|\alpha\bar{\mu}+\beta\bar{\lambda}|^2.
\end{aligned}$$

This gives

$$|\lambda|-|\beta| \leq \sqrt{1-|\alpha\bar{\mu}+\beta\bar{\lambda}|^2}\,,$$

or

$$|\lambda| \leq |\beta|+\sqrt{1-|\alpha\bar{\mu}+\beta\bar{\lambda}|^2}\,,$$

which is the same as

$$\sqrt{1-|\mu|^2} \leq \sqrt{1-|\alpha|^2}+\sqrt{1-|\alpha\bar{\mu}+\beta\bar{\lambda}|^2}\,.$$

Because $|\mu|^2=|(u,v)|^2$, $|\alpha|^2=|(u,w)|^2$, and

$$|\alpha\bar{\mu}+\beta\bar{\lambda}|^2 = |(\alpha u+\beta z, \mu u+\lambda z)|^2 = |(w,v)|^2,$$

the inequality (6.6) is proved.

Equality holds for the overall inequality if and only if equality holds at the two points in our derivation where we invoked the arithmetic-geometric mean inequality and $|c| \geq \mathrm{Re}(c)$. Thus, equality holds if and only if $|\lambda|=|\beta|$ and $|\alpha|=|\mu|$, as well as $\mathrm{Re}(\lambda\bar{\beta}\alpha\bar{\mu})=$

$|\lambda\bar{\beta}\alpha\bar{\mu}|$. The former is equivalent to having $\lambda = e^{i\theta}\beta$ and $\mu = e^{i\phi}\alpha$ for some real numbers θ and ϕ, while the latter is then equivalent to $\operatorname{Re}(|\alpha\beta|^2(e^{i(\theta-\phi)} - 1)) = 0$. Thus, $\alpha = 0$, $\beta = 0$, or $e^{i\theta} = e^{i\phi}$, so equality in (6.6) holds if and only if either w is a multiple of u ($\beta = 0$) or w is a multiple of v ($\alpha = 0$ or $e^{i\theta} = e^{i\phi}$). ∎

Now consider the vector space $\mathbb{M}_n$ of all $n \times n$ complex matrices with the inner product $(A, B)_{\mathbb{M}} = \operatorname{tr}(B^*A)$ for A and B in $\mathbb{M}_n$.

Let U and V be n-square unitary matrices. By putting

$$u = \frac{1}{\sqrt{n}}V, \quad v = \frac{1}{\sqrt{n}}U^*, \quad w = \frac{1}{\sqrt{n}}I$$

in (6.6), and writing $m(X) = \frac{1}{n}\operatorname{tr} X$ for the average of the eigenvalues of the matrix $X \in \mathbb{M}_n$, we have the following result.

Theorem 6.15 *For any unitary matrices U and V,*

$$\sqrt{1 - |m(UV)|^2} \le \sqrt{1 - |m(U)|^2} + \sqrt{1 - |m(V)|^2}$$

with equality if and only if U or V is a unitary scalar matrix.

Problems

1. Let U be an $m \times n$ matrix such that $U^*U = I_n$. Show that

$$\operatorname{tr}(UAU^*) = \operatorname{tr} A, \quad \text{for any } A \in \mathbb{M}_n.$$

How about

$$\operatorname{tr}(U^*AU) = \operatorname{tr} A, \quad \text{for any } A \in \mathbb{M}_n?$$

2. Show that for any square matrix A and positive integers p and q

$$|\operatorname{tr} A^{p+q}|^2 \le \operatorname{tr}\big((A^*)^p A^p\big) \operatorname{tr}\big((A^*)^q A^q\big)$$

and

$$\operatorname{tr}((A^*)^p A^p) \le (\operatorname{tr}(A^*A))^p, \quad \operatorname{tr}(A^*A)^p \le (\operatorname{tr}(A^*A))^p.$$

3. If U is an $n \times n$ nonscalar unitary matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$, show that the following strict inequality holds:

$$\left|\frac{\lambda_1 + \cdots + \lambda_n}{n}\right| < 1.$$

4. Let V be any square submatrix of a unitary matrix U. Show that $|\lambda(V)| \leq 1$ for any eigenvalue $\lambda(V)$ of V.

5. For unit vectors u and v in an inner product space V over $\mathbb{C}$, define $<_{u,v}= \cos^{-1}|(u,v)|$. Show that for any unit vector w in V,
$$\sin <_{u,v} \leq \sin <_{u,w} + \sin <_{w,v} .$$

6. Let $D = \operatorname{diag}(a_1, \ldots, a_n)$. Show that for any n-square unitary U,
$$\min_i |a_i| \leq |\lambda(DU)| \leq \max_i |a_i|,$$
where $\lambda(DU)$ is any eigenvalue of DU.

7. With $\|A\|_2 = \sqrt{(A,A)_{\mathbb{M}}} = \sqrt{\operatorname{tr}(A^*A)}$, show that for any $A \in \mathbb{M}_n$
$$\|A\|_2^2 = \left\|\frac{A+A^*}{2}\right\|_2^2 + \left\|\frac{A-A^*}{2}\right\|_2^2.$$

8. Let U be a unitary matrix. If λ and μ are two different eigenvalues of U, show that their eigenvectors u and v are orthogonal. Further show that $au + bv$ cannot be an eigenvector of U if $ab \neq 0$.

9. Let $\mathbb{P}^2$ be the collection of all the unit vectors in $\mathbb{C}^n$. Define
$$d(x,y) = \sqrt{1 - |(x,y)|^2}, \quad x,\ y \in \mathbb{P}^2.$$
Show that $d(x,y) = d(y,x)$ for all x and y in $\mathbb{P}^2$ and that $d(x,y) = 0$ if and only if $x = cy$ for some complex number c with $|c| = 1$.

10. For nonzero vectors u, $v \in \mathbb{C}^2$, define
$$d(u,v) = \sqrt{1 - \frac{|(u,v)|^2}{\|u\|^2\|v\|^2}} \ .$$
Show that for any u, v, $w \in \mathbb{C}^2$, and λ, $\mu \in \mathbb{C}$,
    (a) $d(\lambda u, \mu v) = d(u,v)$,
    (b) $d(u,v) \leq d(u,w) + d(w,v)$,
    (c) $d(u,v) = d(z_u, z_v)$, where $z_x = \frac{x_2}{x_1}$ if $x = (x_1, x_2)^T \in \mathbb{C}^2$, $x_1 \neq 0$,
$$d(z_u, z_v) = \frac{|z_u - z_v|}{\sqrt{(1+|z_u|^2)(1+|z_v|^2)}}.$$

. ———— ⊙ ———— .

CHAPTER 7

Positive Semidefinite Matrices

Introduction: This chapter studies the positive semidefinite matrices, concentrating primarily on the inequalities of this type of matrix. The main goal is to present the fundamental results and show some often-used techniques. Section 7.1 gives the basic properties, Section 7.2 treats the Löwner partial ordering of positive semidefinite matrices, and Section 7.3 presents some inequalities of principal submatrices. Section 7.4 derives inequalities of partitioned positive semidefinite matrices using Schur complements, and Sections 7.5 and 7.6 investigate the Hadamard product of the positive semidefinite matrices. Finally, Section 7.7 shows the Cauchy–Schwarz type matrix inequalities and the Wielandt and Kantorovich inequalities.

7.1 Positive Semidefinite Matrices

An n-square complex matrix A is said to be *positive semidefinite* or *nonnegative definite*, written as $A \geq 0$, if

$$x^*Ax \geq 0, \quad \text{for all } x \in \mathbb{C}^n. \tag{7.1}$$

A is further called *positive definite*, symbolized $A > 0$, if the strict inequality in (7.1) holds for all nonzero $x \in \mathbb{C}^n$.

F. Zhang, *Matrix Theory: Basic Results and Techniques*, Universitext,
DOI 10.1007/978-1-4614-1099-7_7, © Springer Science+Business Media, LLC 2011

It is immediate that if A is an $n \times n$ complex matrix, then

$$A \geq 0 \quad \Leftrightarrow \quad X^*AX \geq 0 \tag{7.2}$$

for every $n \times m$ complex matrix X. (Note that one may augment a vector $x \in \mathbb{C}^n$ by zero entries to get a matrix of size $n \times m$.)

The following decomposition theorem (see the spectral decomposition theorem in Chapter 3) of positive semidefinite matrices best characterizes positive semidefiniteness under unitary similarity.

Theorem 7.1 *An $n \times n$ complex matrix A is positive semidefinite if and only if there exists an $n \times n$ unitary matrix U such that*

$$A = U^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n)U, \tag{7.3}$$

where the λ_i are the eigenvalues of A and are all nonnegative. In addition, if $A \geq 0$ then $\det A \geq 0$. *A is positive definite if and only if all the λ_i in (7.3) are positive. Besides, if $A > 0$ then* $\det A > 0$.

Positive semidefinite matrices have many interesting and important properties and play a central role in matrix theory.

Theorem 7.2 *Let A be an n-square Hermitian matrix. Then*

1. *A is positive definite if and only if the determinant of every leading principal submatrix (leading minor) of A is positive.*
2. *A is positive semidefinite if and only if the determinant of every (not just leading) principal submatrix of A is nonnegative.*

Proof. Let A_k be a $k \times k$ principal submatrix of $A \in \mathbb{M}_n$. By permuting rows and columns we may place A_k in the upper-left corner of A. In other words, there exists a permutation matrix P such that A_k is the $(1,1)$-block of P^TAP.

If $A \geq 0$, then (7.1) holds. Thus, for any $x \in \mathbb{C}^k$,

$$x^*A_kx = y^*Ay \geq 0, \quad \text{where } y = P\binom{x}{0} \in \mathbb{C}^n.$$

This says that A_k is positive semidefinite. Therefore, $\det A_k \geq 0$. The strict inequalities hold for positive definite matrix A.

Conversely, if every principal submatrix of A has a nonnegative determinant, then the polynomial in λ (Problem 19 of Section 1.3)

$$\det(\lambda I - A) = \lambda^n - \delta_1 \lambda^{n-1} + \delta_2 \lambda^{n-2} - \cdots + (-1)^n \det A,$$

has no negative zeros, since each δ_i, the sum of the determinants of all the principal matrices of order i, is nonnegative.

The case where A is positive definite follows similarly. ∎

Note that A being Hermitian in Theorem 7.2 is necessary. For instance, all the determinants of the principal submatrices of the matrix $A = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ are positive, but A is not positive (semi)definite.

As a side product of the proof, we see that A is positive (semi)definite if and only if all of its principal submatrices are positive (semi)definite.

It is immediate that $A \geq 0 \ \Rightarrow \ a_{ii} \geq 0$ and that $a_{ii}a_{jj} \geq |a_{ij}|^2$ for $i \neq j$ by considering 2×2 principal submatrices

$$\begin{pmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{pmatrix} \geq 0.$$

Thus, if some diagonal entry $a_{ii} = 0$, then $a_{ij} = 0$ for all j, and hence, $a_{hi} = 0$ for all h, in as much as A is Hermitian. We conclude that some diagonal entry $a_{ii} = 0$ if and only if the row and the column containing a_{ii} consist entirely of 0.

Using Theorem 7.1 and the fact that any square matrix is a product of a unitary matrix and an upper-triangular matrix (QR factorization; see Section 3.2), one can prove the next result (Problem 17).

Theorem 7.3 *The following statements for $A \in \mathbb{M}_n$ are equivalent.*

1. *A is positive semidefinite.*
2. *$A = B^*B$ for some matrix B.*
3. *$A = C^*C$ for some upper-triangular matrix C.*
4. *$A = D^*D$ for some upper-triangular matrix D with nonnegative diagonal entries (Cholesky factorization).*
5. *$A = E^* \left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right) E = F^*F$ for some $n \times n$ invertible matrix E and $r \times n$ matrix F, where r is the rank of A (Rank factorization).*

Every nonnegative number has a unique nonnegative square root. The analogous result for positive semidefinite matrices also holds as we saw in Section 3.2 (Theorem 3.5). We restate the theorem and present a proof that works for linear operators.

Theorem 7.4 *For every $A \geq 0$, there exists a unique $B \geq 0$ so that*

$$B^2 = A.$$

Furthermore, B can be expressed as a polynomial in A.

Proof. We may view n-square matrices as linear operators on $\mathbb{C}^n$. The spectral theorem ensures the existence of orthonormal eigenvectors $u_1, u_2, \ldots, u_n$ belonging to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A, respectively. Then $u_1, u_2, \ldots, u_n$ form an orthonormal basis for $\mathbb{C}^n$ and $A(u_i) = \lambda_i u_i$, $\lambda_i \geq 0$. Define a linear operator B by $B(u_i) = \sqrt{\lambda_i}\, u_i$ for $i = 1, 2, \ldots, n$. It is routine to check that $B^2(x) = A(x)$ and $(B(x), x) \geq 0$ for all vectors x; that is, $B^2 = A$ and $B \geq 0$.

To show the uniqueness, suppose C is also a linear operator such that $C^2(x) = A(x)$ and $(C(x), x) = (x, C(x)) \geq 0$ for all vectors x. If v is an eigenvector of C: $Cv = \mu v$, then $C^2 v = \mu^2 v$, i.e., μ^2 is an eigenvalue of A. Hence, the eigenvalues of C are the nonnegative square roots of the eigenvalues of A; that is, $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n}$.

Choose orthonormal eigenvectors $v_1, v_2, \ldots, v_n$ corresponding to the eigenvalues $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n}$ of C, respectively. Then $v_1, v_2, \ldots, v_n$ form an orthonormal basis for $\mathbb{C}^n$. Let $u_i = w_{1i}v_1 + \cdots + w_{ni}v_n$, $i = 1, 2, \ldots, n$. On one hand, $C^2(u_i) = A(u_i) = \lambda_i u_i = w_{1i}\lambda_i v_1 + \cdots + w_{ni}\lambda_i v_n$, however, $C^2(u_i) = w_{1i}\lambda_1 v_1 + \cdots + w_{ni}\lambda_n v_n$. Because $v_1, v_2, \ldots, v_n$ are linearly independent, we have $w_{ti}\lambda_i = w_{ti}\lambda_t$ for each t. It follows that $w_{ti}\sqrt{\lambda_i} = w_{ti}\sqrt{\lambda_t}$, $t = 1, 2, \ldots, n$. Thus,

$$\begin{aligned} C(u_i) &= C(w_{1i}v_1 + \cdots + w_{ni}v_n) \\ &= w_{1i}\sqrt{\lambda_1}v_1 + \cdots + w_{ni}\sqrt{\lambda_n}\, v_n \\ &= w_{1i}\sqrt{\lambda_i}v_1 + \cdots + w_{ni}\sqrt{\lambda_i}\, v_n \\ &= \sqrt{\lambda_i}u_i = B(u_i). \end{aligned}$$

As $u_1, u_2, \ldots, u_n$ constitute a basis for $\mathbb{C}^n$, we conclude $B = C$.

To see that B is a polynomial of A, let $p(x)$ be a polynomial, by interpolation, such that $p(\lambda_i) = \sqrt{\lambda_i}$, $i = 1, 2, \dots, n$ (Problem 4, Section 5.4). Then it is easy to verify that $p(A) = B$. ∎

Such a matrix B is called the *square root* of A, denoted by $A^{1/2}$.

Note that A^*A is positive semidefinite for every complex matrix A and that the eigenvalues of $(A^*A)^{1/2}$ are the singular values of A. We further discuss the matrix $(A^*A)^{1/2}$ in Chapters 8 and 9.

Problems

1. Show that if A is a positive semidefinite matrix, then so are the matrices $\overline{A}$, A^T, $\operatorname{adj}(A)$, and A^{-1} if the inverse exists.

2. Let A be a positive semidefinite matrix. Show that $\operatorname{tr} A \geq 0$. Equality holds if and only if $A = 0$.

3. Let $A = (a_{ij}) \in \mathbb{M}_n$ and $S = \sum_{i,j=1}^n a_{ij}$. If $A \geq 0$, show that $S \geq 0$.

4. Let $A \in \mathbb{M}_n$ be positive semidefinite. Show that $(\det A)^{1/n} \leq \frac{1}{n} \operatorname{tr} A$.

5. Find a 2×2 nonsymmetric real matrix A such that $x^T A x \geq 0$ for every $x \in \mathbb{R}^2$. What if $x \in \mathbb{C}^2$?

6. Let
$$A = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$
   (a) Is B similar to A?
   (b) Is B congruent to A?
   (c) Is B be obtainable from A by elementary operations?

7. For what real number t is the following $n \times n$ matrix with diagonal entries 1 and off-diagonal entries t positive semidefinite?
$$\begin{pmatrix} 1 & & t \\ & \ddots & \\ t & & 1 \end{pmatrix}.$$

8. For what x, y, $z \in \mathbb{C}$ are the following matrices positive semidefinite?
$$\begin{pmatrix} 1 & 1 & x \\ 1 & 1 & 1 \\ \bar{x} & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & y \\ -1 & \bar{y} & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & z & 0 \\ \bar{z} & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

9. Let $x \in \mathbb{C}, u, v \in \mathbb{C}^n,\ \alpha \in [0, 1]$. Show that (if the powers make sense)

$$\begin{pmatrix} |x|^{2\alpha} & x \\ \bar{x} & |x|^{2(1-\alpha)} \end{pmatrix} \geq 0, \quad \begin{pmatrix} u^*u & u^*v \\ v^*u & v^*v \end{pmatrix} \geq 0.$$

10. If λ, $\mu \in \mathbb{C}$, show that the following matrices are positive semidefinite:

$$\begin{pmatrix} |\lambda|^2+1 & \lambda+\mu \\ \bar{\lambda}+\bar{\mu} & |\mu|^2+1 \end{pmatrix}, \quad \begin{pmatrix} |\lambda|^2 & \lambda\mu \\ \bar{\lambda}\bar{\mu} & |\mu|^2 \end{pmatrix}, \quad \begin{pmatrix} |\lambda|^2+|\mu|^2 & \lambda+\mu \\ \bar{\lambda}+\bar{\mu} & 2 \end{pmatrix}.$$

11. Show that if A is positive definite, so is a principal submatrix of A. Conclude that the diagonal entries of A are all positive.

12. Let $A = (a_{ij}) > 0$ be $n \times n$. Show that the matrix with (i, j)-entry

$$\frac{a_{ij}}{\sqrt{a_{ii}a_{jj}}}, \quad i, j = 1, 2, \ldots, n,$$

is positive definite. What are the diagonal entries of this matrix?

13. Let $A \geq 0$. Show that A can be written as a sum of rank 1 matrices

$$A = \sum_{i=1}^{k} u_i u_i^*,$$

where each u_i is a column vector and $k = \operatorname{rank}(A)$.

14. Let $A \in \mathbb{M}_n$. If $x^*Ax = 0$ for some $x \neq 0$, does it follow that $A = 0$? or $Ax = 0$? What if A is positive semidefinite?

15. Let A be an n-square positive semidefinite matrix. Show that

$$\lambda_{\min}(A) \leq x^*Ax \leq \lambda_{\max}(A), \quad \text{for any unit } x \in \mathbb{C}^n.$$

16. Show that $A \geq 0$ if and only if $A = Q^*Q$ for some matrix Q with linearly independent rows. What is the size of Q?

17. Prove Theorem 7.3. Show further that if $A > 0$ then the Cholesky factorization of A is unique.

18. Show that every positive definite matrix is $*$-congruent to itself; that is, if $A > 0$ then $A^{-1} = P^*AP$ for some invertible matrix P.

19. Let A be a Hermitian matrix. Show that all the eigenvalues of A lie in the interval $[a, b]$ if and only if $A - aI \geq 0$ and $bI - A \geq 0$ and that there exist $\alpha > 0$ and $\beta > 0$ such that $\alpha I + A > 0$ and $I + \beta A > 0$.

20. Let A be a Hermitian matrix. If no eigenvalue of A lies in the interval $[a,b]$, show that $A^2 - (a+b)A + abI$ is positive definite.

21. Find a matrix $A \in \mathbb{M}_n$ such that all of its principal submatrices of order not exceeding $n-1$ are positive semidefinite, but A is not.

22. Find a Hermitian matrix A such that the leading minors are all nonnegative, but A is not positive semidefinite.

23. Let $A \in \mathbb{M}_n$. Show that $A \geq 0$ if and only if every leading principal submatrix of A (including A itself) is positive semidefinite.

24. Let $A \in \mathbb{M}_n$ be a singular Hermitian matrix. If A contains a positive definite principal submatrix of order $n-1$, show that $A \geq 0$.

25. Let $A \in \mathbb{M}_n$ be a positive definite matrix. Prove $\operatorname{tr} A \operatorname{tr} A^{-1} \geq n^2$.

26. Does every normal matrix have a normal square root? Is it unique? How about a general complex matrix?

27. Find the square roots for the positive semidefinite matrices

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}, \quad \begin{pmatrix} a & 1 \\ 1 & a^{-1} \end{pmatrix}, \quad a > 0.$$

28. Let A be a Hermitian matrix and $k > 0$ be an odd number. Show that there exists a unique Hermitian matrix B such that $A = B^k$. Show further that if $AP = PA$ for some matrix P, then $BP = PB$.

29. Let A be a nonzero n-square matrix. If A is Hermitian and satisfies

$$\frac{\operatorname{tr} A}{(\operatorname{tr} A^2)^{1/2}} \geq \sqrt{n-1},$$

show that $A \geq 0$. Conversely, if $A \geq 0$, show that

$$\frac{\operatorname{tr} A}{(\operatorname{tr} A^2)^{1/2}} \geq 1.$$

30. Let A be a square complex matrix such that $A + A^* \geq 0$. Show that

$$\det \frac{A + A^*}{2} \leq |\det A|.$$

31. Show that if B commutes with $A \geq 0$, then B commutes with $A^{1/2}$. Thus any positive semidefinite matrix commutes with its square root.

32. Let $A > 0$. Show that $(A^{-1})^{1/2} = (A^{1/2})^{-1}$ (denoted by $A^{-1/2}$).

33. Does a contractive matrix really make matrices "smaller"? To be precise, if A is a positive definite matrix and C is a contraction, i.e., $\sigma_{\max}(C) \leq 1$, both of size $n \times n$, is it true that $A \geq C^*AC$?

34. Let $A \geq 0$ and let B be a principal submatrix of A. Show that B is singular (i.e., $\det B = 0$) if and only if the rows (columns) of A that contain B are linearly dependent.

35. Let A be an $n \times n$ positive definite matrix. Show that

$$(\det A)^{1/n} = \min \frac{\operatorname{tr}(AX)}{n},$$

where the minimum is taken over all n-square $X > 0$ with $\det X = 1$.

36. Let A be a Hermitian matrix. Show that neither A nor $-A$ is positive semidefinite if and only if at least one of the following holds.

    (a) A has a minor of even order with negative sign.

    (b) A has two minors of odd order with opposite signs.

37. Let A and B be $n \times n$ complex matrices. Show that

$$A^*A = B^*B$$

if and only if $B = UA$ for some unitary U. When does

$$\frac{A^* + A}{2} = (A^*A)^{1/2}?$$

38. Let A be a positive definite matrix and $r \in [0, 1]$. Show that

$$rA + (1 - r)I \geq A^r$$

and

$$(Au, u)^r \geq (A^r u, u), \quad \text{for all unit vectors } u.$$

39. Let A and C be $n \times n$ matrices, where A is positive semidefinite and C is contractive. Show that for any real number r, $0 < r < 1$,

$$C^*A^rC \leq (C^*AC)^r.$$

. ———————— ⊙ ———————— .

7.2 A Pair of Positive Semidefinite Matrices

Inequality is one of the main topics in modern matrix theory. In this section we present some inequalities involving two positive semidefinite matrices.

Let A and B be two Hermitian matrices of the same size. If $A-B$ is positive semidefinite, we write

$$A \geq B \quad \text{or} \quad B \leq A.$$

It is easy to see that $\geq$ is a partial ordering, referred to as *Löwner (partial) ordering*, on the set of Hermitian matrices; that is,

1. $A \geq A$ for every Hermitian matrix A.
2. If $A \geq B$ and $B \geq A$, then $A = B$.
3. If $A \geq B$ and $B \geq C$, then $A \geq C$.

Obviously, $A + B \geq B$ if $A \geq 0$. That $A \geq 0 \Leftrightarrow X^*AX \geq 0$ in (7.2) of the previous section immediately generalizes as follows.

$$A \geq B \quad \Leftrightarrow \quad X^*AX \geq X^*BX \tag{7.4}$$

for every complex matrix X of appropriate size. If A and B are both positive semidefinite, then $(A^{1/2})^* = A^{1/2}$ and thus $A^{1/2}BA^{1/2} \geq 0$.

Theorem 7.5 *Let $A \geq 0$ and $B \geq 0$ be of the same size. Then*

1. *The trace of the product AB is less than or equal to the product of the traces $\operatorname{tr} A$ and $\operatorname{tr} B$; that is, $\operatorname{tr}(AB) \leq \operatorname{tr} A \operatorname{tr} B$.*
2. *The eigenvalues of AB are all nonnegative. Furthermore, AB is positive semidefinite if and only if $AB = BA$.*
3. *If α, β are the largest eigenvalues of A, B, respectively, then*

$$-\frac{1}{4}\alpha\beta I \leq AB + BA \leq 2\alpha\beta I.$$

Proof. To show (1), by unitary similarity, with $A = U^*DU$,

$$\operatorname{tr}(AB) = \operatorname{tr}(U^*DUB) = \operatorname{tr}(DUBU^*),$$

we may assume that $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Suppose that $b_{11}, \ldots, b_{nn}$ are the diagonal entries of B. Then

$$\begin{aligned} \operatorname{tr}(AB) &= \lambda_1 b_{11} + \cdots + \lambda_n b_{nn} \\ &\le (\lambda_1 + \cdots + \lambda_n)(b_{11} + \cdots + b_{nn}) \\ &= \operatorname{tr} A \operatorname{tr} B. \end{aligned}$$

For (2), recall that XY and YX have the same eigenvalues if X and Y are square matrices of the same size. Thus, $AB = A^{1/2}(A^{1/2}B)$ has the same eigenvalues as $A^{1/2}BA^{1/2}$, which is positive semidefinite.

AB is not positive semidefinite in general, since it need not be Hermitian. If A and B commute, however, then AB is Hermitian, for

$$(AB)^* = B^*A^* = BA = AB,$$

and thus $AB \ge 0$. Conversely, if $AB \ge 0$, then it is Hermitian, and

$$AB = (AB)^* = B^*A^* = BA.$$

To show (3), we assume that $A \ne 0$ and $B \ne 0$. Dividing through the inequalities by $\alpha\beta$, we see that the statement is equivalent to its case $\alpha = 1$, $\beta = 1$. Thus, we need to show $-\frac{1}{4}I \le AB + BA \le 2I$. Note that $0 \le A \le I$ implies $0 \le A^2 \le A \le I$. It follows that

$$\begin{aligned} 0 &\le (A + B - \tfrac{1}{2}I)^2 \\ &= (A+B)^2 - (A+B) + \tfrac{1}{4}I \\ &= A^2 + B^2 + AB + BA - A - B + \tfrac{1}{4}I \\ &\le AB + BA + \tfrac{1}{4}I; \end{aligned}$$

that is, $AB + BA \ge -\frac{1}{4}I$. To show $AB + BA \le 2I$, we compute

$$0 \le (A-B)^2 = A^2 + B^2 - AB - BA \le 2I - AB - BA. \quad \blacksquare$$

What follows is the main result of this section, which we use to reduce many problems involving a pair of positive definite matrices to a problem involving two diagonal matrices.

Theorem 7.6 *Let A and B be n-square positive semidefinite matrices. Then there exists an invertible matrix P such that*

$$P^*AP \quad \textit{and} \quad P^*BP$$

are both diagonal matrices. In addition, if A is nonsingular, then P can be chosen so that $P^*AP = I$ and P^*BP is diagonal.

Proof. Let $\operatorname{rank}(A + B) = r$ and S be a nonsingular matrix so that

$$S^*(A + B)S = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Conformally partition S^*BS as

$$S^*BS = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

By (7.4), we have $S^*(A + B)S \geq S^*BS$. This implies

$$B_{22} = 0, \quad B_{12} = 0, \quad B_{21} = 0.$$

Now for B_{11}, because $B_{11} \geq 0$, there exists an $r \times r$ unitary matrix T such that $T^*B_{11}T$ is diagonal. Put

$$P = S\begin{pmatrix} T & 0 \\ 0 & I_{n-r} \end{pmatrix}.$$

Then P^*BP and $P^*AP = P^*(A + B)P - P^*BP$ are both diagonal.

If A is invertible, we write $A = C^*C$ for some matrix C. Consider matrix $(C^{-1})^*BC^{-1}$. Since it is positive semidefinite, we have a unitary matrix U such that

$$(C^{-1})^*BC^{-1} = UDU^*,$$

where D is a diagonal matrix with nonnegative diagonal entries.

Let $P = C^{-1}U$. Then $P^*AP = I$ and $P^*BP = D$. ∎

Many results can be derived by reduction of positive semidefinite matrices A and B to diagonal matrices, or further to nonnegative numbers, to which some elementary inequalities may apply. The following two are immediate from the previous theorem by writing $A = P^*D_1P$ and $B = P^*D_2P$, where P is an invertible matrix, and D_1 and D_2 are diagonal matrices with nonnegative entries.

Theorem 7.7 *Let $A \geq 0$, $B \geq 0$ be of the same order (> 1). Then*

$$\det(A+B) \geq \det A + \det B \tag{7.5}$$

with equality if and only if $A+B$ is singular or $A=0$ or $B=0$, and

$$(A+B)^{-1} \leq \frac{1}{4}(A^{-1}+B^{-1}) \tag{7.6}$$

if A and B are nonsingular, with equality if and only if $A=B$.

Theorem 7.8 *If $A \geq B \geq 0$, then*

1. $\operatorname{rank}(A) \geq \operatorname{rank}(B)$,
2. $\det A \geq \det B$, *and*
3. $B^{-1} \geq A^{-1}$ *if A and B are nonsingular.*

Every positive semidefinite matrix has a positive semidefinite square root. The square root is a matrix monotone function for positive semidefinite matrices in the sense that the Löwner partial ordering is preserved when taking the square root.

Theorem 7.9 *Let A and B be positive semidefinite matrices. Then*

$$A \geq B \quad \Rightarrow \quad A^{1/2} \geq B^{1/2}.$$

Proof 1. It may be assumed that A is positive definite by continuity (Problem 5). Let $C = A^{1/2}$, $D = B^{1/2}$, and $E = C - D$. We have to establish $E \geq 0$. For this purpose, it is sufficient to show that the eigenvalues of E are all nonnegative. Notice that

$$0 \leq C^2 - D^2 = C^2 - (C-E)^2 = CE + EC - E^2.$$

It follows that $CE + EC \geq 0$, for E is Hermitian and $E^2 \geq 0$.

On the other hand, let λ be an eigenvalue of E and let u be an eigenvector corresponding to λ. Then λ is real and by (7.1),

$$0 \leq u^*(CE+EC)u = 2\lambda(u^*Cu).$$

Since $C > 0$, we have $\lambda \geq 0$. Hence $E \geq 0$; namely, $C \geq D$.

Proof 2. First notice that $A^{1/2} - B^{1/2}$ is Hermitian. We show that the eigenvalues are all nonnegative. Let $(A^{1/2} - B^{1/2})x = \lambda x,\ x \neq 0$. Then $B^{1/2}x = A^{1/2}x - \lambda x$. By the Cauchy–Schwarz inequality,

$$|x^*y| \le (x^*x)^{1/2}\,(y^*y)^{1/2}, \ \text{ for all } x,\, y \in \mathbb{C}^n.$$

Thus, we have

$$\begin{aligned} x^*Ax &= (x^*Ax)^{1/2}(x^*Ax)^{1/2} \ge (x^*Ax)^{1/2}(x^*Bx)^{1/2} \ge x^*A^{1/2}B^{1/2}x \\ &= x^*A^{1/2}(A^{1/2}x - \lambda x) = x^*Ax - \lambda x^*A^{1/2}x. \end{aligned}$$

It follows that $\lambda x^*A^{1/2}x \ge 0$ for all x in $\mathbb{C}$, so $\lambda \ge 0$.

Proof 3. If A is positive definite, then by using (7.4) and by multiplying both sides of $B \le A$ by $A^{-1/2} = (A^{-1/2})^*$, we obtain

$$A^{-1/2}BA^{-1/2} \le I,$$

rewritten as

$$(B^{1/2}A^{-1/2})^*(B^{1/2}A^{-1/2}) \le I,$$

which gives

$$\sigma_{\max}(B^{1/2}A^{-1/2}) \le 1,$$

where $\sigma_{\max}$ means the largest singular value. Thus, by Problem 10,

$$\lambda_{\max}(A^{-1/4}B^{1/2}A^{-1/4}) = \lambda_{\max}(B^{1/2}A^{-1/2}) \le \sigma_{\max}(B^{1/2}A^{-1/2}) \le 1,$$

where $A^{-1/4}$ is the square root of $A^{-1/2}$, and, by Problem 11,

$$0 \le A^{-1/4}B^{1/2}A^{-1/4} \le I.$$

Multiplying both sides by $A^{1/4}$, the square root of $A^{1/2}$, we see that

$$B^{1/2} \le A^{1/2}.$$

The case for singular A follows from a continuity argument. ∎

Theorem 7.10 *Let A and B be positive semidefinite matrices. Then*

$$A \ge B \quad \Rightarrow \quad A^r \ge B^r, \quad 0 \le r \le 1.$$

This result, due to Löwner and Heinz, can be shown in a similar way as the above proof 3. That is, one can prove that if the inequality holds for $s, t \in [0, 1]$, then it holds for $(s + t)/2$, concluding that the set of numbers in $[0, 1]$ for which the inequality holds is convex.

Problems

1. Show that $A \geq B \Rightarrow \operatorname{tr} A \geq \operatorname{tr} B$. When does equality occur?

2. Give an example where $A \geq 0$ and $B \geq 0$ but AB is not Hermitian.

3. Show by example that $A \geq B \geq 0 \Rightarrow A^2 \geq B^2$ is not true in general. But $AB = BA$ and $A \geq B \geq 0$ imply $A^k \geq B^k$, $k = 1, 2, \ldots$.

4. Referring to Theorem 7.6, give an example showing that the matrix P is not unique in general. Show that matrix P can be chosen to be unitary if and only if $AB = BA$.

5. Show Theorem 7.9 for the singular case by a continuity argument; that is, $\lim_{\epsilon \to 0^+} (A + \epsilon I)^{1/2} = A^{1/2}$ for $A \geq 0$.

6. Complete the proof of Theorem 7.10.

7. If A is an $n \times n$ complex matrix such that $x^*Ax \geq x^*x$ for every $x \in \mathbb{C}^n$, show that A is nonsingular and that $A \geq I \geq A^{-1} > 0$.

8. Let $A, B \in \mathbb{M}_n$. Show that $A > B$ (i.e., $A - B > 0$) if and only if $X^*AX > X^*BX$ for every $n \times m$ matrix X with $\operatorname{rank}(X) = m$.

9. Let A be a nonsingular Hermitian matrix. Show that $A \geq A^{-1}$ if and only if all eigenvalues of A lie in $[-1, 0) \cup [1, \infty)$.

10. Let $A \in \mathbb{M}_n$. If $A \geq 0$, show that $x^*Ax \leq \lambda_{\max}(A)$ for all unit x and that $|\lambda(A)| \leq \sigma_{\max}(A)$ for every eigenvalue $\lambda(A)$ of A.

11. Let $A = A^*$. Show that $0 \leq A \leq I \Leftrightarrow$ every eigenvalue $\lambda(A) \in [0, 1]$.

12. Let $A \geq 0$. Show that $\lambda_{\max}(A)I \geq A$ and $\lambda_{\max}(A) \geq \max\{a_{ii}\}$.

13. Let A and B be n-square positive semidefinite matrices.

    (a) Show that $A > B \geq 0 \Leftrightarrow \lambda_{\max}(A^{-1}B) < 1$.

    (b) Show that $A > B \geq 0 \Rightarrow \det A > \det B$.

    (c) Give an example that $A \geq B \geq 0$, $\det A = \det B$, but $A \neq B$.

14. Let $A \geq 0$ and $B \geq 0$ be of the same size. As is known, the eigenvalues of positive semidefinite matrices are the same as the singular values, and the eigenvalues of AB are nonnegative. Are the eigenvalues of AB in this case necessarily equal to the singular values of AB?

15. Show that the eigenvalues of the product of three positive semidefinite matrices are not necessarily nonnegative by the example

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & i \\ -i & 1 \end{pmatrix}.$$

16. Let $A > 0$ and $B = B^*$ (not necessarily positive semidefinite) be of the same size. Show that there exists a nonsingular matrix P such that $P^*AP = I$, P^*BP is diagonal, and the diagonal entries of P^*BP are the eigenvalues of $A^{-1}B$. Show by example that the assertion is not true in general if $A > 0$ is replaced with $A \geq 0$.

17. Let A, B, C be three n-square positive semidefinite matrices. Give an example showing that there does not necessarily exist an invertible matrix P such that P^*AP, P^*BP, P^*CP are all diagonal.

18. Let A, B be $m \times n$ matrices. Show that for any $m \times m$ matrix $X > 0$

$$|\operatorname{tr}(A^*B)|^2 \leq \operatorname{tr}(A^*XA)\operatorname{tr}(B^*X^{-1}B).$$

19. Let $A = (a_{ij})$ be an $n \times n$ positive semidefinite matrix. Show that

    (a) $\sum_{i=1}^n a_{ii}^2 \leq \operatorname{tr} A^2 \leq \left(\sum_{i=1}^n a_{ii}\right)^2$,

    (b) $\left(\sum_{i=1}^n a_{ii}\right)^{1/2} \leq \operatorname{tr} A^{1/2} \leq \sum_{i=1}^n a_{ii}^{1/2}$,

    (c) $\left(\sum_{i=1}^n a_{ii}\right)^{-1} \leq \sum_{i=1}^n a_{ii}^{-1} \leq \operatorname{tr} A^{-1}$ if $A > 0$.

20. Let $A \geq 0$ and $B \geq 0$ be of the same size. Show that

$$\operatorname{tr}(A^{1/2}B^{1/2}) \leq (\operatorname{tr} A)^{1/2}(\operatorname{tr} B)^{1/2}$$

and

$$\left(\operatorname{tr}(A+B)\right)^{1/2} \leq (\operatorname{tr} A)^{1/2} + (\operatorname{tr} B)^{1/2}.$$

21. Let $A > 0$ and $B > 0$ be of the same size. Show that

$$\operatorname{tr}\left((A^{-1} - B^{-1})(A - B)\right) \leq 0.$$

22. Let $A > 0$ and $B \geq C \geq 0$ be all of the same size. Show that

$$\operatorname{tr}\left((A+B)^{-1}B\right) \geq \operatorname{tr}\left((A+C)^{-1}C\right).$$

23. Construct an example that $A > 0, B > 0$ but $AB + BA \not\geq 0$. Explain why $A^{1/2}BA^{1/2} \leq \frac{1}{2}(AB + BA)$ is not true in general for $A, B \geq 0$.

24. Show that for Hermitian matrices A and B of the same size,

$$A^2 + B^2 \geq AB + BA.$$

25. Let $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & \alpha \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 1 & \beta \\ \beta & 1 \end{smallmatrix}\right)$. Find a condition on α and β so that

$$A^2 + AB + BA \not\geq 0.$$

26. Let A and B be n-square Hermitian matrices. Show that

$$A > 0,\ AB + BA \geq 0 \quad \Rightarrow \quad B \geq 0$$

and

$$A > 0,\ AB + BA > 0 \quad \Rightarrow \quad B > 0.$$

Show that $A > 0$ in fact can be replaced by the weaker condition $A \geq 0$ with positive diagonal entries. Show by example that the assertions do not hold in general if $A > 0$ is replaced by $A \geq 0$.

27. Let A and B be n-square real symmetric invertible matrices. Show that there exists a real n-square invertible matrix P such that $P^T AP$ and $P^T BP$ are both diagonal if and only if all the roots of $p(x) = \det(xA - B)$ and $q(x) = \det(xB - A)$ are real.

28. Let $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. Does there exist a real matrix P such that $P^T AP$ and $P^T BP$ are both diagonal? Does there exist a complex matrix P such that $P^* AP$ and $P^T BP$ are both diagonal?

29. Let $A > 0$ and $B \geq 0$ be of the same size. Show that

$$(A + B)^{-1} \leq A^{-1} \quad (\text{or } B^{-1}).$$

30. Let $A > 0$ and $B > 0$ be of the same size. Show that

$$A^{-1} - (A + B)^{-1} - (A + B)^{-1} B (A + B)^{-1} \geq 0.$$

31. Let $A \geq 0$ and $B \geq 0$ be of the same size. Prove or disprove that

$$(BAB)^\alpha = B^\alpha A^\alpha B^\alpha,$$

where $\alpha = 2,\ \frac{1}{2}$, or -1 if A and B are nonsingular.

32. Let $A \geq 0$ and $B \geq 0$ be of the same size. Show that

$$BA^2 B \leq I \quad \Rightarrow \quad B^{1/2} AB^{1/2} \leq I.$$

33. Let $A \geq 0$ and $B \geq 0$ be of the same size. Consider the inequalities:

$$\text{(a) } A \geq B \quad \text{(b) } A^2 \geq B^2 \quad \text{(c) } BA^2 B \geq B^4 \quad \text{(d) } (BA^2 B)^{1/2} \geq B^2.$$

Show that (a)$\not\Rightarrow$(b). However, (b)$\Rightarrow$(c)$\Rightarrow$(d).

34. Prove Theorem 7.10. Show by example that $A \geq B \geq 0 \not\Rightarrow A^2 \geq B^2$.

35. Let $A \geq 0$ and $B \geq 0$ be of order n, where $n > 1$. Show that

$$\big(\det(A+B)\big)^{1/n} \geq (\det A)^{1/n} + (\det B)^{1/n}.$$

Equality occurs if and only if $A = 0$ or $B = 0$ or $A+B$ is singular or $B = aA$ for some $a > 0$. Conclude that

$$\det(A+B) \geq \det A + \det B$$

with equality if and only if $A = 0$ or $B = 0$ or $A+B$ is singular, and

$$\det(A+B) \geq \det A$$

with equality if and only if $B = 0$ or $A+B$ is singular.

36. Let A, B, and C be $n \times n$ positive semidefinite matrices. Show that

$$\det(A+B) + \det(A+C) \leq \det A + \det(A+B+C).$$

[Hint: Use elementary symmetric functions and compound matrices.]

37. Let $A \geq 0$ and $B \geq 0$ be of the same size. Show that for any $\lambda, \mu \in \mathbb{C}$,

$$|\det(\lambda A + \mu B)| \leq \det(|\lambda| A + |\mu| B).$$

38. Show by example that $A \geq B \geq 0$ does not imply

$$A^{1/2} - B^{1/2} \leq (A-B)^{1/2}.$$

39. Let $A > 0$ and $B \geq 0$ be of the same size. Show that

$$\frac{\operatorname{tr} B}{\operatorname{tr} A} \geq \frac{\det B}{\det A}.$$

40. Let $A \geq 0$ and $B \geq 0$, both $n \times n$. Show that

$$\operatorname{tr}(ABA) \leq \operatorname{tr}(A)\lambda_{\max}(AB)$$

and

$$\operatorname{tr}(A + ABA) \leq \big(n + \operatorname{tr}(AB)\big)\lambda_{\max}(A).$$

41. Show that $A \geq 0$ and $B \geq C \geq 0$ imply the matrix inequality

$$A^{1/2}BA^{1/2} \geq A^{1/2}CA^{1/2} \quad \text{but not} \quad B^{1/2}AB^{1/2} \geq C^{1/2}AC^{1/2}.$$

42. Let A and B be n-square complex matrices. Prove or disprove

$$A^*A \leq B^*B \quad \Rightarrow \quad A^*CA \leq B^*CB, \quad \text{for } C \geq 0.$$

43. Let A and B be n-square complex matrices. Prove or disprove

$$A^*A \le B^*B \quad \Rightarrow \quad AA^* \le BB^*.$$

44. Let $A \ge 0$ and $B \ge 0$ be of the same size. For any $t \in [0,1]$ with $\tilde{t} = 1 - t$, assuming that the involved inverses exist, show that

   (a) $(tA + \tilde{t}B)^{-1} \le tA^{-1} + \tilde{t}B^{-1}$. So $(\frac{A+B}{2})^{-1} \le \frac{A^{-1}+B^{-1}}{2}$.

   (b) $(tA+\tilde{t}B)^{-1/2} \le tA^{-1/2}+\tilde{t}B^{-1/2}$. So $(\frac{A+B}{2})^{-1/2} \le \frac{A^{-1/2}+B^{-1/2}}{2}$.

   (c) $(tA + \tilde{t}B)^{1/2} \ge tA^{1/2} + \tilde{t}B^{1/2}$. So $(\frac{A+B}{2})^{1/2} \ge \frac{A^{1/2}+B^{1/2}}{2}$.

   (d) $(tA + \tilde{t}B)^2 \le tA^2 + \tilde{t}B^2$. So $(\frac{A+B}{2})^2 \le \frac{A^2+B^2}{2}$.
   Show, however, that $(\frac{A+B}{2})^3 \le \frac{A^3+B^3}{2}$ is not true in general.

   (e) $\det(tA+\tilde{t}B) \ge (\det A)^t(\det B)^{\tilde{t}}$. So $\det(\frac{A+B}{2}) \ge \sqrt{\det A \det B}$.

45. Let $A > 0$ and $B > 0$ be matrices of the same size with eigenvalues contained in the closed interval $[m, M]$. Show that

$$\begin{aligned} \frac{2mM}{(m+M)^2}(A+B) &\le 2(A^{-1}+B^{-1})^{-1} \\ &\le A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2} \\ &\le \frac{1}{2}(A+B). \end{aligned}$$

46. Let A, B, and C be Hermitian matrices of the same size. If $A \ge B$ and if C commutes with both AB and $A+B$, show that C commutes with A and B. What if the condition $A \ge B$ is removed?

47. Show that the product AB of a positive definite matrix A and a Hermitian matrix B is diagonalizable. What if A is singular?

48. Let A be an n-square complex matrix. If $A + A^* > 0$, show that every eigenvalue of A has a positive real part. Use this fact to show that $X > 0$ if X is a Hermitian matrix satisfying for some $Y > 0$

$$XY + YX > 0.$$

49. Let A and B be $n \times n$ Hermitian matrices. If $A^k + B^k = 2I_n$ for some positive integer k, show that $A + B \le 2I_n$. [Hint: Consider odd k first. For the even case, show that $A^{2p} + B^{2p} \le 2I \Rightarrow A^p + B^p \le 2I$.]

. ———— ⊙ ———— .

7.3 Partitioned Positive Semidefinite Matrices

In this section we present the Fischer and Hadamard determinant inequalities and the matrix inequalities involving principal submatrices of positive semidefinite matrices.

Let A be a square complex matrix partitioned as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \tag{7.7}$$

where A_{11} is a square submatrix of A. If A_{11} is nonsingular, we have

$$\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix} A \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} = \begin{pmatrix} A_{11} & 0 \\ 0 & \widetilde{A_{11}} \end{pmatrix}, \tag{7.8}$$

where

$$\widetilde{A_{11}} = A_{22} - A_{21}A_{11}^{-1}A_{12}$$

is called the *Schur complement* of A_{11} in A. By taking determinants,

$$\det A = \det A_{11} \ \det \widetilde{A_{11}}.$$

If A is a positive definite matrix, then A_{11} is nonsingular and

$$A_{22} \geq \widetilde{A_{11}} \geq 0.$$

The Fischer determinant inequality follows, for $\det A_{22} \geq \det \widetilde{A_{11}}$.

Theorem 7.11 (Fischer Inequality) *If A is a positive semidefinite matrix partitioned as in (7.7), then*

$$\det A \leq \det A_{11} \det A_{22}$$

with equality if and only if both sides vanish or $A_{12} = 0$. Also

$$|\det A_{12}|^2 \leq \det A_{11} \det A_{22}$$

if the blocks A_{11}, A_{12}, A_{21}, and A_{22} are square matrices of the same size.

Proof. The nonsingular case follows from the earlier discussion. For the singular case, one may replace A with $A + \epsilon I$, $\epsilon > 0$, to obtain the desired inequality by a continuity argument.

If equality holds and both A_{11} and A_{22} are nonsingular, then

$$\det \widetilde{A_{11}} = \det A_{22} = \det(A_{22} - A_{21}A_{11}^{-1}A_{12}).$$

Thus, $A_{21}A_{11}^{-1}A_{12} = 0$ or $A_{12} = 0$ by Theorem 7.7 and Problem 9.

For the second inequality, notice that $A_{22} \geq A_{21}A_{11}^{-1}A_{12} \geq 0$. The assertion follows at once by taking the determinants. ∎

An induction on the size of the matrices gives the following result.

Theorem 7.12 (Hadamard Inequality) *Let A be a positive semidefinite matrix with diagonal entries $a_{11}, a_{22}, \ldots, a_{nn}$. Then*

$$\det A \leq a_{11}a_{22}\cdots a_{nn}.$$

Equality holds if and only if some $a_{ii} = 0$ or A is diagonal.

A direct proof goes as follows. Assume that each $a_{ii} > 0$ and let $D = \operatorname{diag}(a_{11}^{-1/2}, \ldots, a_{nn}^{-1/2})$. Put $B = DAD$. Then B is a positive semidefinite matrix with diagonal entries all equal to 1. By the arithmetic mean–geometric mean inequality, we have

$$n = \operatorname{tr} B = \sum_{i=1}^{n} \lambda_i(B) \geq n\left(\prod_{i=1}^{n} \lambda_i(B)\right)^{1/n} = n(\det B)^{1/n}.$$

This implies $\det B \leq 1$. Thus,

$$\det A = \det(D^{-1}BD^{-1}) = \prod_{i=1}^{n} a_{ii} \det B \leq \prod_{i=1}^{n} a_{ii}.$$

Equality occurs if and only if the eigenvalues of B are identical and $\det B = 1$; that is, B is the identity and A is diagonal.

It follows that for any complex matrix A of size $m \times n$,

$$\det(A^*A) \leq \prod_{j=1}^{n} \sum_{i=1}^{m} |a_{ij}|^2. \tag{7.9}$$

An interesting application of the Hadamard inequality is to show that if $A = B + iC \geq 0$, where B and C are real matrices, then

$$\det A \leq \det B$$

by passing the diagonal entries of A to B through a real orthogonal diagonalization of the real matrix B (Problem 31).

We now turn our attention to the inequalities involving principal submatrices of positive semidefinite matrices.

Let A be an n-square positive semidefinite matrix. We denote in this section by $[A]_\omega$, or simply $[A]$, the $k \times k$ principal submatrix of A indexed by a sequence $\omega = \{i_1, \ldots, i_k\}$, where $1 \leq i_1 < \cdots < i_k \leq n$. We are interested in comparing $f([A])$ and $[f(A)]$, where $f(x)$ is the elementary function x^2, $x^{1/2}$, $x^{-1/2}$, or x^{-1}.

Theorem 7.13 *Let $A \geq 0$ and let $[A]$ be a principal submatrix of the matrix A. Then, assuming that the inverses involved exist,*

$$[A^2] \geq [A]^2, \quad [A^{1/2}] \leq [A]^{1/2}, \quad [A^{-1/2}] \geq [A]^{-1/2}, \quad [A^{-1}] \geq [A]^{-1}.$$

Proof. We may assume that $[A] = A_{11}$ as in (7.7). Otherwise one can carry out a similar argument for $P^T AP$, where P is a permutation matrix so that $[A]$ is in the upper-left corner.

Partition A^2, $A^{1/2}$, and A^{-1} conformally to A in (7.7) as

$$A^2 = \begin{pmatrix} E & F \\ F^* & G \end{pmatrix}, \quad A^{1/2} = \begin{pmatrix} P & Q \\ Q^* & R \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} X & Y \\ Y^* & Z \end{pmatrix}.$$

Upon computation of A^2 using (7.7), we have the first inequality:

$$[A^2] = E = A_{11}^2 + A_{12}A_{21} \geq A_{11}^2 = [A]^2.$$

This yields $[A]^{1/2} \geq [A^{1/2}]$ by replacing A with $A^{1/2}$ and then taking the square root. The third inequality, $[A^{-1/2}] \geq [A]^{-1/2}$, follows from an application of the last inequality to the second.

We are left to show $[A^{-1}] \geq [A]^{-1}$. By Theorem 2.4, we have

$$[A^{-1}] = X = A_{11}^{-1} + A_{11}^{-1}A_{12}\widetilde{A_{11}}^{-1}A_{21}A_{11}^{-1} \geq A_{11}^{-1} = [A]^{-1}. \quad \blacksquare$$

The inequalities in Theorem 7.13 may be unified in the form

$$[f(A)] \geq f([A]), \quad \text{where } f(x) = x^2,\ -x^{1/2},\ x^{-1/2}, \text{ or } x^{-1}.$$

Notice that if A is an n-square positive definite matrix, then for any $n \times m$ matrix B,

$$\begin{pmatrix} A & B \\ B^* & B^*A^{-1}B \end{pmatrix} \geq 0. \tag{7.10}$$

Note also that $B^*A^{-1}B$ is the smallest matrix to make the block matrix positive semidefinite in the Löwner partial ordering sense.

Theorem 7.14 *Let $A \in \mathbb{M}_n$ be a positive definite matrix and let B be an $n \times m$ matrix. Then for any positive semidefinite $X \in \mathbb{M}_m$,*

$$\begin{pmatrix} A & B \\ B^* & X \end{pmatrix} \geq 0 \quad \Leftrightarrow \quad X \geq B^*A^{-1}B.$$

Proof. It is sufficient to notice the matrix identity

$$\begin{pmatrix} I_n & 0 \\ -B^*A^{-1} & I_m \end{pmatrix} \begin{pmatrix} A & B \\ B^* & X \end{pmatrix} \begin{pmatrix} I_n & -A^{-1}B \\ 0 & I_m \end{pmatrix}$$

$$= \begin{pmatrix} A & 0 \\ 0 & X - B^*A^{-1}B \end{pmatrix}. \quad \blacksquare$$

Note that, by Theorem 2.4, $[A^{-1}] = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$. Thus for any positive semidefinite matrix A partitioned as in (7.7),

$$A - \begin{pmatrix} [A^{-1}]^{-1} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} A_{12}A_{22}^{-1}A_{21} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \geq 0. \tag{7.11}$$

Since a principal submatrix of a positive definite matrix is also positive definite, we have, for any $n \times n$ matrices A, B, and C,

$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geq 0 \quad \Rightarrow \quad \begin{pmatrix} [A] & [B] \\ [B^*] & [C] \end{pmatrix} \geq 0.$$

Using the partitioned matrix in (7.10), we obtain the following result.

Theorem 7.15 *Let A be an $n \times n$ positive definite matrix. Then for any $n \times n$ matrix B, with $[\,\cdot\,]$ standing for a principal submatrix,*

$$[B^*][A]^{-1}[B] \le [B^*A^{-1}B].$$

We end the section by presenting a result on partitioned positive semidefinite matrices. It states that for a partitioned positive semidefinite matrix in which each block is square, the resulting matrix of taking determinant of each block is also positive semidefinite. An analogous result for trace is given in Section 7.5.

Theorem 7.16 *Let A be a $kn \times kn$ positive semidefinite matrix partitioned as $A = (A_{ij})$, where each A_{ij} is an $n \times n$ matrix, $1 \le i, j \le k$. Then the $k \times k$ matrix $D = (\det A_{ij})$ is positive semidefinite.*

Proof. By Theorem 7.3, we write $A = R^*R$, where R is $kn \times kn$. Partition $R = (R_1, R_2, \dots, R_k)$, where R_i is $kn \times n$, $i = 1, 2, \dots, k$. Then $A_{ij} = R_i^* R_j$. Applying the Binet–Cauchy formula (Theorem 4.9) to $\det A_{ij} = \det(R_i^* R_j)$ with $\alpha = \{1, 2, \dots, n\}$, we have

$$\det A_{ij} = \det(R_i^* R_j) = \sum_\beta \det R_i^*[\alpha|\beta] \det R_j[\beta|\alpha],$$

where $\beta = \{j_1, \dots, j_n\}$, $1 \le j_1 < \cdots < j_n \le kn$. It follows that

$$\begin{aligned} D = (\det A_{ij}) &= \Big(\sum_\beta \det R_i^*[\alpha|\beta] \det R_j[\beta|\alpha] \Big) \\ &= \sum_\beta \Big(\det R_i^*[\alpha|\beta] \det R_j[\beta|\alpha] \Big) \\ &= \sum_\beta \Big(\det(R_i[\beta|\alpha])^* \det R_j[\beta|\alpha] \Big) \\ &= \sum_\beta T_\beta^* T_\beta \ge 0, \end{aligned}$$

where T_β is the row vector $(\det R_1[\beta|\alpha], \dots, \det R_k[\beta|\alpha])$. ∎

Problems

1. Show that $\begin{pmatrix} A & A \\ A & X \end{pmatrix} \geq 0$ for any $X \geq A \geq 0$.

2. Show that if $A \geq B \geq 0$, then $\begin{pmatrix} A & B \\ B & A \end{pmatrix} \geq 0$.

3. Show that X must be the zero matrix if $\begin{pmatrix} I+X & I \\ I & I-X \end{pmatrix} \geq 0$.

4. Show that $\begin{pmatrix} I & X \\ X^* & X^*X \end{pmatrix} \geq 0$ for any matrix X.

5. Show that $\begin{pmatrix} A & A^{1/2}B^{1/2} \\ B^{1/2}A^{1/2} & B \end{pmatrix} \geq 0$ and $\begin{pmatrix} A & A \\ A^{1/2} & A^{1/2} \end{pmatrix} \geq 0$ if $A, B \geq 0$.

6. Let $A > 0$ and $B > 0$ be of the same size. Show that
$$\begin{pmatrix} A+B & A \\ A & A+X \end{pmatrix} \geq 0 \quad \Leftrightarrow \quad X \geq -(A^{-1} + B^{-1})^{-1}.$$

7. Refer to Theorem 7.11. Show the reversal Fischer inequality
$$\det \widetilde{A_{11}} \det \widetilde{A_{22}} \leq \det A.$$

8. Show the Hadamard determinant inequality by Theorem 7.3(4).

9. Let A be an $n \times m$ complex matrix and B be an $n \times n$ positive definite matrix. If $A^*BA = 0$, show that $A = 0$. Does the assertion hold if B is a nonzero positive definite or general nonsingular matrix?

10. When does equality in (7.9) occur?

11. Show that a square complex matrix A is unitary if and only if each row (column) vector of A has length 1 and $|\det A| = 1$.

12. Show that the following matrices are positive semidefinite.

    (a) $\begin{pmatrix} \sigma_{\max}(A)I_n & A^* \\ A & \sigma_{\max}(A)I_m \end{pmatrix}$ for any $m \times n$ matrix A.

    (b) $\begin{pmatrix} A & I \\ I & A^{-1} \end{pmatrix}$ for any positive definite matrix A.

    (c) $\begin{pmatrix} A^*A & A^*B \\ B^*A & B^*B \end{pmatrix}$ for any A and B of the same size.

    (d) $\begin{pmatrix} I+A^*A & A^*+B^* \\ A+B & I+BB^* \end{pmatrix}$ for any A and B of the same size.

    (e) $\begin{pmatrix} \lambda A & A \\ A & \frac{1}{\lambda}A \end{pmatrix}$ for any $\lambda \in (0, +\infty)$ and $A \geq 0$.

    (f) $\begin{pmatrix} |A^*| & A \\ A^* & |A| \end{pmatrix}$ for any matrix A.

13. Let A be $n \times n$ and B be $n \times m$. If A is nonsingular, verify that

$$\begin{pmatrix} A & B \\ B^* & B^*A^{-1}B \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & B^* \end{pmatrix} \begin{pmatrix} A & I \\ I & A^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix}.$$

14. Let $A > 0$ and $B > 0$ be of the same size. Show that, with $[X]$ standing for the corresponding principal submatrices of X,

    (a) $[(A+B)^{-1}] \le [A^{-1}] + [B^{-1}]$,
    (b) $[A+B]^{-1} \le [A]^{-1} + [B^{-1}]$,
    (c) $[A+B]^{-1} \le [(A+B)^{-1}]$,
    (d) $[A]^{-1} + [B]^{-1} \le [A^{-1}] + [B^{-1}]$.

15. Show that for any square complex matrix A, with $[X]$ standing for the corresponding principal submatrices of X,

$$[A^*A] \ge [A^*][A].$$

16. Let A and B be square complex matrices of the same size. With $[X]$ standing for the corresponding principal submatrices of X, show that

$$[AB][B^*A^*] \le [ABB^*A^*]$$

and

$$[A][B][B^*][A^*] \le [A][BB^*][A^*].$$

Show by example that the inequalities below do not hold in general:

$$[A][B][B^*][A^*] \le [ABB^*A^*].$$

$$[AB][B^*A^*] \le [A][BB^*][A^*].$$

$$[A][BB^*][A^*] \le [ABB^*A^*].$$

Conclude that the following inequality does not hold in general.

$$[B^*][A][B] \le [B^*AB], \quad \text{where } A \ge 0.$$

17. Let A be a positive semidefinite matrix. Show by the given A that

$$[A^4] \ge [A]^4$$

is not true in general, where $[\,\cdot\,]$ stands for a principal submatrix and

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

What is wrong with the proof, using $[A^2] \ge [A]^2$ in Theorem 7.13,

$$[A^4] = [(A^2)^2] \ge [A^2]^2 \ge ([A]^2)^2 = [A]^4?$$

18. Let $A > 0$ and $B > 0$ be of the same size. Show that
$$A - X^*B^{-1}X > 0 \quad \Leftrightarrow \quad B - XA^{-1}X^* > 0.$$

19. Let $A > 0$ and $B \geq 0$ be of the same size. Show that
$$\operatorname{tr}(A^{-1}B) \geq \frac{\operatorname{tr} B}{\lambda_{\max}(A)} \geq \frac{\operatorname{tr} B}{\operatorname{tr} A}$$
and
$$\operatorname{tr}\left((I + AB)^{-1}A\right) \geq \frac{\operatorname{tr} A}{1 + \lambda_{\max}(AB)}.$$

20. Let $A \in \mathbb{M}_n$, $C \in \mathbb{M}_m$, and B be $n \times m$. Show that
$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geq 0 \quad \Rightarrow \quad \operatorname{tr}(B^*B) \leq \operatorname{tr} A \operatorname{tr} C.$$
Does
$$\det(B^*B) \leq \det A \det C?$$

21. Let A, B, and C be n-square matrices. If $AB = BA$, show that
$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geq 0 \quad \Rightarrow \quad A^{1/2}CA^{1/2} \geq B^*B.$$

22. Let A, B, and C be n-square matrices. Show that
$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geq 0 \quad \Rightarrow \quad A \pm (B + B^*) + C \geq 0.$$

23. Let A, B, and C be n-square matrices. Show that
$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geq 0 \quad \Rightarrow \quad \begin{pmatrix} \Sigma(A) & \Sigma(B) \\ \Sigma(B^*) & \Sigma(C) \end{pmatrix} \geq 0,$$
where $\Sigma(X)$ denotes the sum of all entries of matrix X.

24. Give an example of square matrices A, B, C of the same size for which
$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geq 0 \quad \text{but} \quad \begin{pmatrix} A & B^* \\ B & C \end{pmatrix} \ngeq 0.$$

25. Give an example of square matrices A, B, C of the same size for which
$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geq 0 \quad \text{but} \quad \begin{pmatrix} A^2 & B^2 \\ (B^*)^2 & C^2 \end{pmatrix} \ngeq 0.$$

26. Let $A \in \mathbb{M}_n$ be a positive definite matrix. Show that for any $B \in \mathbb{M}_n$,

$$A + B + B^* + B^* A^{-1} B \geq 0.$$

In particular,

$$I + B + B^* + B^* B \geq 0.$$

27. For any nonzero vectors $x_1, x_2, \ldots, x_n$ in an inner product space, let

$$G(x_1, x_2, \ldots, x_n) = ((x_j, x_i)).$$

Such a matrix is called the *Gram matrix* of $x_1, x_2, \ldots, x_n$. Show that

$$G(x_1, x_2, \ldots, x_n) \geq 0$$

and that

$$\det G(x_1, \ldots, x_n) \leq \prod_{i=1}^{n} (x_i, x_i).$$

Equality holds if and only if the vectors are orthogonal. Moreover,

$$\det G(x_1, \ldots, x_n) \leq \det G(x_1, \ldots, x_m) \det G(x_{m+1}, \ldots, x_n).$$

28. Let $u_1, u_2, \ldots, u_n \in \mathbb{C}^m$ be n column vectors of m components. Form four matrices as follows by these vectors. Show that the first three matrices are positive semidefinite, and the last one is not in general.

$$\begin{pmatrix} u_1^* u_1 & u_1^* u_2 & \ldots & u_1^* u_n \\ u_2^* u_1 & u_2^* u_2 & \ldots & u_2^* u_n \\ \cdots & \cdots & \cdots & \cdots \\ u_n^* u_1 & u_n^* u_2 & \ldots & u_n^* u_n \end{pmatrix}, \quad \begin{pmatrix} u_1^* u_1 & u_2^* u_1 & \ldots & u_n^* u_1 \\ u_1^* u_2 & u_2^* u_2 & \ldots & u_n^* u_2 \\ \cdots & \cdots & \cdots & \cdots \\ u_1^* u_n & u_2^* u_n & \ldots & u_n^* u_n \end{pmatrix},$$

$$\begin{pmatrix} u_1 u_1^* & u_1 u_2^* & \ldots & u_1 u_n^* \\ u_2 u_1^* & u_2 u_2^* & \ldots & u_2 u_n^* \\ \cdots & \cdots & \cdots & \cdots \\ u_n u_1^* & u_n u_2^* & \ldots & u_n u_n^* \end{pmatrix}, \quad \begin{pmatrix} u_1 u_1^* & u_2 u_1^* & \ldots & u_n u_1^* \\ u_1 u_2^* & u_2 u_2^* & \ldots & u_n u_2^* \\ \cdots & \cdots & \cdots & \cdots \\ u_1 u_n^* & u_2 u_n^* & \ldots & u_n u_n^* \end{pmatrix}.$$

29. Use the Hadamard inequality to show the Fischer inequality. [Hint: If B is a principal submatrix of A, where A is Hermitian, then there exists a unitary matrix U such that $U^* B U$ is diagonal.]

30. Show Theorem 7.16 by a compound matrix. [Hint: The $k \times k$ matrix $(\det A_{ij})$ is a principal submatrix of the nth compound matrix of A.]

31. Let $A = B + iC \geq 0$, where B and C are real matrices. Show that

    (a) B is positive semidefinite.

    (b) C is skew-symmetric.

    (c) $\det B \geq \det A$; when does equality occur?

    (d) $\operatorname{rank}(B) \geq \max\{\operatorname{rank}(A),\ \operatorname{rank}(C)\}$.

32. Find a positive semidefinite matrix A partitioned as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

such that

$$\det A = \det A_{11} \det A_{22} \quad \text{but} \quad A_{12} = A_{21}^* \neq 0.$$

33. Show that for any complex matrices A and B of the same size,

$$\begin{pmatrix} \det(A^*A) & \det(A^*B) \\ \det(B^*A) & \det(B^*B) \end{pmatrix} \geq 0, \quad \begin{pmatrix} \operatorname{tr}(A^*A) & \operatorname{tr}(A^*B) \\ \operatorname{tr}(B^*A) & \operatorname{tr}(B^*B) \end{pmatrix} \geq 0,$$

and

$$\begin{vmatrix} A^*A & A^*B \\ B^*A & B^*B \end{vmatrix} = 0.$$

34. Let A be an $n \times n$ positive semidefinite matrix partitioned as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad \text{where } A_{11} \text{ and } A_{22} \text{ are square.}$$

Show that, by writing $A = X^*X$, where $X = (S, T)$ for some S, T,

$$\mathcal{C}(A_{12}) \subseteq \mathcal{C}(A_{11}), \quad \mathcal{C}(A_{21}) \subseteq \mathcal{C}(A_{22}),$$

and

$$\mathcal{R}(A_{12}) \subseteq \mathcal{R}(A_{22}), \quad \mathcal{R}(A_{21}) \subseteq \mathcal{R}(A_{11}).$$

Further show that

$$\operatorname{rank}(A_{11}, A_{12}) = \operatorname{rank}(A_{11}), \quad \operatorname{rank}(A_{21}, A_{22}) = \operatorname{rank}(A_{22}).$$

Derive that $A_{12} = A_{11}P$ and $A_{21} = QA_{11}$ for some P and Q. Thus

$$\max\{\operatorname{rank}(A_{12}), \operatorname{rank}(A_{21})\} \leq \min\{\operatorname{rank}(A_{11}), \operatorname{rank}(A_{22})\}.$$

35. Let $[X]$ stand for a principal submatrix of X. If $A \geq 0$, show that

$$\operatorname{rank}[A^k] = \operatorname{rank}[A]^k = \operatorname{rank}[A], \quad k = 1, 2, \ldots.$$

[Hint: $\operatorname{rank}[AB] \leq \operatorname{rank}[A]$ and $\operatorname{rank}[A^2] = \operatorname{rank}[A]$ for $A, B \geq 0$.]

7.4 Schur Complements and Determinant Inequalities

Making use of Schur complements (or type III elementary operations for partitioned matrices) has appeared to be an important technique in many matrix problems and applications in statistics. In this section, we are concerned with matrix and determinant (in)equalities involving matrices in the forms $I + A^*A$, $I - A^*A$, and $I - A^*B$.

As defined in the previous section, the Schur complement of the nonsingular principal submatrix A_{11} in the partitioned matrix

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

is

$$\widetilde{A_{11}} = A_{22} - A_{21}A_{11}^{-1}A_{12}.$$

Note that if A is positive semidefinite and if A_{11} is nonsingular, then

$$A_{22} \geq \widetilde{A_{11}} \geq 0.$$

Theorem 7.17 *Let $A > 0$ be partitioned as above. Then*

$$A^{-1} = \begin{pmatrix} \widetilde{A_{22}}^{-1} & X \\ Y & \widetilde{A_{11}}^{-1} \end{pmatrix}, \tag{7.12}$$

where

$$X = -A_{11}^{-1}A_{12}\widetilde{A_{11}}^{-1} = -\widetilde{A_{22}}^{-1}A_{12}A_{22}^{-1}$$

and

$$Y = -A_{22}^{-1}A_{21}\widetilde{A_{22}}^{-1} = -\widetilde{A_{11}}^{-1}A_{21}A_{11}^{-1}.$$

The proof of this theorem follows from Theorem 2.4 immediately.

The inverse form (7.12) of A in terms of Schur complements is very useful. We demonstrate an application of it to obtain some determinant inequalities and present the Hua inequality at the end.

Theorem 7.18 *For any n-square complex matrices A and B,*

$$\det(I + AA^*)\det(I + B^*B) \geq |\det(A + B)|^2 + |\det(I - AB^*)|^2$$

with equality if and only if $n = 1$ or $A + B = 0$ or $AB^* = I$.

The determinant inequality proceeds from the following key matrix identity, for which we present two proofs.

$$\begin{aligned} I + AA^* &= (A + B)(I + B^*B)^{-1}(A + B)^* \\ &\quad + (I - AB^*)(I + BB^*)^{-1}(I - AB^*)^*. \end{aligned} \tag{7.13}$$

Note that the left-hand side of (7.13) is independent of B.

Proof 1 for the identity (7.13). Use Schur complements. Let

$$X = \begin{pmatrix} I + B^*B & B^* + A^* \\ A + B & I + AA^* \end{pmatrix}.$$

Then the Schur complement of $I + B^*B$ in X is

$$(I + AA^*) - (A + B)(I + B^*B)^{-1}(A + B)^*. \tag{7.14}$$

On the other hand, we write

$$X = \begin{pmatrix} I & B^* \\ A & I \end{pmatrix}\begin{pmatrix} I & A^* \\ B & I \end{pmatrix}.$$

Then by using (7.12), if $I - AB^*$ is invertible (then so is $I - B^*A$),

$$\begin{aligned} X^{-1} &= \begin{pmatrix} I & A^* \\ B & I \end{pmatrix}^{-1}\begin{pmatrix} I & B^* \\ A & I \end{pmatrix}^{-1} \\ &= \begin{pmatrix} (I - A^*B)^{-1} & -(I - A^*B)^{-1}A^* \\ -(I - BA^*)^{-1}B & (I - BA^*)^{-1} \end{pmatrix} \\ &\quad \times \begin{pmatrix} (I - B^*A)^{-1} & -B^*(I - AB^*)^{-1} \\ -A(I - B^*A)^{-1} & (I - AB^*)^{-1} \end{pmatrix}. \end{aligned}$$

Thus, we have the lower-right corner of X^{-1}, after multiplying out the right-hand side and then taking inverses,

$$(I - AB^*)(I + BB^*)^{-1}(I - AB^*)^*. \tag{7.15}$$

Equating (7.14) and (7.15), by (7.12), results in (7.13). The singular case of $I - AB^*$ follows from a continuity argument by replacing A with ϵA such that $I - \epsilon AB^*$ is invertible and by letting $\epsilon \to 1$.

Proof 2 for the identity (7.13). A direct proof by showing that

$$(I + AA^*) - (I - AB^*)(I + BB^*)^{-1}(I - AB^*)^*$$

equals

$$(A + B)(I + B^*B)^{-1}(A + B)^*.$$

Noticing that

$$B(I + B^*B) = (I + BB^*)B,$$

we have, by multiplying the inverses,

$$(I + BB^*)^{-1}B = B(I + B^*B)^{-1} \tag{7.16}$$

and, by taking the conjugate transpose,

$$B^*(I + BB^*)^{-1} = (I + B^*B)^{-1}B^*. \tag{7.17}$$

Furthermore, the identity

$$I = (I + B^*B)(I + B^*B)^{-1}$$

yields

$$I - B^*B(I + B^*B)^{-1} = (I + B^*B)^{-1} \tag{7.18}$$

and, by switching B and B^*,

$$I - (I + BB^*)^{-1} = BB^*(I + BB^*)^{-1}. \tag{7.19}$$

Upon computation, we have

$$\begin{aligned}
&(I + AA^*) - (I - AB^*)(I + BB^*)^{-1}(I - AB^*)^* \\
&\quad = \; AA^* - AB^*(I + BB^*)^{-1}BA^* + AB^*(I + BB^*)^{-1} \\
&\qquad\; + (I + BB^*)^{-1}BA^* + I - (I + BB^*)^{-1} \quad \text{(by expansion)} \\
&\quad = \; AA^* - AB^*B(I + B^*B)^{-1}A^* + A(I + B^*B)^{-1}B^* \\
&\qquad\; + B(I + B^*B)^{-1}A^* + BB^*(I + BB^*)^{-1} \;\; \text{(by 7.16, 7.17, 7.19)} \\
&\quad = \; A(I + B^*B)^{-1}A^* + A(I + B^*B)^{-1}B^* \\
&\qquad\; + B(I + B^*B)^{-1}A^* + B(I + B^*B)^{-1}B^* \;\; \text{(by 7.17, 7.18)} \\
&\quad = \; (A + B)(I + B^*B)^{-1}(A + B)^* \quad \text{(by factoring).}
\end{aligned}$$

The identity (7.13) thus follows.

Now we are ready to prove the determinant inequality. Recall that (Theorem 7.7 and Problem 35 of Section 7.2) for any positive semidefinite matrices X and Y of the same size (more than 1),

$$\det(X+Y) \geq \det X + \det Y$$

with equality if and only if $X+Y$ is singular or $X=0$ or $Y=0$.

Applying this to (7.13) and noticing that $I+AA^*$ is never singular, we have, when A and B are square matrices of the same size,

$$|\det(I-AB^*)|^2 + |\det(A+B)|^2 \leq \det(I+AA^*)\det(I+B^*B);$$

equality holds if and only if $n=1$ or $A+B=0$ or $AB^*=I$. ∎

As consequences of (7.13), we have the Löwner partial orderings

$$I+AA^* \geq (A+B)(I+B^*B)^{-1}(A+B)^* \geq 0$$

and

$$I+AA^* \geq (I-AB^*)(I+BB^*)^{-1}(I-AB^*)^* \geq 0,$$

and thus the determinant inequalities

$$|\det(A+B)|^2 \leq \det(I+AA^*)\det(I+B^*B)$$

and

$$|\det(I-AB^*)|^2 \leq \det(I+AA^*)\det(I+B^*B). \tag{7.20}$$

Using similar ideas, one can derive the *Hua determinant inequality* (Problem 15) for contractive matrices. Recall that a matrix X is contractive if $I-X^*X \geq 0$, and strictly contractive if $I-X^*X>0$. It is readily seen that the product of contractive matrices is a contractive matrix:

$$I \geq A^*A \;\Rightarrow\; B^*B \geq B^*(A^*A)B \;\Rightarrow\; I \geq (AB)^*(AB).$$

To show Hua's result, one shows that

$$(I-B^*A)(I-A^*A)^{-1}(I-A^*B)$$

is equal to

$$(B - A)^*(I - AA^*)^{-1}(B - A) + (I - B^*B);$$

both expressions are positive semidefinite when A and B are strict contractions. This implies the matrix inequality

$$I - B^*B \le (I - B^*A)(I - A^*A)^{-1}(I - A^*B). \tag{7.21}$$

Theorem 7.19 (Hua Determinant Inequality) *Let A and B be $m \times n$ contractive matrices. Then*

$$|\det(I - A^*B)|^2 \ge \det(I - A^*A)\det(I - B^*B). \tag{7.22}$$

Equality holds if and only if $A = B$.

Hua's determinant inequality is a reversal of (7.20) under the condition that A and B be contractive matrices of the same size. Note that the matrix inequality (7.21) is equivalent to saying

$$\begin{pmatrix} (I - A^*A)^{-1} & (I - B^*A)^{-1} \\ (I - A^*B)^{-1} & (I - B^*B)^{-1} \end{pmatrix} \ge 0.$$

Question: Is the above block matrix the same as

$$\begin{pmatrix} (I - A^*A)^{-1} & (I - A^*B)^{-1} \\ (I - B^*A)^{-1} & (I - B^*B)^{-1} \end{pmatrix}?$$

If not, is the latter block matrix positive semidefinite? Note that in general $\left(\begin{smallmatrix} X & Z^* \\ Z & Y \end{smallmatrix}\right) \ge 0 \not\Rightarrow \left(\begin{smallmatrix} X & Z \\ Z^* & Y \end{smallmatrix}\right) \ge 0$ (see Problem 24, Section 7.3).

Problems

1. Let A_{11} be a principal submatrix of a square matrix A. Show that

$$A > 0 \quad \Leftrightarrow \quad A_{11} > 0 \text{ and } \widetilde{A_{11}} > 0.$$

2. Show by writing $A = (A^{-1})^{-1}$ that for any principal submatrix A_{11},

$$\widetilde{\widetilde{A_{11}}^{-1}}^{\,-1} = A_{11} \quad \text{or} \quad \widetilde{\widetilde{A_{11}}^{-1}} = A_{11}^{-1}.$$

3. Let A_{11} and B_{11} be the corresponding principal submatrices of the $n \times n$ positive semidefinite matrices A and B, respectively. Show that

$$4(A_{11}+B_{11})^{-1} \le \widetilde{A_{22}}^{-1} + \widetilde{B_{22}}^{-1}.$$

4. Let A and B be positive definite. Show that $A \le B \Rightarrow \widetilde{A_{11}} \le \widetilde{B_{11}}$.

5. Let $A \ge 0$. Show that for any matrices X and Y of appropriate sizes,

$$\begin{pmatrix} X^*AX & X^*AY \\ Y^*AX & Y^*AY \end{pmatrix} \ge 0.$$

6. Use the block matrix $\left(\begin{smallmatrix} I & A \\ A^* & I \end{smallmatrix}\right)$ to show the matrix identities

$$(I - A^*A)^{-1} = I + A^*(I - AA^*)^{-1}A$$

and

$$I + AA^*(I - AA^*)^{-1} = (I - AA^*)^{-1}.$$

7. Show that if P is the elementary matrix of a type III operation on

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}; \quad \text{that is,} \quad P = \begin{pmatrix} I & X \\ 0 & I \end{pmatrix},$$

then the Schur complements of A_{11} in A and in P^TAP are the same.

8. Let A, B, C, and D be $n \times n$ nonsingular complex matrices. Show that

$$\begin{vmatrix} A^{-1} & B^{-1} \\ C^{-1} & D^{-1} \end{vmatrix} = \frac{(-1)^n}{\det(ACBD)} \begin{vmatrix} A & C \\ B & D \end{vmatrix}.$$

9. Let A, C be $m \times n$ matrices, and B, D be $m \times p$ matrices. Show that

$$\begin{pmatrix} AA^* + BB^* & AC^* + BD^* \\ CA^* + DB^* & CC^* + DD^* \end{pmatrix} \ge 0.$$

10. Show that for matrices A, B, C, and D of appropriate sizes,

$$|\det(AC + BD)|^2 \le \det(AA^* + BB^*)\det(C^*C + D^*D).$$

In particular, for any two square matrices X and Y of the same size,

$$|\det(X + Y)|^2 \le \det(I + XX^*)\det(I + Y^*Y)$$

and

$$|\det(I + XY)|^2 \le \det(I + XX^*)\det(I + Y^*Y).$$

11. Prove or disprove that for any n-square complex matrices A and B,

$$\det(A^*A + B^*B) = \det(A^*A + BB^*)$$

or

$$\det(A^*A + B^*B) = \det(AA^* + BB^*).$$

12. Let A and B be $m \times n$ matrices. Show that for any $n \times n$ matrix X,

$$\begin{aligned} AA^* + BB^* &= (B + AX)(I + X^*X)^{-1}(B + AX)^* \\ &\quad + (A - BX^*)(I + XX^*)^{-1}(A - BX^*)^*. \end{aligned}$$

13. Denote $\mathcal{H}(X) = \frac{1}{2}(X^* + X)$ for a square matrix X. For any n-square matrices A and B, explain why $(A - B)^*(A - B) \geq 0$. Show that

$$\mathcal{H}(I - A^*B) \geq \frac{1}{2}\Big((I - A^*A) + (I - B^*B)\Big).$$

14. Let A and B be square contractive matrices of the same size. Derive

$$\det(I - A^*A)\det(I - B^*B) + |\det(A^* - B^*)|^2 \leq |\det(I - A^*B)|^2$$

by applying Theorem 7.17 to the block matrix

$$\begin{pmatrix} I - A^*A & I - A^*B \\ I - B^*A & I - B^*B \end{pmatrix} = \begin{pmatrix} I & A^* \\ I & B^* \end{pmatrix} \begin{pmatrix} I & I \\ -A & -B \end{pmatrix}.$$

Show that the determinant of the block matrix on the left-hand side is $(-1)^n|\det(A - B)|^2$. As a consequence of the inequality,

$$|\det(I - A^*B)|^2 \geq \det(I - A^*A)\det(I - B^*B)$$

with equality if and only if $A = B$ when A, B are strict contractions.

15. Show Theorem 7.19 by the method in the second proof of (7.13).

16. Let A, B, C, and D be square matrices of the same size. Show that

$$\begin{aligned} &I + D^*C - (I + D^*B)(I + A^*B)^{-1}(I + A^*C) \\ &= (D - A)^*(I + BA^*)^{-1}(C - B) \end{aligned}$$

if the inverses involved exist, by considering the block matrix

$$\begin{pmatrix} I + A^*B & I + A^*C \\ I + D^*B & I + D^*C \end{pmatrix}.$$

7.5 The Kronecker and Hadamard Products of Positive Semidefinite Matrices

The Kronecker product and Hadamard product were introduced in Chapter 4 and basic properties were presented there. In this section we are interested in the matrix inequalities of the Kronecker and Hadamard products of positive semidefinite matrices.

Theorem 7.20 *Let $A \geq 0$ and $B \geq 0$. Then $A \otimes B \geq 0$.*

Proof. Let $A = U^*CU$ and $B = V^*DV$, where C and D are diagonal matrices with nonnegative entries on the main diagonals, and U and V are unitary matrices. Thus, by Theorem 4.6,

$$A \otimes B = (U^*CU) \otimes (V^*DV) = (U^* \otimes V^*)(C \otimes D)(U \otimes V) \geq 0. \quad \blacksquare$$

Note that A and B in the theorem may have different sizes.

Our next celebrated theorem of Schur on Hadamard products is used repeatedly in deriving matrix inequalities that involve Hadamard products of positive semidefinite matrices.

Theorem 7.21 (Schur) *Let A and B be n-square matrices. Then*

$$A \geq 0, \ B \geq 0 \quad \Rightarrow \quad A \circ B \geq 0$$

and

$$A > 0, \ B > 0 \quad \Rightarrow \quad A \circ B > 0.$$

Proof 1. Since the Hadamard product $A \circ B$ is a principal submatrix of the Kronecker product $A \otimes B$, which is positive semidefinite by the preceding theorem, the positive semidefiniteness of $A \circ B$ follows.

For the positive definite case, it is sufficient to notice that a principal submatrix of a positive definite matrix is also positive definite.

Proof 2. Write, by Theorem 7.3, $A = U^*U$ and $B = V^*V$, and let u_i and v_i be the ith columns of matrices U and V, respectively. Then

$$a_{ij} = u_i^* u_j = (u_j, u_i), \quad b_{ij} = v_i^* v_j = (v_j, v_i)$$

for each pair i and j, and thus (Problem 7, Section 4.3)

$$A \circ B = (a_{ij} b_{ij}) = \Big((u_j, u_i)(v_j, v_i) \Big) = \Big((u_j \otimes v_j, u_i \otimes v_i) \Big) \geq 0.$$

Proof 3. Let $A = \sum_{i=1}^n \lambda_i u_i u_i^*$, where λ_is are the eigenvalues of A, thus nonnegative; u_is are orthonormal column vectors. Denote by U_i the diagonal matrix with the components of u_i on the main diagonal of U_i. Note that $(u_i u_i^*) \circ B = U_i B U_i^*$. We have

$$A \circ B = \Big(\sum_{i=1}^n \lambda_i u_i u_i^* \Big) \circ B = \sum_{i=1}^n \lambda_i (u_i u_i^*) \circ B = \sum_{i=1}^n \lambda_i U_i B U_i^* \geq 0.$$

Proof 4. For vector $x \in \mathbb{C}^n$, denote by $\operatorname{diag} x$ the n-square diagonal matrix with the components of x on the diagonal. We have

$$\begin{aligned} x^*(A \circ B)x &= \operatorname{tr}(\operatorname{diag} x^* \ A \ \operatorname{diag} x \ B^T) \\ &= \operatorname{tr} \Big((B^{1/2})^T \ \operatorname{diag} x^* \ A^{1/2} A^{1/2} \ \operatorname{diag} x \ (B^{1/2})^T \Big) \\ &= \operatorname{tr} \Big(A^{1/2} \ \operatorname{diag} x \ (B^{1/2})^T \Big)^* \Big(A^{1/2} \operatorname{diag} x \ (B^{1/2})^T \Big) \geq 0. \quad \blacksquare \end{aligned}$$

We are now ready to compare the pairs involving squares and inverses such as $(A \circ B)^2$ and $A^2 \circ B^2$, and $(A \circ B)^{-1}$ and $A^{-1} \circ B^{-1}$.

Theorem 7.22 *Let $A \geq 0$ and $B \geq 0$ be of the same size. Then*

$$A^2 \circ B^2 \geq (A \circ B)^2.$$

Moreover, if A and B are nonsingular, then

$$A^{-1} \circ B^{-1} \geq (A \circ B)^{-1} \quad \text{and} \quad A \circ A^{-1} \geq I.$$

Proof. Let a_i and b_i be the ith columns of matrices A and B, respectively. It is easy to verify by a direct computation that

$$(AA^*) \circ (BB^*) = (A \circ B)(A^* \circ B^*) + \sum_{i \neq j} (a_i \circ b_j)(a_i^* \circ b_j^*).$$

It follows that for any matrices A and B of the same size

$$(AA^*) \circ (BB^*) \geq (A \circ B)(A^* \circ B^*).$$

In particular, if A and B are positive semidefinite, then

$$A^2 \circ B^2 \geq (A \circ B)^2.$$

If A and B are nonsingular matrices, then, by Theorem 4.6,

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

Noticing that $A \circ B$ and $A^{-1} \circ B^{-1}$ are principal submatrices of $A \otimes B$ and $A^{-1} \otimes B^{-1}$ in the same position, respectively, we have by Theorem 7.13, with $[X]$ representing a principal submatrix of X,

$$A^{-1} \circ B^{-1} = [(A \otimes B)^{-1}] \geq [A \otimes B]^{-1} = (A \circ B)^{-1}.$$

For the last inequality, replacing B with A^{-1} in the above inequality, we get $A^{-1} \circ A \geq (A \circ A^{-1})^{-1}$, which implies $(A^{-1} \circ A)^2 \geq I$. Taking the square roots of both sides reveals $A^{-1} \circ A \geq I$. ∎

The last inequality can also be proven by induction on the size of the matrices as follows. Partition A and A^{-1} conformally as

$$A = \begin{pmatrix} a & \alpha \\ \alpha^* & A_1 \end{pmatrix} \quad \text{and} \quad A^{-1} = \begin{pmatrix} b & \beta \\ \beta^* & B_1 \end{pmatrix}.$$

By inequality (7.11)

$$A - \begin{pmatrix} \frac{1}{b} & 0 \\ 0 & 0 \end{pmatrix} \geq 0, \quad A^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & A_1^{-1} \end{pmatrix} \geq 0,$$

and by Theorem 7.20,

$$\left(A - \begin{pmatrix} \frac{1}{b} & 0 \\ 0 & 0 \end{pmatrix} \right) \circ \left(A^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & A_1^{-1} \end{pmatrix} \right) \geq 0,$$

which yields

$$A \circ A^{-1} \geq \begin{pmatrix} 1 & 0 \\ 0 & A_1 \circ A_1^{-1} \end{pmatrix}.$$

An induction hypothesis on $A_1 \circ A_1^{-1}$ reveals $A \circ A^{-1} \geq I$.

Theorem 7.23 *Let A, B, and C be n-square matrices.*

$$\textit{If} \ \begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geq 0, \quad \textit{then} \ \ A \circ C \geq \pm B \circ B^*.$$

Proof.

$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geq 0 \Rightarrow \begin{pmatrix} C & B^* \\ B & A \end{pmatrix} \geq 0 \Rightarrow \begin{pmatrix} A \circ C & B \circ B^* \\ B \circ B^* & A \circ C \end{pmatrix} \geq 0.$$

The desired inequalities are immediate from the fact (Problem 4) that $\left(\begin{smallmatrix} H & K \\ K & H \end{smallmatrix}\right) \geq 0 \Leftrightarrow H \geq \pm K$, where $H \geq 0$ and K is Hermitian. ∎

Our next result is an analogue of Theorem 7.16 of Section 7.3.

Theorem 7.24 *Let A be a $kn \times kn$ positive semidefinite matrix partitioned as $A = (A_{ij})$, where each A_{ij} is an $n \times n$ matrix, $1 \leq i, j \leq k$. Then the $k \times k$ matrix $T = (\operatorname{tr} A_{ij})$ is positive semidefinite.*

Proof 1. Let $A = R^*R$, where R is nk-by-nk. Partition $R = (R_1, R_2, \dots, R_k)$, where each R_i is $nk \times n$, $i = 1, 2, \dots, k$. Then $A_{ij} = R_i^* R_j$. Note that $\operatorname{tr}(R_i^* R_j) = \langle R_j, R_i \rangle$, an inner product of the space of $nk \times n$ matrices. Thus $T = (\operatorname{tr}(A_{ij})) = (\operatorname{tr}(R_i^* R_j)) = (\langle R_j, R_i \rangle)$ is a Gram matrix. So $T \geq 0$.

Proof 2. Let $e_t \in \mathbb{C}^n$ denote the column vector with the tth component 1 and 0 elsewhere, $t = 1, 2, \dots, n$. Then for any $n \times n$ matrix $X = (x_{ij})$, $x_{tt} = e_t^* X e_t$. Thus, $\operatorname{tr} A_{ij} = \sum_{t=1}^n e_t^* A_{ij} e_t$. Therefore,

$$\begin{aligned} T = (\operatorname{tr} A_{ij}) &= \left(\sum_{t=1}^n e_t^* A_{ij} e_t \right) = \sum_{t=1}^n (e_t^* A_{ij} e_t) \\ &= \sum_{t=1}^n E_t^* (A_{ij}) E_t = \sum_{t=1}^n E_t^* A E_t \geq 0, \end{aligned}$$

where $E_t = \operatorname{diag}(e_t, \dots, e_t)$ is $nk \times k$, with k copies of e_t. ∎

One can also prove the theorem using a similar idea by first extracting the diagonal entries of each A_{ij} through $B = J_k \otimes I_n$, where J_k is the $k \times k$ matrix all of whose entries are 1. Note that $A \circ B \geq 0$.

Problems

1. Show that $A \geq 0 \;\Leftrightarrow\; \operatorname{tr}(A \circ B) \geq 0$ for all $B \geq 0$, where $A,\ B \in \mathbb{M}_n$.
2. Let $A \geq 0$ and $B \geq 0$ be of the same size. Show that

$$A \geq B \quad \Leftrightarrow \quad A \otimes I \geq B \otimes I \quad \Leftrightarrow \quad A \otimes A \geq B \otimes B.$$

3. Let $A \geq 0$ and $B \geq 0$ be of the same size. Show that

   (a) $\operatorname{tr}(AB) \leq \operatorname{tr}(A \otimes B) = \operatorname{tr} A \operatorname{tr} B \leq \frac{1}{4}(\operatorname{tr} A + \operatorname{tr} B)^2$.

   (b) $\operatorname{tr}(A \circ B) \leq \frac{1}{2} \operatorname{tr}(A \circ A + B \circ B)$.

   (c) $\operatorname{tr}(A \otimes B) \leq \frac{1}{2} \operatorname{tr}(A \otimes A + B \otimes B)$.

   (d) $\det(A \otimes B) \leq \frac{1}{2}(\det(A \otimes A) + \det(B \otimes B))$.

4. Let H and K be n-square Hermitian matrices. Show that

$$\begin{pmatrix} H & K \\ K & H \end{pmatrix} \geq 0 \quad \Leftrightarrow \quad H \geq \pm K.$$

5. Let $A \geq 0$ and $B \geq 0$ be of the same size. Show that

$$\operatorname{rank}(A \circ B) \leq \operatorname{rank}(A) \operatorname{rank}(B).$$

   Show further that if $A > 0$, then $\operatorname{rank}(A \circ B)$ is equal to the number of nonzero diagonal entries of the matrix B.

6. Let $A > 0$ and let λ be any eigenvalue of $A^{-1} \circ A$. Show that

$$\text{(a) } \lambda \geq 1, \quad \text{(b) } A^{-1} + A \geq 2I, \quad \text{(c) } A^{-1} \circ A^{-1} \geq (A \circ A)^{-1}.$$

7. Let A, B, C, and D be $n \times n$ positive semidefinite matrices. Show that

$$A \geq B \quad \Rightarrow \quad A \circ C \geq B \circ C, \quad A \otimes C \geq B \otimes C;$$

$$A \geq B, \ C \geq D \quad \Rightarrow \quad A \circ C \geq B \circ D, \quad A \otimes C \geq B \otimes D.$$

8. Let $A > 0$ and $B > 0$ be of the same size. Show that

$$(A^{-1} \circ B^{-1})^{-1} \leq A \circ B \leq (A^2 \circ B^2)^{1/2}.$$

9. Prove or disprove that for $A \geq 0$ and $B \geq 0$ of the same size

$$A^3 \circ B^3 \geq (A \circ B)^3 \quad \text{or} \quad A^{1/2} \circ B^{1/2} \leq (A \circ B)^{1/2}.$$

10. Show that for any square matrices A and B of the same size

$$(A \circ B)(A^* \circ B^*) \leq (\sigma^2 I) \circ (AA^*),$$

   where $\sigma = \sigma_{\max}(B)$ is the largest singular value of B.

11. Let A and B be complex matrices of the same size. Show that

$$\begin{pmatrix} (AA^*) \circ I & A \circ B \\ A^* \circ B^* & (B^*B) \circ I \end{pmatrix} \geq 0, \quad \begin{pmatrix} (AA^*) \circ (BB^*) & A \circ B \\ A^* \circ B^* & I \end{pmatrix} \geq 0.$$

12. Let $A \geq 0$ and $B \geq 0$ be of the same size. Let λ be the largest eigenvalue of A and μ be the largest diagonal entry of B. Show that

$$A \circ B \leq \lambda I \circ B \leq \lambda \mu I \quad \text{and} \quad \begin{pmatrix} \lambda\mu I & A \circ B \\ A \circ B & \lambda\mu I \end{pmatrix} \geq 0.$$

13. Let A, B, C, D, X, Y, U, and V be $n \times n$ complex matrices and let

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \; N = \begin{pmatrix} X & Y \\ U & V \end{pmatrix}, \; M \odot N = \begin{pmatrix} A \otimes X & B \otimes Y \\ C \otimes U & D \otimes V \end{pmatrix}.$$

Show that $M \circ N$ is a principal submatrix of $M \odot N$ and that $M \odot N$ is a principal submatrix of $M \otimes N$. Moreover, $M \odot N \geq 0$ if M, $N \geq 0$.

14. Let $A = (a_{ij}) \geq 0$. Show that $A \circ A = (a_{ij}^2) \geq 0$ and $A \circ A^T = A \circ \overline{A} = (|a_{ij}|^2) \geq 0$. Show also that $(a_{ij}^3) \geq 0$. How about $(|a_{ij}|^3)$ and $(\sqrt{|a_{ij}|}\,)$? Show, however, that the matrix $\hat{A} = (|a_{ij}|)$ is not necessarily positive semidefinite as one checks for

$$A = \begin{pmatrix} 1 & \alpha & 0 & -\alpha \\ \alpha & 1 & \alpha & 0 \\ 0 & \alpha & 1 & \alpha \\ -\alpha & 0 & \alpha & 1 \end{pmatrix}, \quad \alpha = \frac{1}{\sqrt{2}}.$$

15. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be positive numbers. Use Cauchy matrices to show that the following matrices are positive semidefinite.

$$\left(\frac{1}{\lambda_i + \lambda_j}\right), \quad \left(\frac{1}{\lambda_i \lambda_j}\right), \quad \left(\frac{\lambda_i \lambda_j}{\lambda_i + \lambda_j}\right),$$

$$\left(\frac{1}{\lambda_i^2 + \lambda_j^2}\right), \quad \left(\sqrt{\lambda_i \lambda_j}\right), \quad \left(\frac{2}{\lambda_i^{-1} + \lambda_j^{-1}}\right),$$

$$\left(\frac{1}{\lambda_i(\lambda_i + \lambda_j)\lambda_j}\right), \quad \left(\frac{\sqrt{\lambda_i \lambda_j}}{\lambda_i + \lambda_j}\right), \quad \left(\frac{\lambda_i \lambda_j}{\sqrt{\lambda_i} + \sqrt{\lambda_j}}\right).$$

16. Let $A = (A_{ij})$ be an $nk \times nk$ partitioned matrix, where each block A_{ij} is $n \times n$. If A is positive semidefinite, show that the matrices $C = (C_m(A_{ij}))$ and $E = (E_m(A_{ij}))$ are positive semidefinite, where $C_m(X)$ and $E_m(X)$, $1 \leq m \leq n$, denote the mth compound matrix and the mth elementary symmetric function of $n \times n$ matrix X, respectively. Deduce that the positivity of $A = (A_{ij})$ implies that of $D = (\det(A_{ij}))$. [Hint: See Section 4.4 on compound matrices.]

. ———————— ⊙ ———————— .

7.6 Schur Complements and the Hadamard Product

The goal of this section is to obtain some inequalities for matrix sums and Hadamard products using Schur complements.

As we saw earlier (Theorem 7.7 and Theorem 7.22), for any positive definite matrices A and B of the same size

$$(A+B)^{-1} \le \frac{1}{4}(A^{-1}+B^{-1})$$

and

$$(A \circ B)^{-1} \le A^{-1} \circ B^{-1}.$$

These are special cases of the next theorem whose proof uses the fact

$$\begin{pmatrix} A & B \\ B^* & B^*A^{-1}B \end{pmatrix} \ge 0 \quad \text{if } A > 0.$$

Theorem 7.25 *Let A and B be n-square positive definite matrices, and let C and D be any matrices of size $m \times n$. Then*

$$(C+D)(A+B)^{-1}(C+D)^* \le CA^{-1}C^* + DB^{-1}D^*, \tag{7.23}$$

$$(C \circ D)(A \circ B)^{-1}(C \circ D)^* \le (CA^{-1}C^*) \circ (DB^{-1}D^*). \tag{7.24}$$

Proof. Note that $X \ge 0$, $Y \ge 0$, $X+Y \ge 0$, and $X \circ Y \ge 0$, where

$$X = \begin{pmatrix} A & C^* \\ C & CA^{-1}C^* \end{pmatrix}, \quad Y = \begin{pmatrix} B & D^* \\ D & DB^{-1}D^* \end{pmatrix}.$$

The inequalities are immediate by taking the Schur complement of the $(1,1)$-block in $X+Y \ge 0$ and $X \circ Y \ge 0$, respectively. ∎

An alternative approach to proving (7.24) is to use Theorem 7.15 with the observation that $X \circ Y$ is a principal submatrix of $X \otimes Y$.

By taking $A = B = I_n$ in (7.23) and (7.24), we have

$$\frac{1}{2}(C+D)(C+D)^* \le CC^* + DD^*$$

and

$$(C \circ D)(C \circ D)^* \le (CC^*) \circ (DD^*).$$

Theorem 7.26 *Let A and B be positive definite matrices of the same size partitioned conformally as*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

Then

$$\widetilde{A_{11} + B_{11}} \geq \widetilde{A_{11}} + \widetilde{B_{11}} \tag{7.25}$$

and

$$\widetilde{A_{11} \circ B_{11}} \geq \widetilde{A_{11}} \circ \widetilde{B_{11}}. \tag{7.26}$$

Proof. Let

$$\hat{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{21}A_{11}^{-1}A_{12} \end{pmatrix}, \quad \hat{B} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{21}B_{11}^{-1}B_{12} \end{pmatrix}.$$

The inequality (7.25) is obtained by taking the Schur complement of $A_{11} + B_{11}$ in $\hat{A} + \hat{B}$ and using (7.23). For (7.26), notice that

$$A_{22} \geq A_{21}A_{11}^{-1}A_{12}, \quad B_{22} \geq B_{21}B_{11}^{-1}B_{12}.$$

Therefore,

$$\begin{aligned} &A_{22} \circ (B_{21}B_{11}^{-1}B_{12}) + B_{22} \circ (A_{21}A_{11}^{-1}A_{12}) \\ &\quad \geq 2\Big((A_{21}A_{11}^{-1}A_{12}) \circ (B_{21}B_{11}^{-1}B_{12})\Big). \end{aligned}$$

It follows that

$$\begin{aligned} \widetilde{A_{11}} \circ \widetilde{B_{11}} &= (A_{22} - A_{21}A_{11}^{-1}A_{12}) \circ (B_{22} - B_{21}B_{11}^{-1}B_{12}) \\ &\leq A_{22} \circ B_{22} - (A_{21}A_{11}^{-1}A_{12}) \circ (B_{21}B_{11}^{-1}B_{12}). \end{aligned}$$

Applying (7.24) to the right-hand side of the above inequality yields

$$\begin{aligned} &A_{22} \circ B_{22} - (A_{21}A_{11}^{-1}A_{12}) \circ (B_{21}B_{11}^{-1}B_{12}) \\ &\quad \leq A_{22} \circ B_{22} - (A_{21} \circ B_{21})(A_{11} \circ B_{11})^{-1}(A_{12} \circ B_{12}) \\ &\quad = \widetilde{A_{11} \circ B_{11}}. \end{aligned}$$

Thus

$$\widetilde{A_{11}} \circ \widetilde{B_{11}} \leq \widetilde{A_{11} \circ B_{11}}. \quad \blacksquare$$

We end this section with a determinant inequality of Oppenheim. The proof uses a Schur complement technique.

As we recall, for $A \geq 0$ and any principal submatrix A_{11} of A

$$\det A = \det A_{11} \det \widetilde{A_{11}}.$$

Theorem 7.27 (Oppenheim) *Let A and B be $n \times n$ positive semidefinite matrices with diagonal entries a_{ii} and b_{ii}, respectively. Then*

$$\prod_{i=1}^{n} a_{ii} b_{ii} \geq \det(A \circ B) \geq a_{11} \cdots a_{nn} \det B \geq \det A \det B.$$

Proof. The first and last inequalities are immediate from the Hadamard determinant inequality. We show the second inequality.

Let $\hat{B}$ be as in the proof of the preceding theorem. Consider $A \circ \hat{B}$ this time and use induction on n, the order of the matrices.

If $n = 2$, then it is obvious. Suppose $n > 2$. Notice that

$$B_{21} B_{11}^{-1} B_{12} = B_{22} - \widetilde{B_{11}}.$$

Take the Schur complement of $A_{11} \circ B_{11}$ in $A \circ \hat{B}$ to get

$$A_{22} \circ (B_{22} - \widetilde{B_{11}}) - (A_{21} \circ B_{21})(A_{11} \circ B_{11})^{-1}(A_{12} \circ B_{12}) \geq 0$$

or

$$A_{22} \circ B_{22} - (A_{21} \circ B_{21})(A_{11} \circ B_{11})^{-1}(A_{12} \circ B_{12}) \geq A_{22} \circ \widetilde{B_{11}}.$$

Observe that the left-hand side of the above inequality is the Schur complement of $A_{11} \circ B_{11}$ in $A \circ B$. By taking determinants, we have

$$\det(\widetilde{A_{11} \circ B_{11}}) \geq \det(A_{22} \circ \widetilde{B_{11}}).$$

Multiply both sides by $\det(A_{11} \circ B_{11})$ to obtain

$$\det(A \circ B) \geq \det(A_{11} \circ B_{11}) \det(A_{22} \circ \widetilde{B_{11}}).$$

The assertion then follows from the induction hypothesis on the two determinants on the right-hand side. ∎

Let $\lambda_i(X)$ be the eigenvalues of $n\times n$ matrix X, $i = 1, \ldots, n$. The inequalities in Theorem 7.27 are rewritten in terms of eigenvalues as

$$\prod_{i=1}^{n} a_{ii}b_{ii} \geq \prod_{i=1}^{n} \lambda_i(A\circ B) \geq \prod_{i=1}^{n} a_{ii}\lambda_i(B) \geq \prod_{i=1}^{n} \lambda_i(AB) = \prod_{i=1}^{n} \lambda_i(A)\lambda_i(B).$$

Problems

1. A *correlation matrix* is a positive semidefinite matrix all of whose diagonal entries are equal to 1. Let A be an $n\times n$ positive semidefinite matrix. Show that $\min_X \det(A\circ X) = \det A$, where the minimal value is taken over all $n \times n$ correlation matrices X.

2. Let $A > 0$. Use the identity $\det \widetilde{A_{11}} = \frac{\det A}{\det A_{11}}$ to show that

$$\frac{\det(A+B)}{\det(A_{11}+B_{11})} \geq \frac{\det A}{\det A_{11}} + \frac{\det B}{\det B_{11}}.$$

3. Let A, B, and $A+B$ be invertible matrices. Find the inverse of $\left(\begin{smallmatrix} A & A \\ A & A+B \end{smallmatrix}\right)$. Use the Schur complement of $A+B$ to verify that

$$A - A(A+B)^{-1}A = (A^{-1}+B^{-1})^{-1}.$$

4. Let $A > 0$ and $B > 0$ be of the same size. Show that for all x, $y \in \mathbb{C}^n$,

$$(x+y)^*(A+B)^{-1}(x+y) \leq x^*A^{-1}x + y^*B^{-1}y$$

and

$$(x\circ y)^*(A\circ B)^{-1}(x\circ y) \leq (x^*A^{-1}x)\circ(y^*B^{-1}y).$$

5. Show that for any $m \times n$ complex matrices A and B

$$\begin{pmatrix} A^*A & A^* \\ A & I_m \end{pmatrix} \geq 0, \quad \begin{pmatrix} B^*B & B^* \\ B & I_m \end{pmatrix} \geq 0.$$

Derive the following inequalities using the Schur complement:

$$I_m \geq A(A^*A)^{-1}A^*, \ \text{ if rank}\,(A) = n;$$

$$(A^*A)\circ(B^*B) \geq (A^*\circ B^*)(A\circ B);$$

$$A^*A + B^*B \geq \frac{(A^*+B^*)(A+B)}{2}.$$

6. Let $A \in \mathbb{M}_n$ be a positive definite matrix. Show that for any $B \in \mathbb{M}_n$,
$$\begin{pmatrix} A \circ (B^* A^{-1} B) & B^* \circ B \\ B^* \circ B & A \circ (B^* A^{-1} B) \end{pmatrix} \geq 0.$$
In particular,
$$\begin{pmatrix} A \circ A^{-1} & I \\ I & A \circ A^{-1} \end{pmatrix} \geq 0.$$
Derive the following inequalities using the Schur complement:
$$(A \circ A^{-1})^{-1} \leq A \circ A^{-1};$$
$$\det(B^* \circ B) \leq \det\big(A \circ (B^* A^{-1} B)\big);$$
$$\big(\operatorname{tr}(B^* \circ B)^2\big)^{1/2} \leq \operatorname{tr}\big(A \circ (B^* A^{-1} B)\big);$$
$$I \circ B^* B \geq (B^* \circ B)(I \circ B^* B)^{-1}(B^* \circ B)$$
if B has no zero row or column. Discuss the analogue for sum (+).

7. Let A, B, and C be $n \times n$ complex matrices such that $\left(\begin{smallmatrix} A & B \\ B^* & C \end{smallmatrix}\right) \geq 0$. With $\star$ denoting the sum $+$ or the Hadamard product $\circ$, show that
$$\big(\operatorname{tr}(B^* \star B)^2\big)^{1/2} \leq \operatorname{tr}(A \star C)$$
and
$$\det(B^* \star B) \leq \det(A \star C).$$

8. Let A be a positive definite matrix partitioned as $\left(\begin{smallmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{smallmatrix}\right)$, where A_{11} and A_{22} are square matrices (maybe of different sizes). Show that
$$A \circ A^{-1} \geq \begin{pmatrix} A_{11} \circ A_{11}^{-1} & 0 \\ 0 & \widetilde{A_{11}} \circ \widetilde{A_{11}}^{-1} \end{pmatrix} \geq 0.$$
Conclude that $A \circ A^{-1} \geq I$. Similarly show that $A^T \circ A^{-1} \geq I$.

9. Let A_1 and A_2 be $n \times n$ real contractive matrices. Show that
   (a) $\det(I - A_i) \geq 0$ for $i = 1, 2$.
   (b) $H = (h_{ij}) \geq 0$, where $h_{ij} = 1/\det(I - A_i^* A_j)$, $1 \leq i, j \leq 2$.
   (c) $L = (l_{ij}) \geq 0$, where $l_{ij} = 1/\det(I - A_i^* A_j)^k$, $1 \leq i, j \leq 2$, and k is any positive integer.

[Hint: Use the Hua determinant inequality in Section 7.4.]

7.7 The Wielandt and Kantorovich Inequalities

The Cauchy–Schwarz inequality is one of the most useful and fundamental inequalities in mathematics. It states that for any vectors x and y in an inner product vector space with inner product $(\cdot\,,\,\cdot)$,

$$|(x,y)|^2 \le (x,x)(y,y)$$

and equality holds if and only if x and y are linearly dependent. Thus

$$|y^*x|^2 \le (x^*x)(y^*y)$$

for all column vectors $x, y \in \mathbb{C}^n$. An easy proof of this is to observe

$$(x,y)^*(x,y) = \begin{pmatrix} x^* \\ y^* \end{pmatrix} (x,y) = \begin{pmatrix} x^*x & x^*y \\ y^*x & y^*y \end{pmatrix} \ge 0.$$

In this section we give a refined version of the Cauchy–Schwarz inequality, show a Cauchy–Schwarz inequality involving matrices, and present the Wielandt and Kantorovich inequalities.

Lemma 7.1 *Let $A = \left(\begin{smallmatrix} a & b \\ \bar{b} & c \end{smallmatrix}\right)$ be a nonzero 2×2 positive semidefinite matrix with eigenvalues α and β, $\alpha \ge \beta$. Then*

$$|b|^2 \le \left(\frac{\alpha - \beta}{\alpha + \beta}\right)^2 ac. \tag{7.27}$$

Equality holds if and only if $a = c$ or $\beta = 0$, i.e., A is singular.

Proof. Solving the equation $\det(\lambda I - A) = 0$ reveals the eigenvalues

$$\alpha,\, \beta = \frac{(a+c) \pm \sqrt{(a-c)^2 + 4|b|^2}}{2}.$$

Computing $\frac{\alpha-\beta}{\alpha+\beta}$ and squaring it, we get an equivalent form of (7.27),

$$(a-c)^2(ac - |b|^2) \ge 0.$$

The conclusions follow immediately ∎

Theorem 7.28 *Let x and y be vectors in an inner product space. Let $A = \begin{pmatrix} (x,x) & (x,y) \\ (y,x) & (y,y) \end{pmatrix}$ have eigenvalues α and β, $\alpha \geq \beta$, $\alpha > 0$. Then*

$$|(x,y)|^2 \leq \left(\frac{\alpha - \beta}{\alpha + \beta}\right)^2 (x,x)(y,y). \tag{7.28}$$

Equality holds if and only if x and y have the same length (i.e., $\|x\| = \|y\|$) or are linearly dependent (i.e., A is singular or $\beta = 0$).

Proof. The inequality follows from the lemma at once. For the equality case, we may assume that $\|y\| = 1$. It is sufficient to notice that $(x - ty, x - ty) = 0$; that is, $x = ty$, when $t = (x,y)$. ∎

We proceed to derive more related inequalities in which matrices are involved. To this end, we use the fact that if $0 < r \leq p \leq q \leq s$, then $\frac{q-p}{q+p} \leq \frac{s-r}{s+r}$. This is because $\frac{t-1}{t+1}$ is an increasing function.

Theorem 7.29 *Let A, B, and C be n-square matrices such that*

$$\begin{pmatrix} A & B^* \\ B & C \end{pmatrix} \geq 0.$$

Denote by M the (nonzero) partitioned matrix and let α and β be the largest and smallest eigenvalues of M, respectively. Then

$$|(Bx,y)|^2 \leq \left(\frac{\alpha - \beta}{\alpha + \beta}\right)^2 (Ax,x)(Cy,y), \quad x,\ y \in \mathbb{C}^n. \tag{7.29}$$

Proof. We may assume that x and y are unit (column) vectors. Let

$$N = \begin{pmatrix} x^* & 0 \\ 0 & y^* \end{pmatrix} M \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} = \begin{pmatrix} x^*Ax & x^*B^*y \\ y^*Bx & y^*Cy \end{pmatrix}.$$

Then N is a 2×2 positive semidefinite matrix. Let λ and μ be the eigenvalues of N with $\lambda \geq \mu$. By Lemma 7.1, we have

$$|(Bx,y)|^2 \leq \left(\frac{\lambda - \mu}{\lambda + \mu}\right)^2 (Ax,x)(Cy,y).$$

To get the desired inequality, with $\beta I \leq M \leq \alpha I$, pre- and post-multiplying by $\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}^*$ and $\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}$, respectively, as $\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}^* \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} = I$, we obtain that $\beta I \leq N \leq \alpha I$. It follows that $\beta \leq \mu \leq \lambda \leq \alpha$. Consequently, $0 \leq \frac{\lambda-\mu}{\lambda+\mu} \leq \frac{\alpha-\beta}{\alpha+\beta}$. Inequality (7.29) then follows. ∎

Theorem 7.30 (Weilandt) *Let A be an $n\times n$ positive semidefinite matrix and λ_1 and λ_n be the largest and smallest eigenvalues of A, respectively. Then for all orthogonal n-column vectors x and $y \in \mathbb{C}^n$,*

$$|x^*Ay|^2 \le \left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2 (x^*Ax)(y^*Ay).$$

Proof 1. Let $M=(x,y)^*A(x,y)$. M is positive semidefinite and

$$M=\begin{pmatrix} x^*Ax & x^*Ay \\ y^*Ax & y^*Ay \end{pmatrix}.$$

Because $\lambda_n I \le A \le \lambda_1 I$, multiplying by $(x,y)^*$ from the left and by (x,y) from the right, as $(x,y)^*(x,y)=I_2$, we have $\lambda_n I \le M \le \lambda_1 I$. If α and β are the eigenvalues of the 2×2 matrix M with $\alpha\ge\beta$, $\alpha>0$, then $\lambda_n \le \beta \le \alpha \le \lambda_1$. By Lemma 7.1, we have

$$|y^*Ax|^2 \le \left(\frac{\alpha-\beta}{\alpha+\beta}\right)^2 (x^*Ax)(y^*Ay).$$

Using the fact that $\frac{\alpha-\beta}{\alpha+\beta} \le \frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}$, we obtain the inequality.

Proof 2. Let x and y be orthogonal unit vectors and let θ be a real number such that $e^{i\theta}(Ay,x)=|(Ay,x)|=|x^*Ay|=|y^*Ax|$. Because $\lambda_n I \le A \le \lambda_1 I$, we have, for any complex number c,

$$\lambda_n\|x+cy\|^2 \le (A(x+cy),x+cy) \le \lambda_1\|x+cy\|^2.$$

Expanding the inequalities and setting $c=te^{i\theta}$, $t\in\mathbb{R}$, we have

$$t^2(y^*Ay-\lambda_n)+2t|x^*Ay|+x^*Ax-\lambda_n \ge 0 \tag{7.30}$$

and

$$t^2(\lambda_1-y^*Ay)+2t|x^*Ay|+\lambda_1-x^*Ax \ge 0. \tag{7.31}$$

Multiply (7.30) by λ_1 and (7.31) by λ_n; then add to get

$$t^2(\lambda_1-\lambda_n)y^*Ay+2t(\lambda_1+\lambda_n)|x^*Ay|+(\lambda_1-\lambda_n)x^*Ax \ge 0.$$

Since this is true for all real t, by taking the discriminant, we have

$$(\lambda_1+\lambda_n)^2|x^*Ay|^2 \le (\lambda_1-\lambda_n)^2(x^*Ax)(y^*Ay). \quad \blacksquare$$

In the classical Cauchy–Schwarz inequality $|y^*x| \le (x^*x)(y^*y)$, if we substitute x and y with $A^{1/2}x$ and $A^{-1/2}y$, respectively, where A is an n-square positive definite matrix, we then have

$$|y^*x|^2 \le (x^*Ax)(y^*A^{-1}y).$$

In particular, for any unit column vector x, we obtain

$$1 \le (x^*Ax)(x^*A^{-1}x).$$

A reversal of this is the well-known *Kantorovich inequality*, which is a special case of the following more general result.

Let $f : \mathbb{M}_n \mapsto \mathbb{M}_k$ be a linear transformation. f is said to be *positive* if $f(A) \ge 0$ whenever $A \ge 0$; f is *strictly positive* if $f(A) > 0$ when $A > 0$; and f is *unital* if $f(I_n) = I_k$. Here are some examples:

1. f: $A \mapsto \operatorname{tr} A$ is strictly positive from $\mathbb{M}_n$ to $\mathbb{M}_1 = \mathbb{C}$.
2. g: $A \mapsto X^*AX$ is positive, where X is a fixed $n \times k$ matrix.
3. h: $A \mapsto A_k$ is positive and unital, where A_k is the $k \times k$ leading principal submatrix of A.
4. p: $A \mapsto A \otimes X$ and q: $A \mapsto A \circ X$ are both positive, where X and Y are positive semidefinite matrices.

Theorem 7.31 *Let f be strictly positive and unital and A be a positive definite matrix. Let α and β be positive numbers, $\alpha < \beta$, such that all the eigenvalues of A are contained in the interval $[\alpha, \beta]$. Then*

$$f(A^{-1}) \le \frac{(\alpha+\beta)^2}{4\alpha\beta}(f(A))^{-1}.$$

Proof. Since all the eigenvalues of A are contained in $[\alpha, \beta]$, the matrices $A - \alpha I$ and $\beta I - A$ are both positive semidefinite. As they commute, $(A - \alpha I)(\beta I - A) \ge 0$. This implies

$$\alpha\beta I \le (\alpha+\beta)A - A^2 \quad \text{or} \quad \alpha\beta A^{-1} \le (\alpha+\beta)I - A.$$

Applying f, we have

$$\alpha\beta f(A^{-1}) \le (\alpha+\beta)I - f(A).$$

For any real numbers c and x, $(c - 2x)^2 \geq 0$. It follows that $c - x \leq \frac{1}{4}c^2x^{-1}$ for any real c and positive number x. Thus

$$\alpha\beta f(A^{-1}) \leq (\alpha + \beta)I - f(A) \leq \frac{(\alpha + \beta)^2}{4}(f(A))^{-1}. \quad \blacksquare$$

If we take f in the theorem to be $f(A) = x^*Ax$, where $x \in \mathbb{C}^n$ is a unit vector, then we have the *Kantorovich inequality.*

Theorem 7.32 (Kantorovich) *Let $A \in \mathbb{M}_n$ be positive definite and λ_1, λ_n be its largest and smallest eigenvalues, respectively. Then*

$$(x^*Ax)(x^*A^{-1}x) \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \quad x^*x = 1, \tag{7.32}$$

The Kantorovich inequality has made appearances in a variety of forms. A matrix version is as follows. Let $A \in \mathbb{M}_n$ be a positive definite matrix. Then for any $n \times m$ matrix X satisfying $X^*X = I_m$,

$$(X^*AX)^{-1} \leq X^*A^{-1}X \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}(X^*AX)^{-1}.$$

The first inequality is proven by noting that $I - Y(Y^*Y)^{-1}Y^* \geq 0$ for any matrix Y with columns linearly independent. For the inequalities on the Hadamard product of positive definite matrices, we have

$$(A \circ B)^{-1} \leq A^{-1} \circ B^{-1} \leq \frac{(\lambda + \mu)^2}{4\lambda\mu}(A \circ B)^{-1},$$

where λ is the largest and μ is the smallest eigenvalue of $A \otimes B$.

We leave the proofs to the reader (Problems 18 and 19).

Problems

1. Let $r \geq 1$. Show that for any positive number t such that $\frac{1}{r} \leq t \leq r$,

$$t + \frac{1}{t} \leq r + \frac{1}{r}.$$

2. Let $0 < a < b$. Show that for any $x \in [a, b]$,

$$\frac{1}{x} \leq \frac{a+b}{ab} - \frac{x}{ab}.$$

3. Let x, $y \in \mathbb{C}^n$ be unit vectors. Show that the eigenvalues of the matrix $\begin{pmatrix} 1 & (x,y) \\ (y,x) & 1 \end{pmatrix}$ are contained in $[1-t, 1+t]$, where $t = |(x,y)|$.

4. Let $A = \begin{pmatrix} a & b \\ \bar{b} & c \end{pmatrix}$ be a nonzero 2×2 Hermitian matrix with (necessarily real) eigenvalues α and β, $\alpha \geq \beta$. Show that $2|b| \leq \alpha - \beta$.

5. Let A be an $n \times n$ positive definite matrix. Show that

$$|y^*x|^2 = (x^*Ax)(y^*A^{-1}y)$$

if and only if $y = 0$ or $Ax = cy$ for some constant c.

6. Let $\lambda_1, \ldots, \lambda_n$ be positive numbers. Show that for any x, $y \in \mathbb{C}^n$,

$$\left| \sum_{i=1}^n x_i \overline{y_i} \right|^2 \leq \left(\sum_{i=1}^n \lambda_i |x_i|^2 \right) \left(\sum_{i=1}^n \lambda_i^{-1} |y_i|^2 \right).$$

7. Show that for positive numbers $a_1, a_2, \ldots, a_n$ and any $t \in [0,1]$,

$$\left(\sum_{i=1}^n a_i^{1/2} \right)^2 \leq \left(\sum_{i=1}^n a_i^t \right) \left(\sum_{i=1}^n a_i^{1-t} \right).$$

Equality occurs if and only if $t = \frac{1}{2}$ or all the a_i are equal.

8. Let $A \in \mathbb{M}_n$ be positive semidefinite. Show that for any unit $x \in \mathbb{C}^n$

$$(Ax, x)^2 \leq (A^2x, x).$$

9. Let A, B, and C be $n \times n$ matrices. Assume that A and C are positive definite. Show that the following statements are equivalent.

   (a) $\begin{pmatrix} A & B^* \\ B & C \end{pmatrix} \geq 0$.

   (b) $\lambda_{\max}(BA^{-1}B^*C^{-1}) \leq 1$.

   (c) $|(Bx, y)| \leq \frac{1}{2}\big((Ax, x) + (Cy, y)\big)$ for all $x,\ y \in \mathbb{C}^n$.

10. Show that for any $n \times n$ matrix $A \geq 0$, $m \times n$ matrix B, x, $y \in \mathbb{C}^n$,

$$|(Bx, y)|^2 \leq (Ax, x)(BA^{-1}B^*y, y)$$

and that for any $m \times n$ matrices A, B, and $x \in \mathbb{C}^n,\ y \in \mathbb{C}^m$,

$$\big|\big((A+B)x, y\big)\big|^2 \leq \big((I + A^*A)x, x\big)\big((I + BB^*)y, y\big).$$

11. Let A be an $n \times n$ positive definite matrix. Show that for all $x \neq 0$,

$$1 \leq \frac{(x^*Ax)(x^*A^{-1}x)}{(x^*x)^2} \leq \frac{\lambda_1}{\lambda_n}.$$

12. Let A and B be $n \times n$ Hermitian. If $A \leq B$ or $B \leq A$, show that

$$A \circ B \leq \frac{1}{2}(A \circ A + B \circ B).$$

13. Let A and B be $n \times n$ positive definite matrices with eigenvalues contained in $[m, M]$, where $0 < m < M$. Show that for any $t \in [0, 1]$,

$$tA^2 + (1-t)B^2 - \big(tA + (1-t)B\big)^2 \leq \frac{1}{4}(M-m)^2 I.$$

14. Show the Kantorovich inequality by the Wielandt inequality with

$$y = \|x\|^2(A^{-1}x) - (x^*A^{-1}x)x.$$

15. Show the Kantorovich inequality following the line:

    (a) If $0 < m \leq t \leq M$, then $0 \leq (m + M - t)t - mM$.

    (b) If $0 \leq m \leq \lambda_i \leq M$, $i = 1, \ldots, n$, and $\sum_{i=1}^n |x_i|^2 = 1$, then

    $$S\left(\frac{1}{\lambda}\right) \leq \frac{m + M - S(\lambda)}{mM},$$

    where $S(\frac{1}{\lambda}) = \sum_{i=1}^n \frac{1}{\lambda_i}|x_i|^2$ and $S(\lambda) = \sum_{i=1}^n \lambda_i |x_i|^2$.

    (c) The Kantorovich inequality follows from the inequality

    $$S(\lambda)S\left(\frac{1}{\lambda}\right) \leq \frac{(m+M)^2}{4mM}.$$

16. Let A be an $n \times n$ positive definite matrix having the largest eigenvalue λ_1 and the smallest eigenvalue λ_n. Show that

$$I \leq A \circ A^{-1} \leq \frac{\lambda_1^2 + \lambda_n^2}{2\lambda_1\lambda_n} I.$$

17. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be positive numbers. Show that

$$\max_{i,\, j} \frac{(\lambda_i + \lambda_j)^2}{4\lambda_i\lambda_j} = \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}.$$

18. Let A be an $n \times n$ positive definite matrix with $\lambda_1 = \lambda_{\max}(A)$ and $\lambda_n = \lambda_{\min}(A)$. Show that for all $n \times m$ matrices X, $X^*X = I_m$,

$$(X^*AX)^{-1} \le X^*A^{-1}X \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}(X^*AX)^{-1}.$$

Use the first one to derive the inequality for principal submatrices:

$$[A]^{-1} \le [A^{-1}]$$

and then show the following matrix inequalities.

(a) $X^*AX - (X^*A^{-1}X)^{-1} \le (\sqrt{\lambda_1} - \sqrt{\lambda_n}\,)^2 I$.

(b) $(X^*AX)^2 \le X^*A^2X \le \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}(X^*AX)^2$.

(c) $X^*A^2X - (X^*AX)^2 \le \frac{(\lambda_1-\lambda_n)^2}{4} I$.

(d) $X^*AX \le (X^*A^2X)^{1/2} \le \frac{\lambda_1+\lambda_n}{2\sqrt{\lambda_1\lambda_n}}(X^*AX)$.

(e) $(X^*A^2X)^{1/2} - X^*AX \le \frac{(\lambda_1-\lambda_n)^2}{4(\lambda_1+\lambda_n)} I$.

19. Let A and B be $n \times n$ positive definite matrices. Show that

$$\lambda_{\max}(A \otimes B) = \lambda_{\max}(A)\lambda_{\max}(B) \quad \text{(denoted by } \lambda)$$

and

$$\lambda_{\min}(A \otimes B) = \lambda_{\min}(A)\lambda_{\min}(B) \quad \text{(denoted by } \mu).$$

Derive the following inequalities from the previous problem.

(a) $(A \circ B)^{-1} \le A^{-1} \circ B^{-1} \le \frac{(\lambda+\mu)^2}{4\lambda\mu}(A \circ B)^{-1}$.

(b) $A \circ B - (A^{-1} \circ B^{-1})^{-1} \le (\sqrt{\lambda} - \sqrt{\mu})^2 I$.

(c) $(A \circ B)^2 \le A^2 \circ B^2 \le \frac{(\lambda+\mu)^2}{4\lambda\mu}(A \circ B)^2$.

(d) $(A \circ B)^2 - A^2 \circ B^2 \le \frac{(\lambda-\mu)^2}{4} I$.

(e) $A \circ B \le (A^2 \circ B^2)^{1/2} \le \frac{\lambda+\mu}{2\sqrt{\lambda\mu}} A \circ B$.

(f) $(A^2 \circ B^2)^{1/2} - A \circ B \le \frac{(\lambda-\mu)^2}{4(\lambda+\mu)} I$.

CHAPTER 8

Hermitian Matrices

Introduction: This chapter contains fundamental results of Hermitian matrices and demonstrates the basic techniques used to derive the results. Section 8.1 presents equivalent conditions to matrix Hermitity, Section 8.2 gives some trace inequalities and discusses a necessary and sufficient condition for a square matrix to be a product of two Hermitian matrices, and Section 8.3 develops the min-max theorem and the interlacing theorem for eigenvalues. Section 8.4 deals with the eigenvalue and singular value inequalities for the sum of Hermitian matrices, and Section 8.5 shows a matrix triangle inequality.

8.1 Hermitian Matrices and Their Inertias

A square complex matrix A is said to be *Hermitian* if A is equal to its transpose conjugate, symbolically, $A^* = A$.

Theorem 8.1 *An n-square complex matrix A is Hermitian if and only if there exists a unitary matrix U such that*

$$A = U^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n) U, \tag{8.1}$$

where the λ_i are real numbers (and they are the eigenvalues of A).

F. Zhang, *Matrix Theory: Basic Results and Techniques*, Universitext,
DOI 10.1007/978-1-4614-1099-7_8, © Springer Science+Business Media, LLC 2011

In other words, A is Hermitian if and only if A is unitarily similar to a real diagonal matrix. This is the Hermitian case of the spectral decomposition theorem (Theorem 3.4). The decomposition (8.1) is often used when a trace or norm inequality is under investigation.

Theorem 8.2 *The following statements for $A \in \mathbb{M}_n$ are equivalent.*

1. *A is Hermitian.*
2. *$x^*Ax \in \mathbb{R}$ for all $x \in \mathbb{C}^n$.*
3. *$A^2 = A^*A$.*
4. *$\operatorname{tr} A^2 = \operatorname{tr}(A^*A)$.*

We show that (1)⇔(2) and (1)⇔(3). (1)⇔(4) is similar. It is not difficult to see that (1) and (2) are equivalent, because a complex number a is real if and only if $a^* = a$ and (Problem 16)

$$A^* = A \quad \Leftrightarrow \quad x^*(A^* - A)x = 0 \text{ for all } x \in \mathbb{C}^n.$$

We present four different proofs for (3)⇒(1), each of which shows a common technique of linear algebra and matrix theory. The first proof gives (4)⇒(1) immediately. Other implications are obvious.

Proof 1. Use Schur decomposition. Write $A = U^*TU$, where U is unitary and T is upper-triangular with the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A on the main diagonal. Then $A^2 = A^*A$ implies $T^2 = T^*T$.

By comparison of the main diagonal entries of the matrices on both sides of $T^2 = T^*T$, we have, for each $j = 1, \ldots, n$,

$$\lambda_j^2 = |\lambda_j|^2 + \sum_{i<j} |t_{ij}|^2.$$

It follows that each λ_j is real and that $t_{ij} = 0$ whenever $i < j$. Therefore, T is real diagonal, and thus

$$A^* = (U^*TU)^* = U^*T^*U = U^*TU = A.$$

The trace identity $\operatorname{tr} T^2 = \operatorname{tr}(T^*T)$ yields (4)⇒(1) in the same way.

Proof 2. Use the fact that $\operatorname{tr}(XX^*) = 0 \;\Leftrightarrow\; X = 0$. We show that $\operatorname{tr}(A - A^*)(A - A^*)^* = 0$ to conclude that $A - A^* = 0$ or $A = A^*$.

Upon computation we have

$$(A - A^*)(A - A^*)^* = AA^* - A^2 + A^*A - (A^*)^2,$$

which reveals, by using $\operatorname{tr}(AA^*) = \operatorname{tr}(A^*A)$ and $A^*A = A^2 = (A^*)^2$,

$$\operatorname{tr}(A - A^*)(A - A^*)^* = 0.$$

Proof 3. Use eigenvalues. Let $B = i(A - A^*)$. Then B is Hermitian. We show that B has only zero eigenvalues; consequently, $B = 0$.

Suppose λ is a nonzero eigenvalue of B with eigenvector x:

$$Bx = \lambda x, \quad \lambda \neq 0,\ x \neq 0.$$

Note that the condition $A^*A = A^2$ implies $BA = 0$. We have

$$\lambda A^*x = A^*(Bx) = (BA)^*x = 0.$$

Thus, $A^*x = 0$. But $Bx = \lambda x$ yields $A^*x = Ax + i\lambda x$. Therefore,

$$0 = x^*A^*x = x^*Ax + i\lambda x^*x = \overline{x^*A^*x} + i\lambda x^*x = i\lambda x^*x.$$

It follows that $\lambda = 0$, a contradiction to the assumption $\lambda \neq 0$.

Proof 4. Use inner product. Note that (Problem 16, Section 1.4)

$$\mathbb{C}^n = \operatorname{Ker} A^* \oplus \operatorname{Im} A.$$

Thus, to show $A^* = A$, it suffices to show $A^*x = Ax$ for every $x \in \operatorname{Ker} A^*$ and $x \in \operatorname{Im} A$. If $x \in \operatorname{Ker} A^*$, then, by $A^2 = (A^*)^2 = A^*A$,

$$(Ax, Ax) = (A^*Ax, x) = ((A^*)^2x, x) = 0.$$

This forces $Ax = 0$; namely, $Ax = A^*x$ for every $x \in \operatorname{Ker} A^*$. If $x \in \operatorname{Im} A$, write $x = Ay$, $y \in \mathbb{C}^n$. We then have

$$A^*x = (A^*A)y = A^2y = A(Ay) = Ax. \quad \blacksquare$$

Let A be an $n \times n$ Hermitian matrix. The *inertia* of A is defined to be the ordered triple $(i_+(A), i_-(A), i_0(A))$, where $i_+(A)$, $i_-(A)$, and $i_0(A)$ are the numbers of positive, negative, and zero eigenvalues of A, respectively (including multiplicities). Denote the inertia of A by $\mathrm{In}(A)$, i.e., $\mathrm{In}(A) = (i_+(A), i_-(A), i_0(A))$, or simply (i_+, i_-, i_0) if no confusion is caused. Obviously, $\operatorname{rank}(A) = i_+(A) + i_-(A)$.

The inertias of a nonsingular Hermitian matrix and its inverse are the same since their (necessarily nonzero) eigenvalues are reciprocals of each other. The inertias of similar Hermitian matrices are the same because their eigenvalues are identical. The inertias of $*$-congruent matrices are also the same; this is *Sylvester's law of inertia.* We say that two $n \times n$ complex matrices A and B are *$*$-congruent* if there exists a nonsingular $n \times n$ matrix S such that $B = S^*AS (= \bar{B}^T BS)$.

Theorem 8.3 (Sylvester's Law of Inertia) *Let A and B be Hermitian matrices of the same size. Then A and B are $*$-congruent if and only if they have the same inertia; that is,* $\mathrm{In}\,(A) = \mathrm{In}\,(B)$.

Proof. The spectral theorem ensures that there are positive diagonal matrices E and F with respective sizes $i_+(A)$ and $i_-(A)$ such that A is unitarily similar (*-congruent) to $E \oplus (-F) \oplus 0_{i_0(A)}$. Setting $G = E^{-1/2} \oplus F^{-1/2} \oplus I_{i_0(A)}$ and upon computation, we have

$$G^* \left(E \oplus (-F) \oplus 0_{i_0(A)}\right) G = I_{i_+(A)} \oplus (-I_{i_-(A)}) \oplus 0_{i_0(A)}.$$

A similar argument shows that B is *-congruent to $I_{i_+(B)} \oplus (-I_{i_-(B)}) \oplus 0_{i_0(B)}$. If $\mathrm{In}(A) = \mathrm{In}(B)$, transitivity of *-congruence implies that A and B are *-congruent, i.e., $B = S^*AS$ for some nonsingular S.

For the converse, suppose that $B = S^*AS$ for some nonsingular matrix S. Let $A = U^*MU$ and $B = VNV^*$, where U and V are nonsingular matrices, $M = I_{i_+(A)} \oplus (-I_{i_-(A)}) \oplus 0_{i_0(A)}$ and $N = I_{i_+(B)} \oplus (-I_{i_-(B)}) \oplus 0_{i_0(B)}$. Then $B = S^*AS$ implies that $N = W^*MW$, where $W = USV$ is a nonsingular matrix. Denote the first $i_+(A)$ rows of W by W_1 and the rest of the rows by W_2. Then $N = W^*MW$ reveals $N = W_1{}^*W_1 - W_2^*(I_{i_-(A)} \oplus 0)W_2$. So $N \leq W_1{}^*W_1$. Since $N = I_{i_+(B)} \oplus (-I_{i_-(B)}) \oplus 0_{i_0(B)}$, the leading principal submatrix of $W_1{}^*W_1$ corresponding to $I_{i_+(B)}$ is positive definite. Thus $i_+(B) \leq \operatorname{rank}(W_1{}^*W_1) = \operatorname{rank}(W_1) = i_+(A)$. It follows that $i_-(B) \geq i_-(A)$ as A and B have the same rank. On the other hand, applying the above argument to $-A$ and $-B$ and noting that $i_+(-H) = i_-(H)$ for any Hermitian matrix H, we conclude that $\mathrm{In}(A) = \mathrm{In}(B)$. ∎

Below is a result on Schur complements of Hermitian matrices.

Theorem 8.4 *Let A be Hermitian, A_{11} be a nonsingular principal submatrix of A, and $\widetilde{A_{11}}$ be the Schur complement of A_{11} in A. Then*

$$\mathrm{In}(A) = \mathrm{In}(A_{11}) + \mathrm{In}(\widetilde{A_{11}}).$$

Proof. By permutation similarity (if necessary), we may assume that

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{and define} \quad G \equiv \begin{pmatrix} I & -A_{12}A_{11}^{-1} \\ 0 & I \end{pmatrix}.$$

Then

$$G^*AG = \begin{pmatrix} A_{11} & 0 \\ 0 & \widetilde{A_{11}} \end{pmatrix}.$$

This says the eigenvalues of G^*AG are those of A_{11} and $\widetilde{A_{11}}$, or $\mathrm{In}(G^*AG) = \mathrm{In}(A_{11}) + \mathrm{In}(\widetilde{A_{11}})$, where the triples are added as vectors. The conclusion follows from Sylvester's law of inertia. ∎

Problems

1. What are the differences in the spectral decompositions of normal, Hermitian, positive semidefinite, and unitary matrices?

2. Show that the diagonal entries of a Hermitian matrix are all real.

3. Show that $A^* + A$ and A^*A are Hermitian for any square matrix A.

4. Show that if A and B are Hermitian matrices of the same size, then so are $A + B$ and $A - B$. What about AB and ABA?

5. Let A be an n-square Hermitian matrix. Show that C^*AC is also Hermitian for any $n \times m$ complex matrix C.

6. Show that if A and B are Hermitian matrices of the same size, then $AB = 0 \Leftrightarrow BA = 0$. What if A and B are not Hermitian?

7. Is a matrix similar to a Hermitian matrix necessarily Hermitian? What if unitary similarity is assumed?

8. Show that if matrix A is *skew-Hermitian*, that is, $A^* = -A$, then $A = iB$ for some Hermitian matrix B.

9. Let A be an $n \times n$ positive definite matrix and B be an $n \times n$ skew-Hermitian matrix. Show that the rank of B^*AB is an even number.

10. Let A be a nonsingular skew-Hermitian matrix. Show that $A^2 + A^{-1}$ is nonsingular and that $B = (A^2 - A^{-1})(A^2 + A^{-1})^{-1}$ is unitary.

11. Show that a square complex matrix A can be uniquely written as
$$A = B + iC = S - iT,$$
where B and C are Hermitian, and S and T are skew-Hermitian.

12. Show directly the implication (4)$\Rightarrow$(1) in Theorem 8.2.

13. If A is Hermitian, show that A^2 is positive semidefinite.

14. Find a unitary matrix U such that U^*HU is diagonal, where
$$H = \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}.$$

15. Let A and B be Hermitian matrices of the same size. If $AB - BA$ and $A - B$ commute, show that A and B commute.

16. Let $A \in \mathbb{M}_n$. Show that A is Hermitian if and only if
$$(Ax, y) = (x, Ay), \quad x,\ y \in \mathbb{C}^n,$$
and if and only if
$$(Ax, x) = (x, Ax), \quad x \in \mathbb{C}^n.$$
Is it true that A is Hermitian if
$$(Ax, x) = (x, Ax), \qquad x \in \mathbb{R}^n?$$

17. Let A be an n-square complex matrix. Show that
    (a) $\operatorname{tr}(AX) = 0$ for all Hermitian $X \in \mathbb{M}_n$ if and only if $A = 0$.
    (b) $\operatorname{tr}(AX) \in \mathbb{R}$ for all Hermitian $X \in \mathbb{M}_n$ if and only if $A = A^*$.
    (c) If A is Hermitian and $\operatorname{tr} A \geq \operatorname{Re}\operatorname{tr}(AU)$ for all unitary $U \in \mathbb{M}_n$, then $A \geq 0$.

18. Show that the rank of a Hermitian matrix is the same as the number of nonzero eigenvalues of the matrix and that the rank of a general matrix A equals the number of nonzero singular values of A, but not the number of nonzero eigenvalues (in general).

19. Show that if A is a Hermitian matrix of rank r, then A has a nonsingular principal submatrix of order r. How about a general matrix?

20. Let A be a Hermitian matrix with rank r. Show that all nonzero $r \times r$ principal minors of A have the same sign.

21. Let $A = A^*$ and $\operatorname{tr} A = 0$. If the sum of all 2×2 principal minors of A is zero, show that $A = 0$. [Hint: Use Problem 16 of Section 5.4.]

22. Show that the numbers of positive, negative, and zero eigenvalues in Theorem 8.1 do not depend on the choice of unitary matrix U.

23. Find the rank and inertia of each of following matrices:

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

24. Let A be a Hermitian matrix and B be a principal submatrix of A. Show that $i_\pm(B) \le i_\pm(A)$. Is it true that $i_0(B) \le i_0(A)$?

25. Let A and B be Hermitian matrices of the same size. Show that if $A \le B$ then $i_\pm(A) \le i_\pm(B)$ and that if $A \le B$ and $\operatorname{rank}(A) = \operatorname{rank}(B)$ then A and B are $*$-congruent.

26. Let A be an $n \times n$ Hermitian matrix. Show that for any $n \times m$ matrix Q with rank r, $i_-(Q^*AQ) \le r$ and $i_+(Q^*AQ) \le r$.

27. Let A and B be $n \times n$ nonsingular Hermitian matrices. If the smallest eigenvalue $\lambda_{\min}(B^{-1}A) > 0$, show that $A \ge B \Leftrightarrow B^{-1} \ge A^{-1}$.

28. Compute the inertias of the following partitioned matrices in which I is the $n \times n$ identity matrix and A and B are any $n \times n$ matrices:

$$\begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad \begin{pmatrix} I & I \\ I & 0 \end{pmatrix}, \quad \begin{pmatrix} I + A^*A & I + A^*B \\ I + B^*A & I + B^*B \end{pmatrix}.$$

29. Show that for any $m \times n$ complex matrix A,

$$\operatorname{In}\begin{pmatrix} I_m & A \\ A^* & I_n \end{pmatrix} = (m, 0, 0) + \operatorname{In}(I_n - A^*A) = (n, 0, 0) + \operatorname{In}(I_m - AA^*).$$

30. Show that two unitary matrices are *-congruent if and only if they are similar, i.e., if U and V are unitary, then $U = W^*VW$ for some nonsingular W if and only if $U = R^{-1}VR$ for some nonsingular R.

8.2 The Product of Hermitian Matrices

This section concerns the product of two Hermitian matrices. As is known, the product of two Hermitian matrices is not necessarily Hermitian in general. For instance, take

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Note that the eigenvalues of AB are the nonreal numbers $\frac{1}{2}(3 \pm \sqrt{7}\, i)$.

We first show a trace inequality of the product of two Hermitian matrices, and then we turn our attention to discussing when a matrix product is Hermitian. Note that the trace of a product of two Hermitian matrices is always real although the product is not Hermitian. This is seen as follows. If A and B are Hermitian, then

$$\operatorname{tr}(AB) = \operatorname{tr}(A^*B^*) = \operatorname{tr}(BA)^* = \overline{\operatorname{tr}(BA)} = \overline{\operatorname{tr}(AB)}.$$

That is, $\operatorname{tr}(AB)$ is real. Is this true for three Hermitian matrices?

Theorem 8.5 *Let A and B be n-square Hermitian matrices. Then*

$$\operatorname{tr}(AB)^2 \le \operatorname{tr}(A^2B^2). \tag{8.2}$$

Equality occurs if and only if A and B commute; namely, $AB = BA$.

Proof 1. Let $C = AB - BA$. Using the fact that $\operatorname{tr}(XY) = \operatorname{tr}(YX)$ for any square matrices X and Y of the same size, we compute

$$\begin{aligned} \operatorname{tr}(C^*C) &= \operatorname{tr}(BA - AB)(AB - BA) \\ &= \operatorname{tr}(BA^2B) + \operatorname{tr}(AB^2A) - \operatorname{tr}(BABA) - \operatorname{tr}(ABAB) \\ &= 2\operatorname{tr}(A^2B^2) - 2\operatorname{tr}(AB)^2. \end{aligned}$$

Note that $\operatorname{tr}(A^2B^2) = \operatorname{tr}(AB^2A)$ is real because AB^2A is Hermitian. Thus $\operatorname{tr}(AB)^2$ is real. The inequality (8.2) then follows immediately from the fact that $\operatorname{tr}(C^*C) \ge 0$ with equality if and only if $C = 0$; that is, $AB = BA$.

Proof 2. Since for any unitary matrix $U \in \mathbb{M}_n$, U^*AU is also Hermitian, inequality (8.2) holds if and only if

$$\operatorname{tr}\Big((U^*AU)B\Big)^2 \le \operatorname{tr}\Big((U^*AU)^2B^2\Big).$$

Thus we assume $A = \operatorname{diag}(a_1, \dots, a_n)$ by Schur decomposition. Then

$$\begin{aligned}\operatorname{tr}(A^2B^2) - \operatorname{tr}(AB)^2 &= \sum_{i,j} a_i^2|b_{ij}|^2 - \sum_{i,j} a_ia_j|b_{ij}|^2 \\ &= \sum_{i<j}(a_i - a_j)^2|b_{ij}|^2 \ge 0\end{aligned}$$

with equality if and only if $a_ib_{ij} = a_jb_{ij}$. This implies $AB = BA$.

Proof 3 for the Equality Case. We use the fact that a matrix X is Hermitian if and only if $\operatorname{tr} X^2 = \operatorname{tr}(XX^*)$ (see Theorem 7.2(4)).

Notice that the Hermitity of A and B gives

$$\operatorname{tr}(A^2B^2) = \operatorname{tr}(ABBA) = \operatorname{tr}(AB)(AB)^*.$$

Thus,

$$\operatorname{tr}(AB)^2 = \operatorname{tr}(A^2B^2) \quad\Rightarrow\quad \operatorname{tr}(AB)^2 = \operatorname{tr}(AB)(AB)^*.$$

It follows that AB is Hermitian. Hence,

$$AB = (AB)^* = B^*A^* = BA. \quad \blacksquare$$

Clearly, the product of two Hermitian matrices is Hermitian if and only if these two matrices commute (Problem 4). We are now interested in the following question. When can a given matrix be written as a product of two Hermitian matrices?

Let A be given. Suppose $A = BC$ is a product of two Hermitian matrices B and C of the same size. If B is nonsingular, then

$$A = BC = B(CB)B^{-1} = BA^*B^{-1};$$

namely, A is similar to A^*. This is in fact a necessary and sufficient condition for a matrix to be a product of two Hermitian matrices.

Theorem 8.6 *A square matrix A is a product of two Hermitian matrices if and only if A is similar to A^*.*

Proof. Necessity: Let $A = BC$, where B and C are n-square Hermitian matrices. Then we have at once

$$AB = BCB = B(BC)^* = BA^*$$

and inductively for every positive integer k

$$A^k B = B(A^*)^k. \tag{8.3}$$

We may write, without loss of generality via similarity (Problem 7),

$$A = \begin{pmatrix} J & 0 \\ 0 & K \end{pmatrix},$$

where J and K contain the Jordan blocks of eigenvalues 0 and nonzero, respectively. Note that J is nilpotent and K is invertible.

Partition B and C conformally with A as

$$B = \begin{pmatrix} L & M \\ M^* & N \end{pmatrix}, \quad C = \begin{pmatrix} P & Q \\ Q^* & R \end{pmatrix}.$$

Then (8.3) implies that for each positive integer k

$$K^k M^* = M^*(J^*)^k.$$

Notice that $(J^*)^k = 0$ when $k \geq n$, for J is nilpotent. It follows that $M = 0$, since K is nonsingular. Thus $A = BC$ is the same as

$$\begin{pmatrix} J & 0 \\ 0 & K \end{pmatrix} = \begin{pmatrix} L & 0 \\ 0 & N \end{pmatrix} \begin{pmatrix} P & Q \\ Q^* & R \end{pmatrix}.$$

This yields $K = NR$, and hence N and R are nonsingular.

Taking $k = 1$ in (8.3), we have

$$\begin{pmatrix} J & 0 \\ 0 & K \end{pmatrix} \begin{pmatrix} L & 0 \\ 0 & N \end{pmatrix} = \begin{pmatrix} L & 0 \\ 0 & N \end{pmatrix} \begin{pmatrix} J^* & 0 \\ 0 & K^* \end{pmatrix},$$

which gives $KN = NK^*$, or, because N is invertible,

$$N^{-1}KN = K^*.$$

In other words, K is similar to K^*. On the other hand, any square matrix is similar to its transpose (Theorem 3.14(1)). Thus J is similar to $J^T = J^*$, and it follows that A is similar to A^*.

Sufficiency: We show that if A is similar to A^*, then A can be expressed as a product of two Hermitian matrices. Notice that

$$A = P^{-1}H_1H_2P \quad \Rightarrow \quad A = P^{-1}H_1(P^{-1})^*P^*H_2P.$$

This says if A is similar to a product of Hermitian matrices, then A is in fact a product of Hermitian matrices.

Recall from Theorem 3.14(2) that A is similar to A^* if and only if the Jordan blocks of the nonreal eigenvalues λ of A occur in conjugate pairs. Thus it is sufficient to show that the paired Jordan block

$$\begin{pmatrix} J(\lambda) & 0 \\ 0 & J(\overline{\lambda}) \end{pmatrix},$$

where $J(\lambda)$ is a Jordan block with λ on the diagonal, is similar to a product of two Hermitian matrices. This is seen as follows: matrices

$$\begin{pmatrix} J(\lambda) & 0 \\ 0 & J(\overline{\lambda}) \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} J(\lambda) & 0 \\ 0 & (J(\overline{\lambda}))^T \end{pmatrix}$$

are similar, since any square matrix is similar to its transpose. But

$$\begin{pmatrix} J(\lambda) & 0 \\ 0 & (J(\overline{\lambda}))^T \end{pmatrix} = \begin{pmatrix} J(\lambda) & 0 \\ 0 & (J(\lambda))^* \end{pmatrix},$$

which is equal to a product of two Hermitian matrices:

$$\begin{pmatrix} 0 & J(\lambda) \\ (J(\lambda))^* & 0 \end{pmatrix} \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}. \quad \blacksquare$$

Problems

1. Let A and B be Hermitian matrices of the same size. Show that $AB - BA$ is skew-Hermitian and $ABA - BAB$ is Hermitian.

2. Let A, B, and C be $n \times n$ Hermitian matrices. Prove or disprove

$$\operatorname{tr}(ABC) = \operatorname{tr}(BCA) \quad \text{or} \quad \operatorname{tr}(ABC) = \operatorname{tr}(CBA).$$

Is $\operatorname{tr}(ABC)$ necessarily real? How about $\det(ABC)$? Eigenvalues?

3. Let A and B be $n \times n$ Hermitian matrices. Show that $\operatorname{tr}(A^k B^k)$ and $\operatorname{tr}(AB)^k$ are real for any positive integer k.

4. Let A and B be n-square Hermitian matrices. Show that the product AB is Hermitian if and only if $AB = BA$. What if A, B are normal?

5. Let A and B be Hermitian matrices of the same size. Show that AB and BA are similar. What if A, B are normal?

6. If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of a Hermitian matrix A, what are the singular eigenvalues of A?

7. Give in detail the reason why the matrix A may be assumed to be a Jordan form in the proof of Theorem 8.6.

8. Let $A \in \mathbb{M}_n$. If $A^k = A^{k+1}$ for some positive integer k, show that
$$\operatorname{tr} A = \operatorname{tr} A^2 = \cdots = \operatorname{tr} A^n = \cdots.$$

9. Show that for any square complex matrices A and B of the same size
$$\operatorname{tr}(AB - BA) = 0 \quad \text{and} \quad \operatorname{tr}(AB - BA)(AB + BA) = 0.$$

10. Let A and B be Hermitian matrices of the same size. Show that
$$|\operatorname{tr}(AB)| \leq (\operatorname{tr} A^2)^{1/2} (\operatorname{tr} B^2)^{1/2} \leq \operatorname{tr}\left(\frac{A^2 + B^2}{2}\right)$$
and
$$(\operatorname{tr}(A + B)^2)^{1/2} \leq (\operatorname{tr} A^2)^{1/2} + (\operatorname{tr} B^2)^{1/2}.$$

11. Let A, B, and C be Hermitian matrices of the same size. Show that
$$|\operatorname{tr}(ABC)| \leq |\operatorname{tr}(A^2 B^2 C^2)|^{1/2}$$
is not true in general. [Hint: Assume that A is a diagonal matrix.]

12. Let A be a square matrix with all eigenvalues real (A is not necessarily Hermitian), k of which are nonzero, $k \geq 1$. Show that
$$\frac{(\operatorname{tr} A)^2}{\operatorname{tr} A^2} \leq k \leq \operatorname{rank}(A).$$

13. Let A, $B \in \mathbb{M}_n$ be Hermitian matrices of positive traces. Show that
$$\frac{\operatorname{tr}(A + B)^2}{\operatorname{tr}(A + B)} \leq \frac{\operatorname{tr} A^2}{\operatorname{tr} A} + \frac{\operatorname{tr} B^2}{\operatorname{tr} B}.$$

14. Consider two or three $n \times n$ Hermitian matrices. Prove or disprove

    (a) The determinant of a product of Hermitian matrices is real.

    (b) The trace of a product of Hermitian matrices is real.

    (c) The eigenvalues of a product of Hermitian matrices are real.

15. Let A and B be Hermitian matrices of the same size. Show that there exists a unitary matrix U such that U^*AU and U^*BU are both diagonal if and only if $AB = BA$.

16. Let A and B be Hermitian matrices of the same size. If $AB = BA$, show that for any $a, b \in \mathbb{C}$ the eigenvalues of $aA + bB$ are in the form $a\lambda + b\mu$, where λ and μ are some eigenvalues of A and B, respectively.

17. Let $A \in \mathbb{M}_n$ and S be an invertible matrix so that $S^{-1}AS = A^*$. Set

$$H_c = cS + \bar{c}S^*.$$

    Show that H_c and AH_c are Hermitian matrices. Also show that

$$A = (AH_c)H_c^{-1}$$

    is a product of two Hermitian matrices for some c such that H_c is invertible. Why does such an invertible matrix H_c exist?

18. Show that a matrix is diagonalizable (not necessarily unitarily diagonalizable) with real eigenvalues if and only if it can be written as a product of a positive definite matrix and a Hermitian matrix. For the singular case, is the product of a singular positive semidefinite matrix and a Hermitian matrix always diagonalizable?

19. Show that any square matrix is a product of two symmetric matrices.

20. Let $A \in \mathbb{M}_n$ be positive definite and let $B \in \mathbb{M}_n$ be Hermitian such that AB is a Hermitian matrix. Show that AB is positive definite if and only if the eigenvalues of B are all positive.

21. Let $A \in \mathbb{M}_n$ be a Hermitian matrix. Show that $\operatorname{tr} A > 0$ if and only if $A = B + B^*$ for some B similar to a positive definite matrix.

22. Show that a matrix A is a product of two positive semidefinite matrices if and only if A is similar to a positive semidefinite matrix.

8.3 The Min-Max Theorem and Interlacing Theorem

In this section we use some techniques on vector spaces to derive eigenvalue inequalities for Hermitian matrices. The idea is to choose vectors in certain subspaces spanned by eigenvectors in order to obtain the min-max representations that yield the desired inequalities.

Let H be an $n \times n$ Hermitian matrix with (necessarily real) eigenvalues $\lambda_i(H)$, or simply λ_i, $i = 1, 2, \dots, n$. By Theorem 8.1, there is a unitary matrix U such that

$$U^*HU = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

or

$$HU = U \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

The column vectors $u_1, u_2, \dots, u_n$ of U are orthonormal eigenvectors of H corresponding to $\lambda_1, \lambda_2, \dots, \lambda_n$, respectively, that is,

$$Hu_i = \lambda_i u_i, \quad u_i^* u_j = \delta_{ij}, \quad i, j = 1, 2, \dots, n, \tag{8.4}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise (*Kronecker delta*).

We assume that the eigenvalues and singular values of a Hermitian matrix H are arranged in decreasing order:

$$\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = \lambda_{\min};$$

$$\sigma_{\max} = \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n = \sigma_{\min}.$$

The following theorem is of fundamental importance to the rest of this chapter. The idea and result are employed frequently.

Theorem 8.7 *Let H be an $n \times n$ Hermitian matrix. Let u_1, u_2, $\dots$, u_n be orthonormal eigenvectors of H corresponding to the (not necessarily different) eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of H, respectively. Let $W = \operatorname{Span}\{u_p, \dots, u_q\}$, $1 \leq p \leq q \leq n$. Then for any unit $x \in W$*

$$\lambda_q(H) \leq x^*Hx \leq \lambda_p(H). \tag{8.5}$$

Proof. Let $x = x_p u_p + \cdots + x_q u_q$. Then by using (8.4)

$$\begin{aligned} x^*Hx &= x^*(x_p H u_p + \cdots + x_q H u_q) \\ &= x^*(\lambda_p x_p u_p + \cdots + \lambda_q x_q u_q) \\ &= \lambda_p x_p x^* u_p + \cdots + \lambda_q x_q x^* u_q \\ &= \lambda_p |x_p|^2 + \cdots + \lambda_q |x_q|^2. \end{aligned}$$

The inequality follows since x is a unit vector: $\sum_{i=p}^{q} |x_i|^2 = 1$. ∎

Theorem 8.8 (Rayleigh–Ritz) *Let $H \in \mathbb{M}_n$ be Hermitian. Then*

$$\lambda_{\min}(H) = \min_{x^*x=1} x^*Hx$$

and

$$\lambda_{\max}(H) = \max_{x^*x=1} x^*Hx.$$

Proof. The eigenvectors of H in (8.4) form an orthonormal basis for $\mathbb{C}^n$. By (8.5), it is sufficient to observe that

$$\lambda_{\min}(H) = u_n^* H u_n \quad \text{and} \quad \lambda_{\max}(H) = u_1^* H u_1. \ ∎$$

Recall the dimension identity (Theorem 1.1 of Section 1.1): If S_1 and S_2 are subspaces of an n-dimensional vector space, then

$$\dim(S_1 \cap S_2) = \dim S_1 + \dim S_2 - \dim(S_1 + S_2).$$

It follows that $S_1 \cap S_2$ is nonempty if

$$\dim S_1 + \dim S_2 > n \tag{8.6}$$

and that for three subspaces S_1, S_2, and S_3,

$$\dim(S_1 \cap S_2 \cap S_3) \geq \dim S_1 + \dim S_2 + \dim S_3 - 2n. \tag{8.7}$$

We use these inequalities to obtain the min-max theorem and derive eigenvalue inequalities for Hermitian matrices.

Theorem 8.9 (Courant–Fischer) *Let $H \in \mathbb{M}_n$ be Hermitian and let S represent a subspace of $\mathbb{C}^n$ (of complex column vectors). Then*

$$\begin{aligned}\lambda_k(H) &= \max_{(\dim S=k)} \min_{(x\in S,\, x^*x=1)} x^*Hx \\ &= \max_{(\dim S=n-k)} \min_{(x\in S^\perp,\, x^*x=1)} x^*Hx \\ &= \min_{(\dim S=n-k+1)} \max_{(x\in S,\, x^*x=1)} x^*Hx \\ &= \min_{(\dim S=k-1)} \max_{(x\in S^\perp,\, x^*x=1)} x^*Hx.\end{aligned}$$

Proof. We show the first max-min representation. The second one follows from the first immediately as $\dim S^\perp = k$ if $\dim S = n-k$. The rest of the min-max representations are proven similarly. Let u_i be orthonormal eigenvectors belonging to λ_i, $i = 1, 2\ldots, n$. We set

$$S_1 = \operatorname{Span}\{u_k, \ldots, u_n\}, \quad \dim S_1 = n-k+1,$$

and let $S_2 = S$ be any k-dimensional subspace of $\mathbb{C}^n$. By (8.6), there exists a vector x such that $x \in S_1 \cap S_2$, $x^*x = 1$, and for this x, by (8.5), $\lambda_k \geq x^*Hx$. Thus, for any k-dimensional subspace S of $\mathbb{C}^n$,

$$\lambda_k \geq \min_{v\in S,\; v^*v=1} v^*Hv.$$

It follows that

$$\lambda_k \geq \max_{(\dim S=k)} \min_{(v\in S,\; v^*v=1)} v^*Hv.$$

However, for any unit vector $v \in \operatorname{Span}\{u_1, u_2, \ldots, u_k\}$, which has dimension k, we have, by (8.5) again, $v^*Hv \geq \lambda_k$ and $u_k^*Hu_k = \lambda_k$. Thus, for $S = \operatorname{Span}\{u_1, \ldots, u_k\}$, we have

$$\min_{v\in S,\; v^*v=1} v^*Hv \geq \lambda_k.$$

It follows that

$$\max_{(\dim S=k)} \min_{(v\in S,\; v^*v=1)} v^*Hv \geq \lambda_k.$$

Putting these together,

$$\max_{(\dim S=k)} \min_{(v\in S,\, v^*v=1)} v^*Hv = \lambda_k. \quad \blacksquare$$

The following theorem is usually referred to as the *eigenvalue interlacing theorem*, also known as the *Cauchy, Poincaré,* or *Sturm interlacing theorem*. It states, simply put, that the eigenvalues of a principal submatrix of a Hermitian matrix interlace those of the underlying matrix. This is used to obtain many matrix inequalities.

Theorem 8.10 (Eigenvalue Interlacing Theorem) *Let H be an $n \times n$ Hermitian matrix partitioned as*

$$H = \begin{pmatrix} A & B \\ B^* & C \end{pmatrix},$$

where A is an $m \times m$ principal submatrix of H, $1 \le m \le n$. Then

$$\lambda_{k+n-m}(H) \le \lambda_k(A) \le \lambda_k(H), \quad k = 1, 2, \ldots, m.$$

In particular, when $m = n - 1$,

$$\lambda_n(H) \le \lambda_{n-1}(A) \le \lambda_{n-1}(H) \le \cdots \le \lambda_2(H) \le \lambda_1(A) \le \lambda_1(H).$$

We present three different proofs in the following. For convenience, we denote the eigenvalues of H and A, respectively, by

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n, \quad \mu_1 \ge \mu_2 \ge \cdots \ge \mu_m.$$

Proof 1. Use subspaces spanned by certain eigenvectors. Let u_i and v_i be orthonormal eigenvectors of H and A belonging to the eigenvalues λ_i and μ_i, respectively. Symbolically,

$$Hu_i = \lambda_i u_i, \quad u_i^* u_j = \delta_{ij}, \quad i, j = 1, 2, \ldots, n, \ u_i \in \mathbb{C}^n,$$

$$Av_i = \mu_i v_i, \quad v_i^* v_j = \delta_{ij}, \quad i, j = 1, 2, \ldots, m, \ v_i \in \mathbb{C}^m,$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Let

$$w_i = \begin{pmatrix} v_i \\ 0 \end{pmatrix} \in \mathbb{C}^n, \quad i = 1, 2, \ldots, m.$$

Note that the w_i are eigenvectors belonging to the respective eigenvalues μ_i of the partitioned matrix $A \oplus 0$. For $1 \le k \le m$, we set

$$S_1 = \operatorname{Span}\{u_k, \dots, u_n\}$$

and

$$S_2 = \operatorname{Span}\{w_1, \dots, w_k\}.$$

Then

$$\dim S_1 = n - k + 1 \quad \text{and} \quad \dim S_2 = k.$$

We thus have a vector $x \in S_1 \cap S_2$, $x^*x = 1$, and for this x, by (8.5),

$$\lambda_k \ge x^* H x \ge \mu_k. \tag{8.8}$$

An application of this inequality to $-H$ gives $\mu_k \ge \lambda_{k+n-m}$.

Proof 2. Use the adjoint matrix and continuity of functions. Reduce the proof to the case $m = n - 1$ by considering a sequence of leading principal submatrices, two consecutive ones differing in size by one.

We may assume that $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. The case in which some of the λ_i are equal follows from a continuity argument (replacing λ_i with $\lambda_i + \epsilon_i$). Let U be an n-square unitary matrix such that

$$H = U^* \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) U.$$

Then

$$tI - H = U^* \operatorname{diag}(t - \lambda_1,\ t - \lambda_2, \dots, t - \lambda_n) U \tag{8.9}$$

and for $t \ne \lambda_i,\ i = 1, 2, \dots, n$,

$$\operatorname{adj}(tI - H) = \det(tI - H)(tI - H)^{-1}. \tag{8.10}$$

Upon computation, the (n, n)-entry of $(tI - H)^{-1}$ by using (8.9) is

$$\frac{|u_{1n}|^2}{t - \lambda_1} + \frac{|u_{2n}|^2}{t - \lambda_2} + \cdots + \frac{|u_{nn}|^2}{t - \lambda_n}$$

and the (n, n)-entry of $\operatorname{adj}(tI - H)$ is $\det(tI - A)$. Thus by (8.10)

$$\frac{\det(tI - A)}{\det(tI - H)} = \frac{|u_{1n}|^2}{t - \lambda_1} + \frac{|u_{2n}|^2}{t - \lambda_2} + \cdots + \frac{|u_{nn}|^2}{t - \lambda_n}. \tag{8.11}$$

Notice that the function of t defined in (8.11) is continuous except at the points λ_i, and that it is decreasing on each interval $(\lambda_{i+1}, \lambda_i)$. On the other hand, since $\mu_1, \mu_2, \ldots, \mu_n$ are the roots of the numerator $\det(tI - A)$, by considering the behavior of the function over the intervals divided by the eigenvalues λ_i, it follows that

$$\mu_i \in [\lambda_{i+1}, \lambda_i], \quad i = 1, 2, \ldots, n-1.$$

The preceding proof is surely a good example of applications of calculus to linear algebra and matrix theory.

Proof 3. Use the Courant–Fischer theorem. Let $1 \le k \le m$. Then

$$\lambda_k(A) = \max_{S_m^k} \min_{(x \in S_m^k,\, x^*x=1)} x^*Ax,$$

where S_m^k is an arbitrary k-dimensional subspace of $\mathbb{C}^m$, and

$$\lambda_k(H) = \max_{S_n^k} \min_{(x \in S_n^k,\, x^*x=1)} x^*Hx,$$

where S_n^k is an arbitrary k-dimensional subspace of $\mathbb{C}^n$.

Denote by S_0^k the k-dimensional subspace of $\mathbb{C}^n$ of the vectors

$$y = \begin{pmatrix} x \\ 0 \end{pmatrix}, \quad \text{where } x \in S_m^k.$$

Noticing that $y^*Hy = x^*Ax$, we have, by a simple computation,

$$\begin{aligned} \lambda_k(H) &= \max_{S_n^k} \min_{(x \in S_n^k,\, x^*x=1)} x^*Hx \\ &\ge \max_{S_0^k} \min_{(y \in S_0^k,\, y^*y=1)} y^*Hy \\ &= \max_{S_m^k} \min_{(x \in S_m^k,\, x^*x=1)} x^*Ax \\ &= \lambda_k(A). \end{aligned}$$

The other inequality is obtained by replacing H with $-H$. ∎

As an application of the interlacing theorem, we present a result due to Poincaré: if $A \in \mathbb{M}_m$ is a Hermitian matrix, then for any $m \times n$ matrix V satisfying $V^*V = I_n$ and for each $i = 1, 2, \ldots, m$,

$$\lambda_{i+m-n}(A) \le \lambda_i(V^*AV) \le \lambda_i(A). \tag{8.12}$$

To see this, first notice that $V^*V = I_n \Rightarrow m \geq n$ (Problem 1). Let U be an $m \times (m-n)$ matrix such that (V, U) is unitary. Then

$$(V,U)^*A(V,U) = \begin{pmatrix} V^*AV & V^*AU \\ U^*AV & U^*AU \end{pmatrix}.$$

Thus by applying the interlacing theorem, we have

$$\begin{aligned} \lambda_{i+m-n}(A) &= \lambda_{i+m-n}\Big((V,U)^*A(V,U)\Big) \\ &\leq \lambda_i(V^*AV) \\ &\leq \lambda_i\Big((V,U)^*A(V,U)\Big) \\ &= \lambda_i(A). \end{aligned}$$

Problems

1. Let V be an $m \times n$ matrix. Show that if $V^*V = I_n$, then $m \geq n$.

2. Let $[A]$ be a principal submatrix of A. If A is Hermitian, show that

$$\lambda_{\min}(A) \leq \lambda_{\min}([A]) \leq \lambda_{\max}([A]) \leq \lambda_{\max}(A).$$

3. Let $\lambda_k(A)$ denote the kth largest eigenvalue of an n-square positive definite matrix A. Show that

$$\lambda_k(A^{-1}) = \frac{1}{\lambda_{n-k+1}(A)}.$$

4. Let $A \in \mathbb{M}_n$ be Hermitian. Show that for every nonzero $x \in \mathbb{C}^n$,

$$\lambda_{\min}(A) \leq \frac{x^*Ax}{x^*x} \leq \lambda_{\max}(A)$$

and for all diagonal entries a_{ii} of A,

$$\lambda_{\min}(A) \leq a_{ii} \leq \lambda_{\max}(A).$$

5. For any Hermitian matrices A and B of the same size, show that

$$\lambda_{\max}(A - B) + \lambda_{\min}(B) \leq \lambda_{\max}(A).$$

6. Let $A \in \mathbb{M}_n$ be Hermitian and $B \in \mathbb{M}_n$ be positive definite. Show that the eigenvalues of AB^{-1} are all real, that

$$\lambda_{\max}(AB^{-1}) = \max_{x \neq 0} \frac{x^*Ax}{x^*Bx},$$

and that

$$\lambda_{\min}(AB^{-1}) = \min_{x \neq 0} \frac{x^*Ax}{x^*Bx}.$$

7. Let A be an $n \times n$ Hermitian matrix and X be an $n \times p$ matrix such that $X^*X = I_p$. Then

$$\sum_{i=n-p+1}^{n} \lambda_i(A) \leq \operatorname{tr}(X^*AX) \leq \sum_{i=1}^{p} \lambda_i(A)$$

and

$$\sum_{i=1}^{p} \lambda_i^{-1}(A) \leq \operatorname{tr}(X^*AX)^{-1} \leq \sum_{i=n-p+1}^{n} \lambda_i^{-1}(A).$$

8. Let A be an n-square positive semidefinite matrix. If V is an $n \times m$ complex matrix such that $V^*V = I_m$, show that

$$\prod_{i=1}^{m} \lambda_{n-m+i}(A) \leq \det(V^*AV) \leq \prod_{i=1}^{m} \lambda_i(A).$$

9. Let A be a positive semidefinite matrix partitioned as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where A_{11} is square, and let $\widetilde{A_{11}} = A_{22} - A_{21}A_{11}^{-1}A_{12}$ be the Schur complement of A_{11} in A when A_{11} is nonsingular. Show that

$$\lambda_{\min}(A) \leq \lambda_{\min}(\widetilde{A_{11}}) \leq \lambda_{\min}(A_{22}).$$

10. Let $H = \left(\begin{smallmatrix} X & Z \\ Z^* & Y \end{smallmatrix}\right)$ be Hermitian, where X is the $k \times k$ leading principal submatrix of H. If $\lambda_i(H) = \lambda_i(X)$, $i = 1, 2, \ldots, k$, show that $Z = 0$.

11. Let A be an n-square matrix and U be an $n \times k$ matrix, $k \leq n$, such that $U^*U = I_k$. Show that $\sigma_i(U^*AU) \leq \sigma_i(A)$, $i = 1, 2, \ldots, k$.

12. Show the min-max representations in Theorem 8.9.

8.4 Eigenvalue and Singular Value Inequalities

This section presents some basic eigenvalue and singular value inequalities by using the min-max representation theorem and the eigenvalue interlacing theorem. We assume that the eigenvalues $\lambda_i(A)$, singular values $\sigma_i(A)$, and diagonal entries $d_i(A)$ of a Hermitian matrix A are arranged in decreasing order.

The following theorem on comparing two Hermitian matrices best characterizes the *Löwner ordering* in terms of eigenvalues.

Theorem 8.11 *Let A, $B \in \mathbb{M}_n$ be Hermitian matrices. Then*

$$A \geq B \quad \Rightarrow \quad \lambda_i(A) \geq \lambda_i(B), \;\; i = 1, 2, \ldots, n.$$

This follows from the Courant–Fischer theorem immediately, for

$$A \geq B \quad \Rightarrow \quad x^*Ax \geq x^*Bx, \;\; x \in \mathbb{C}^n.$$

Our next theorem compares the eigenvalues of sum, ordinary, and Hadamard products of matrices to those of the individual matrices.

Theorem 8.12 *Let A, $B \in \mathbb{M}_n$ be Hermitian matrices. Then*

1. $\lambda_i(A) + \lambda_n(B) \leq \lambda_i(A+B) \leq \lambda_i(A) + \lambda_1(B)$.
2. $\lambda_i(A)\lambda_n(B) \leq \lambda_i(AB) \leq \lambda_i(A)\lambda_1(B)$ *if* $A \geq 0$ *and* $B \geq 0$.
3. $d_i(A)\lambda_n(B) \leq \lambda_i(A \circ B) \leq d_i(A)\lambda_1(B)$ *if* $A \geq 0$ *and* $B \geq 0$.

Proof. Let x be unit vectors in $\mathbb{C}^n$. Then we have

$$x^*Ax + \min_x x^*Bx \leq x^*(A+B)x \leq x^*Ax + \max_x x^*Bx.$$

Thus

$$x^*Ax + \lambda_n(B) \leq x^*(A+B)x \leq x^*Ax + \lambda_1(B).$$

An application of the min-max theorem results in (1). For (2), we write $\lambda_i(AB) = \lambda_i(B^{1/2}AB^{1/2})$. Notice that $\lambda_1(A)I - A \geq 0$. Thus

$$B^{1/2}AB^{1/2} \leq B^{1/2}AB^{1/2} + B^{1/2}(\lambda_1(A)I - A)B^{1/2} = \lambda_1(A)B.$$

An application of Theorem 8.11 gives (2). For (3), recall that the Hadamard product of two positive semidefinite matrices is positive semidefinite (Schur theorem, Section 7.5). Since $\lambda_1(B)I - B \geq 0$ and $B - \lambda_n(B)I \geq 0$, by taking the Hadamard product with A, we have

$$A \circ \Big(\lambda_1(B)I - B\Big) \geq 0, \quad A \circ \Big(B - \lambda_n(B)I\Big) \geq 0,$$

which reveals

$$\lambda_n(B)(I \circ A) \leq A \circ B \leq \lambda_1(B)(I \circ A).$$

Note that $I \circ A = \operatorname{diag}(a_{11}, \ldots, a_{nn})$. Theorem 8.11 gives for each i,

$$d_i(A)\lambda_n(B) \leq \lambda_i(A \circ B) \leq d_i(A)\lambda_1(B). \quad \blacksquare$$

It is natural to ask the question: if $A \geq 0$ and $B \geq 0$, is

$$\lambda_i(A)\lambda_n(B) \leq \lambda_i(A \circ B) \leq \lambda_i(A)\lambda_1(B)?$$

The answer is negative (Problem 8). For singular values, we have

Theorem 8.13 *Let A and B be complex matrices. Then*

$$\sigma_i(A) + \sigma_n(B) \leq \sigma_i(A + B) \leq \sigma_i(A) + \sigma_1(B) \tag{8.13}$$

if A and B are of the same size $m \times n$, and

$$\sigma_i(A)\sigma_m(B) \leq \sigma_i(AB) \leq \sigma_i(A)\sigma_1(B) \tag{8.14}$$

if A is $m \times n$ and B is $n \times m$.

Proof. To show (8.13), notice that the Hermitian matrix

$$\begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix},$$

where X is an $m \times n$ matrix with rank r, has eigenvalues

$$\sigma_1(X), \ \ldots, \ \sigma_r(X), \ \overbrace{0, \ldots, 0}^{m+n-2r}, \ -\sigma_r(X), \ \ldots, \ -\sigma_1(X).$$

Applying the previous theorem to the Hermitian matrices

$$\begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}$$

and to their sum, one gets the desired singular value inequalities.

For the inequalities on product, it suffices to note that

$$\sigma_i(AB) = \sqrt{\lambda_i(B^*A^*AB)} = \sqrt{\lambda_i(A^*ABB^*)}. \quad \blacksquare$$

Some stronger results can be obtained by using the min-max theorem (Problem 18). For submatrices of a general matrix, we have the following inequalities on the singular values.

Theorem 8.14 *Let B be any submatrix of an $m \times n$ matrix A obtained by deleting s rows and t columns, $s + t = r$. Then*

$$\sigma_{r+i}(A) \le \sigma_i(B) \le \sigma_i(A), \quad i = 1, 2, \dots, \min\{m, n\}.$$

Proof. We may assume that $r = 1$ and B is obtained from A by deleting a column, say b; that is, $A = (B, b)$. Otherwise one may place B in the upper-left corner of A by permutation and consider a sequence of submatrices of A that contain B, two consecutive ones differing by a row or column (see the second proof of Theorem 8.10).

Notice that B^*B is a principal submatrix of A^*A. Using the eigenvalue interlacing theorem (Theorem 8.10), we have for each i,

$$\lambda_{i+1}(A^*A) \le \lambda_i(B^*B) \le \lambda_i(A^*A).$$

The proof is completed by taking square roots. $\blacksquare$

Theorem 8.12 can be generalized to eigenvalues with indices r and s, $r+s \le n-1$. As an example, we present two such inequalities, one for the sum of Hermitian matrices and one for the product of positive semidefinite matrices, and leave others to the reader (Problem 18).

Theorem 8.15 *Let A and B be n-square Hermitian matrices. If r and s are nonnegative integers such that $r + s \le n - 1$, then*

$$\lambda_{r+s+1}(A + B) \le \lambda_{r+1}(A) + \lambda_{s+1}(B).$$

Proof. Let u_i, v_i, and w_i, $i = 1, 2, \ldots, n$, be orthonormal eigenvectors corresponding to the eigenvalues $\lambda_i(A)$, $\lambda_i(B)$, and $\lambda_i(A+B)$, of the matrices A, B, and $A+B$, respectively. Let

$$\begin{aligned} S_1 &= \operatorname{Span}\{u_{r+1}, \ldots, \ldots, u_n\}, & \dim S_1 &= n - r, \\ S_2 &= \operatorname{Span}\{v_{s+1}, \ldots, \ldots, s_n\}, & \dim S_2 &= n - s, \\ S_3 &= \operatorname{Span}\{w_1, \ldots, \ldots, w_{r+s+1}\}, & \dim S_3 &= r + s + 1. \end{aligned}$$

Then, by Problem 23 of Section 1.1,

$$\dim(S_1 \cap S_2 \cap S_3) \geq \dim S_1 + \dim S_2 + \dim S_3 - 2n = 1.$$

Thus, there exists a nonzero unit vector $x \in S_1 \cap S_2 \cap S_3$. By (8.5),

$$\lambda_{r+s+1}(A+B) \leq x^*(A+B)x = x^*Ax + x^*Bx \leq \lambda_{r+1}(A) + \lambda_{s+1}(B). \quad \blacksquare$$

Setting $r = 0$ in the theorem, we have for any $1 \leq k \leq n$,

$$\lambda_k(A+B) \leq \lambda_1(A) + \lambda_k(B).$$

Applying the theorem to $-A$ and $-B$, one gets

$$\lambda_{n-r-s}(A+B) \geq \lambda_{n-r}(A) + \lambda_{n-s}(B).$$

Theorem 8.16 *Let G and H be $n \times n$ positive semidefinite matrices. If r and s are nonnegative integers such that $r + s \leq n - 1$, then*

$$\lambda_{r+s+1}(GH) \leq \lambda_{r+1}(G)\lambda_{s+1}(H).$$

Proof. We may assume that both G and H are positive definite; otherwise use a continuity argument. Let u_i and v_i be the orthonormal eigenvectors corresponding to $\lambda_i(G)$ and $\lambda_i(H)$, respectively, $i = 1, \ldots, n$. Let W_1 be the subspace of $\mathbb{C}^n$ spanned by $u_1, \ldots, u_r$ and W_2 be the subspace spanned by $v_1, \ldots, v_s$.

Let $W_3 = \operatorname{Span}\{H^{-1/2}u_1, \ldots, H^{-1/2}u_r\}$. Then $\dim(W_2 + W_3) \leq r + s$. Let S be a subspace of dimension $r + s$ containing $W_2 + W_3$. Then $S^\perp \subseteq (W_2 + W_3)^\perp = W_2^\perp \cap W_3^\perp$. It follows that

$$\begin{aligned}
\lambda_{r+s+1}(GH) &= \lambda_{r+s+1}(H^{1/2}GH^{1/2}) \\
&= \min_{(\dim W=r+s)} \max_{(x\in W^\perp,\, x^*x=1)} x^*(H^{1/2}GH^{1/2})x \\
&\le \max_{x\in S^\perp,\, x^*x=1} x^*(H^{1/2}GH^{1/2})x \\
&\le \max_{x\in W_2^\perp\cap W_3^\perp,\, x^*x=1} \frac{x^*(H^{1/2}GH^{1/2})x}{x^*Hx}\cdot x^*Hx \\
&\le \max_{x\in W_3^\perp,\, x^*x=1} \frac{x^*(H^{1/2}GH^{1/2})x}{x^*Hx}\cdot \max_{x\in W_2^\perp,\, x^*x=1} x^*Hx \\
&\le \max_{y\in W_1^\perp} \frac{y^*Gy}{y^*y}\cdot \max_{x\in W_2^\perp,\, x^*x=1} x^*Hx \\
&= \lambda_{r+1}(G)\lambda_{s+1}(H). \quad \blacksquare
\end{aligned}$$

Problems

1. Use Theorem 8.11 to show that $A \ge B \ge 0$ implies

$$\operatorname{rank}(A) \ge \operatorname{rank}(B), \quad \det A \ge \det B.$$

2. Let $A \ge 0$, $B \ge 0$. Prove or disprove $\lambda_i(A) \ge \lambda_i(B) \Rightarrow A \ge B$.

3. Show that $A > B \;\Rightarrow\; \lambda_i(A) > \lambda_i(B)$ for each i.

4. Let $A \ge B \ge 0$. Show that for $X \ge 0$ of the same size as A and B

$$\lambda_i(AX) \ge \lambda_i(BX), \quad \text{for each } i.$$

5. Let $A \in \mathbb{M}_n$ be a positive semidefinite matrix. Show that

$$\lambda_1(A)I - A \ge 0 \ge \lambda_n(A)I - A, \;\text{ i.e., }\; \lambda_n(A)I \le A \le \lambda_1(A)I.$$

6. Let A and B be $n \times n$ positive semidefinite matrices. Show that

$$\lambda_n(A) + \lambda_n(B) \le \lambda_n(A+B) \le \lambda_1(A+B) \le \lambda_1(A) + \lambda_1(B);$$
$$\lambda_n(A)\lambda_n(B) \le \lambda_n(AB) \le \lambda_1(AB) \le \lambda_1(A)\lambda_1(B);$$
$$\lambda_n(A)\lambda_n(B) \le \lambda_n(A\circ B) \le \lambda_1(A\circ B) \le \lambda_1(A)\lambda_1(B).$$

7. Show by example that for $A \geq 0$ and $B \geq 0$ of the same size the following inequalities may both occur.

$$\lambda_n(AB) > \lambda_n(A) \min_i b_{ii}, \quad \lambda_n(AB) < \lambda_n(A) \min_i b_{ii}.$$

8. Show by example that for $A \geq 0$ and $B \geq 0$ of the same size the following inequalities do not hold in general.

$$\lambda_i(A)\lambda_n(B) \leq \lambda_i(A \circ B) \leq \lambda_i(A)\lambda_1(B).$$

9. Let A, $B \in \mathbb{M}_n$ be Hermitian matrices. Prove or disprove

$$\lambda_i(A + B) \leq \lambda_i(A) + \lambda_i(B).$$

10. Let A, $B \in \mathbb{M}_n$ be Hermitian matrices. Show that for any $\alpha \in [0, 1]$,

$$\lambda_{\min}(\alpha A + (1 - \alpha)B) \geq \alpha\lambda_{\min}(A) + (1 - \alpha)\lambda_{\min}(B)$$

and

$$\lambda_{\max}(\alpha A + (1 - \alpha)B) \leq \alpha\lambda_{\max}(A) + (1 - \alpha)\lambda_{\max}(B).$$

11. Let A be Hermitian and $B \geq 0$ be of the same size. Show that

$$\lambda_{\min}(B)\lambda_i(A^2) \leq \lambda_i(ABA) \leq \lambda_{\max}(B)\lambda_i(A^2).$$

12. Let $A \in \mathbb{M}_n$ be Hermitian. Show that for any row vector $u \in \mathbb{C}^n$,

$$\lambda_{i+2}(A + u^*u) \leq \lambda_{i+1}(A) \leq \lambda_i(A + u^*u)$$

and

$$\lambda_{i+2}(A) \leq \lambda_{i+1}(A + u^*u) \leq \lambda_i(A).$$

13. When does $d(A) = \lambda(A)$? In other words, for what matrices are the diagonal entries equal to the eigenvalues? How about

$$d(A) = \sigma(A) \quad \text{or} \quad \lambda(A) = \sigma(A)?$$

14. Let A be an $n \times n$ positive semidefinite matrix. If V is an $n \times m$ matrix so that $V^*V = \operatorname{diag}(\delta_1, \delta_2, \ldots, \delta_m)$, each $\delta_i > 0$, show that

$$\lambda_n(A) \min_i \delta_i \leq \lambda_i(V^*AV) \leq \lambda_1(A) \max_i \delta_i.$$

15. Let A and B be n-square complex matrices. Show that

$$\sigma_n(A)\sigma_n(B) \leq |\lambda(AB)| \leq \sigma_1(A)\sigma_1(B)$$

for any eigenvalue $\lambda(AB)$ of AB. In particular, $|\lambda(A)| \leq \sigma_{\max}(A)$.

16. Let A, X, B be $m \times p$, $p \times q$, $q \times n$ matrices, respectively. Show that

$$\sigma_i(AXB) \leq \sigma_1(A)\sigma_i(X)\sigma_1(B) \text{ for every } i \leq \min\{m, p, q, n\}.$$

17. Let X and Y be $n \times n$ matrices, $t \in [0,1]$, and $\tilde{t} = 1 - t$. Show that

$$\sigma_i(tX + \tilde{t}Y) \leq \sigma_i(X \oplus Y).$$

[Hint: Consider $(\alpha I, \beta I)(X \oplus Y)(\alpha I, \beta I)^T$ and use Problem 16.]

18. Let A and B be $n \times n$ matrices and $r + s \leq n - 1$. Show the following.

(a) If A and B are Hermitian, then

$$\lambda_{n-r-s}(A + B) \geq \lambda_{n-r}(A) + \lambda_{n-s}(B).$$

(b) If A and B are positive semidefinite, then

$$\lambda_{n-r-s}(AB) \geq \lambda_{n-r}(A)\lambda_{n-s}(B).$$

(c) For singular values,

$$\sigma_{r+s+1}(A + B) \leq \sigma_{r+1}(A) + \sigma_{s+1}(B),$$

$$\sigma_{r+s+1}(AB) \leq \sigma_{r+1}(A)\sigma_{s+1}(B),$$

and

$$\sigma_{n-r-s}(AB) \geq \sigma_{n-r}(A)\sigma_{n-s}(B),$$

but it is false that

$$\sigma_{n-r-s}(A + B) \geq \sigma_{n-r}(A) + \sigma_{n-s}(B).$$

8.5 Eigenvalues of Hermitian Matrices A, B, and $A+B$

This section is devoted to the relationship between the eigenvalues of Hermitian matrices A and B and those of the sum $A+B$. It has been evident that min-max representations play an important role in the study. We start with a simple, but neat result of min-max type.

Theorem 8.17 (Fan) *Let H be an $n \times n$ Hermitian matrix. Denote by S_k a set of any k orthonormal vectors $x_1, x_2, \ldots, x_k \in \mathbb{C}^n$. Then*

$$\sum_{i=1}^{k} \lambda_i(H) = \max_{S_k} \sum_{i=1}^{k} x_i^* H x_i, \quad 1 \le k \le n. \tag{8.15}$$

Proof. Let $U = (V, W)$ be a unitary matrix, where V consists of the orthonormal vectors $x_1, x_2, \ldots, x_k$. Then by (8.12)

$$\sum_{i=1}^{k} x_i^* H x_i = \operatorname{tr}(V^* H V) = \sum_{i=1}^{k} \lambda_i(V^* H V) \le \sum_{i=1}^{k} \lambda_i(H).$$

Identity (8.15) follows by choosing the unit eigenvectors x_i of $\lambda_i(H)$:

$$\sum_{i=1}^{k} x_i^* H x_i = \sum_{i=1}^{k} \lambda_i(H). \quad \blacksquare$$

Our main result of this section is the following theorem, which results in a number of majorization inequalities (Chapter 10).

Theorem 8.18 (Thompson) *Let A and B be $n \times n$ Hermitian matrices and let $C = A + B$. If $\alpha_1 \ge \cdots \ge \alpha_n$, $\beta_1 \ge \cdots \ge \beta_n$, and $\gamma_1 \ge \cdots \ge \gamma_n$ are the eigenvalues of A, B, and C, respectively, then for any sequence $1 \le i_1 < \cdots < i_k \le n$,*

$$\sum_{t=1}^{k} \alpha_{i_t} + \sum_{t=1}^{k} \beta_{n-k+t} \le \sum_{t=1}^{k} \gamma_{i_t} \le \sum_{t=1}^{k} \alpha_{i_t} + \sum_{t=1}^{k} \beta_t. \tag{8.16}$$

This theorem is proved by using the following min-max expression for the sum of eigenvalues that is in turn shown via a few lemmas.

Theorem 8.19 *Let H be an $n \times n$ Hermitian matrix and let W_t represent a subspace of $\mathbb{C}^n$. Then, for $1 \le i_1 < \cdots < i_k \le n$,*

$$\sum_{t=1}^{k} \lambda_{i_t}(H) = \max_{\dim W_t = i_t} \min_{x_t \in W_t} \sum_{t=1}^{k} x_t^* H x_t, \tag{8.17}$$

where $x_1, \ldots, x_k$ (in $W_1, \ldots, W_k$, respectively) are orthonormal.

Lemma 8.1 *Let $U_1, \ldots, U_k$ be subspaces of an inner product space such that $\dim U_t \ge t$, $t = 1, \ldots, k$. If $u_2, \ldots, u_k$ are linearly independent vectors in $U_2, \ldots, U_k$, respectively, then there exist linearly independent vectors $y_1, y_2, \ldots, y_k$ in $U_1, U_2, \ldots, U_k$, respectively, such that $\operatorname{Span}\{u_2, \ldots, u_k\} \subseteq \operatorname{Span}\{y_1, y_2, \ldots, y_k\}$. If, in addition, $U_1 \subset \cdots \subset U_k$, then $y_1, y_2, \ldots, y_k$ can be chosen to be orthonormal.*

Proof. We use induction on k. Let $k = 2$ and $u_2 \in U_2$ be nonzero. If $u_2 \in U_1$, then we set $y_1 = u_2$ and take y_2 from U_2 so that y_2 is not a multiple of u_2 (i.e., y_1 and y_2 are linearly independent). This is possible because $\dim U_2 \ge 2$. If $u_2 \notin U_1$, there must exist a nonzero vector $y_1 \in U_1$ that is not a multiple of u_2. Now set $y_2 = u_2$. y_1 and y_2 are linearly independent. In either case, $\operatorname{Span}\{u_2\} \subseteq \operatorname{Span}\{y_1, y_2\}$.

Now suppose it is true for the case of $k - 1$; that is, given $u_2, \ldots, u_{k-1}$, there exist linearly independent $y_1, y_2, \ldots, y_{k-1}$ so that the span of the us is contained in the span of the ys. We show the case of k. If $u_k \notin \operatorname{Span}\{y_1, y_2, \ldots, y_{k-1}\}$, then we take $y_k = u_k$. Otherwise, we take a nonzero $y_k \in U_k \cap (\operatorname{Span}\{y_1, y_2, \ldots, y_{k-1}\})^{\perp}$. In any case, we have $\operatorname{Span}\{u_2, \ldots, u_k\} \subseteq \operatorname{Span}\{y_1, y_2, \ldots, y_k\}$, where $y_t \in U_t$ are linearly independent, $t = 1, \ldots, k$. Due to the Gram–Schmidt orthonormalization (Problem 21, Section 1.4), $y_1, y_2, \ldots, y_k$ can be chosen to be orthonormal. ∎

By putting $Q_t = U_{k-t+1}$ for each t in the above lemma, we obtain an equivalent statement to the lemma: if Q_t are k subspaces of an inner product space, $\dim Q_t \ge k - t + 1$, and if v_t are linearly independent vectors, where $v_t \in Q_t$, $t = 1, \ldots, k - 1$, there exist linearly independent vectors y_t, where $y_t \in Q_t$, $t = 1, \ldots, k$, such that $\operatorname{Span}\{v_1, \ldots, v_{k-1}\} \subseteq \operatorname{Span}\{y_1, y_2, \ldots, y_k\}$.

Lemma 8.2 *Let* $1 \le i_1 < \cdots < i_k \le n$. *Let* $W_1, \dots, W_k$ *and* $V_1, \dots, V_k$ *be subspaces of an inner product space such that* $\dim W_t \ge i_t$, $\dim V_t \ge n - i_t + 1$, $t = 1, \dots, k$, *and* $V_1 \supseteq \cdots \supseteq V_k$. *Then there exist orthonormal* $x_t \in W_t$, $t = 1, \dots, k$, *and orthonormal* $y_t \in V_t$, $t = 1, \dots, k$, *such that* $\operatorname{Span}\{x_1, \dots, x_k\} = \operatorname{Span}\{y_1, \dots, y_k\}$.

Proof. Since $\dim W_1 + \dim V_1 = n + 1$, by the dimension identity (Theorem 1.1), $W_1 \cap V_1 \neq \emptyset$. A unit vector $x_1 = y_1 \in W_1 \cap V_1$ exists. So the conclusion is true when $k = 1$. Suppose it is true for $k - 1$; that is, there exist orthonormal $x_t \in W_t$ and $v_t \in V_t$, $t = 1, \dots, k-1$, such that $\operatorname{Span}\{x_1, \dots, x_{k-1}\} = \operatorname{Span}\{v_1, \dots, v_{k-1}\}$. Now let

$$W = \operatorname{Span}\{x_1, \dots, x_{k-1}\} + W_k \cap (\operatorname{Span}\{x_1, \dots, x_{k-1}\})^\perp. \quad (8.18)$$

Then $\dim W \ge \dim W_k \ge i_k$ (Problem 3). Note that $i_k - i_t \ge k - t$; we have $\dim W + \dim V_t \ge i_k + n - i_t + 1 \ge n + k - t + 1$.

Set $Q_t = W \cap V_t$, $t = 1, \dots, k$. By the dimension identity again, $\dim Q_t \ge k - t + 1$. Moreover, $Q_1 \supseteq \cdots \supseteq Q_k$, and $v_t \in Q_t$, $t = 1, \dots, k-1$. By the discussion following Lemma 8.1, there exist orthonormal $y_t \in Q_t$, $t = 1, \dots, k$, such that $\operatorname{Span}\{x_1, \dots, x_{k-1}\} = \operatorname{Span}\{v_1, \dots, v_{k-1}\} \subseteq \operatorname{Span}\{y_1, \dots, y_k\} \subseteq W$. Choose a unit vector x_k from the space $(\operatorname{Span}\{x_1, \dots, x_{k-1}\})^\perp \cap \operatorname{Span}\{y_1, \dots, y_k\} \neq \emptyset$. By (8.18), $x_k \in W_k$ and $\operatorname{Span}\{x_1, \dots, x_k\} = \operatorname{Span}\{y_1, \dots, y_k\}$. ∎

Lemma 8.3 *Let* H *be an* $n \times n$ *Hermitian matrix and let* $1 \le i_1 < \cdots < i_k \le n$. *Then for any subspaces* $W_1, \dots, W_k$ *of* $\mathbb{C}^n$ *such that* $\dim W_t = i_t$, $t = 1, \dots, k$, *there exist orthonormal* $x_t \in W_t$, $t = 1, \dots, k$, *such that*

$$\sum_{t=1}^{k} x_t^* H x_t \le \sum_{t=1}^{k} \lambda_{i_t}(H).$$

Proof. Let $V_t = \operatorname{Span}\{u_{i_t}, u_{i_t+1}, \dots, u_n\}$, where u_j are the orthonormal eigenvectors of H belonging to the eigenvalues λ_j, respectively. Then $\dim V_t = n - i_t + 1$, $t = 1, \dots, k$, and $V_1 \supseteq \cdots \supseteq V_k$. By Lemma 8.2, there exist orthonormal $x_1, \dots, x_k$ and orthonormal $y_1, \dots, y_k$, where $x_t \in W_t$ and $y_t \in V_t$, $t = 1, \dots, k$, such that $\operatorname{Span}\{x_1, \ldots, x_k\} = \operatorname{Span}\{y_1, \ldots, y_k\}$. By Problem 4 and Theorem 8.7, we have

$$\sum_{t=1}^{k} x_t^* H x_t = \sum_{t=1}^{k} y_t^* H y_t \leq \sum_{t=1}^{k} \lambda_{i_t}(H). \quad \blacksquare$$

Now we are ready to prove Theorem 8.19.

Proof of Theorem 8.19. Let $W_t = \operatorname{Span}\{u_1, u_2, \ldots, u_{i_t}\}$, where u_j are the orthonormal eigenvectors of H belonging to the eigenvalues λ_j, respectively. (This is possible because H is Hermitian.) Then $\dim W_t = i_t$, $t = 1, \ldots, k$. For any unit vectors $x_t \in W_t$, we have

$$\sum_{t=1}^{k} x_t^* H x_t \geq \sum_{t=1}^{k} \lambda_{i_t}(H).$$

Combining Lemma 8.3, we accomplish the proof. $\blacksquare$

Proof of Theorem 8.18. By Theorem 8.19, we may assume that $\sum_{t=1}^{k} \gamma_{i_t}$ is attained by the subspaces S_t, $\dim S_t = i_t$, $t = 1, \ldots, k$, with orthonormal $y_t \in S_t$, $t = 1, \ldots, k$. Then Lemma 8.3 ensures

$$\sum_{t=1}^{k} y_t^* A y_t \leq \sum_{t=1}^{k} \lambda_{i_t}(A).$$

However, Theorem 8.17 yields $\sum_{t=1}^{k} y_t^* B y_t \leq \sum_{t=1}^{k} \lambda_t(B)$. Therefore,

$$\sum_{t=1}^{k} \gamma_{i_t} = \sum_{t=1}^{k} y_t^* (A+B) y_t \leq \sum_{t=1}^{k} \lambda_{i_t}(A) + \sum_{t=1}^{k} \lambda_t(B).$$

This is the inequality on the right-hand side of (8.16). For the first inequality, let $D = -B$ and denote the eigenvalues of D by $\delta_1 \geq \cdots \geq \delta_n$. Then $A = C + D$. The inequality we just proved reveals

$$\sum_{t=1}^{k} \alpha_{i_t} \leq \sum_{t=1}^{k} \gamma_{i_t} + \sum_{t=1}^{k} \delta_t.$$

Note that $\delta_t = -\beta_{n-t+1}$, $t = 1, \ldots, n$. The inequality follows. $\blacksquare$

Problems

1. Let $A \in \mathbb{M}_n$ be a Hermitian matrix with eigenvalues ordered $|\lambda_1(A)| \geq \cdots \geq |\lambda_n(A)|$. Show that $|\lambda_1(A)| = \max_{x^*x=1} |x^*Ax|$. However, $|\lambda_n(A)| = \min_{x^*x=1} |x^*Ax|$ is not true in general.

2. Let A be an $m \times n$ matrix. For any index i, show that
$$\begin{aligned} \sigma_i(A) &= \max_{(\dim W=i)} \min_{(x\in W,\, x^*x=1)} x^*(A^*A)^{1/2}x \\ &= \max_{(\dim W=i)} \min_{(x\in W,\, x^*x=1)} (x^*A^*Ax)^{1/2}. \end{aligned}$$

3. Let V and S be subspaces of an inner product space of finite dimension. Let $W = S + S^{\perp} \cap V$. Show that $\dim W \geq \dim V$.

4. Let $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_m\}$ be orthonormal sets in $\mathbb{C}^n$ so that
$$\mathrm{Span}\{x_1, \ldots, x_m\} = \mathrm{Span}\{y_1, \ldots, y_m\}.$$
Show that for any n-square complex matrix A,
$$\sum_{t=1}^{m} x_t^* A x_t = \sum_{t=1}^{m} y_t^* A y_t.$$

5. Let H be an $n \times n$ Hermitian matrix. For any $1 \leq k \leq n$, show that
$$\sum_{i=1}^{k} \lambda_{n-i+1}(H) = \min_{U^*U=I_k} \mathrm{tr}\, U^*HU = \min_{x_i^*x_j=\delta_{ij}} \sum_{i=1}^{k} x_i^* H x_i,$$
where $\delta_{ij} = 1$ if $i = j$ and 0 if $i \neq j$.

6. Let $H \in \mathbb{M}_n$ be positive semidefinite. For any $1 \leq k \leq n$, show that
$$\prod_{i=1}^{k} \lambda_i(H) = \max_{U^*U=I_k} \det U^*HU = \max_{x_i^*x_j=\delta_{ij}} \det(x_i^* H x_j),$$
where $\delta_{ij} = 1$ if $i = j$ and 0 if $i \neq j$, and
$$\prod_{i=1}^{k} \lambda_{n-i+1}(H) = \min_{U^*U=I_k} \det U^*HU = \min_{x_i^*x_j=\delta_{ij}} \det(x_i^* H x_j).$$

7. Let A and B be $n \times n$ Hermitian matrices and $B > 0$. Let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$ be the eigenvalues of $\det(A - \mu B) = 0$. Show that
$$\mu_1 = \max_{x\neq 0} \frac{x^*Ax}{x^*Bx}, \quad \mu_n = \min_{x\neq 0} \frac{x^*Ax}{x^*Bx}.$$

Show more generally that for $1 \le k \le n$,

$$\sum_{i=1}^{k} \mu_i = \max_{x_i^* B x_j = 0,\, i \neq j} \sum_{i=1}^{k} \frac{x_i^* A x_i}{x_i^* B x_i}$$

and

$$\prod_{i=n-k+1}^{n} \mu_i = \min_{x_i^* B x_j = 0,\, i \neq j} \prod_{i=1}^{k} \frac{x_i^* A x_i}{x_i^* B x_i}.$$

8. Let A be an $m \times n$ complex matrix. For any $1 \le k \le n$, show that

$$\prod_{i=1}^{k} \sigma_i(A) = \max_{U^*U = I_k} (\det(U^* A^* A U))^{1/2}$$

and

$$\prod_{i=1}^{k} \sigma_{n-i+1}(A) = \min_{U^*U = I_k} (\det(U^* A^* A U))^{1/2}.$$

9. Let A be an $m \times n$ complex matrix. For any $1 \le k \le n$, show that

$$\sum_{i=1}^{k} \sigma_i(A) = \max_{U^*U = V^*V = I_k} |\operatorname{tr}(UAV)| = \max_{U^*U = V^*V = I_k} \operatorname{Re}(\operatorname{tr}(UAV)).$$

However, neither of the following holds:

$$\sum_{i=1}^{k} \sigma_{n-i+1}(A) = \min_{U^*U = V^*V = I_k} |\operatorname{tr}(UAV)|;$$

$$\sum_{i=1}^{k} \sigma_{n-i+1}(A) = \min_{U^*U = V^*V = I_k} \operatorname{Re}(\operatorname{tr}(UAV)).$$

10. Let A and B be $n \times n$ Hermitian matrices such that $\operatorname{tr}(A+B)^k = \operatorname{tr}(A^k) + \operatorname{tr}(B^k)$ for all positive integers k. Show that

   (a) $\operatorname{rank}(A+B) = \operatorname{rank} A + \operatorname{rank} B$.

   (b) $\operatorname{Im}(A+B) = \operatorname{Im} A + \operatorname{Im} B$.

   (c) $AB = 0$.

   [Hint: For (a), use Problem 7 of Section 5.4. For (c), view A and B as linear transformations on $\operatorname{Im}(A+B)$. Find the matrix of $A+B$ on $\operatorname{Im}(A+B)$ and use the equality case of the Hadamard inequality.]

· ———— ⊙ ———— ·

8.6 A Triangle Inequality for the Matrix $(A^*A)^{1/2}$

This section studies the positive semidefinite matrix $(A^*A)^{1/2}$, denoted by $|A|$, where A is any complex matrix. We call $|A|$ the *modulus* of matrix A. The main result is that for any square matrices A and B of the same size, there exist unitary matrices U and V such that

$$|A + B| \le U^*|A|U + V^*|B|V.$$

As is known (Theorem 8.11), for Hermitian matrices A and B,

$$A \ge B \quad \Rightarrow \quad \lambda_i(A) \ge \lambda_i(B).$$

Our first observation is on the converse of the statement. The inequalities $\lambda_i(A) \ge \lambda_i(B)$ for all i cannot ensure $A \ge B$. For example,

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}.$$

Then $\lambda_1(A) = 3 \ge \lambda_1(B) = 3$ and $\lambda_2(A) = 2 \ge \lambda_2(B) = 1$. But $A - B$, having a negative eigenvalue -1, is obviously not positive semidefinite. We have, however, the following result.

Theorem 8.20 *Let A, $B \in \mathbb{M}_n$ be Hermitian matrices. If*

$$\lambda_i(A) \ge \lambda_i(B)$$

for all $i = 1, 2, \ldots, n$, then there exists a unitary matrix U such that

$$U^*AU \ge B.$$

Proof. Let P and Q be unitary matrices such that

$$A = P^* \operatorname{diag}(\lambda_1(A), \ldots, \lambda_n(A))P,$$

$$B = Q^* \operatorname{diag}(\lambda_1(B), \ldots, \lambda_n(B))Q.$$

The condition $\lambda_i(A) \ge \lambda_i(B)$, $i = 1, \ldots, n$, implies that

$$\operatorname{diag}(\lambda_1(A), \ldots, \lambda_n(A)) - \operatorname{diag}(\lambda_1(B), \ldots, \lambda_n(B)) \ge 0.$$

Multiply both sides by Q^* from the left and Q from the right to get

$$Q^* \operatorname{diag}(\lambda_1(A), \ldots, \lambda_n(A))Q - B \geq 0.$$

Take $U = P^*Q$. Then we have $U^*AU - B \geq 0$, as desired. ∎

We now turn our attention to the positive semidefinite matrix $|A|$. Note that $|A|$ is the unique positive semidefinite matrix satisfying

$$|A|^2 = A^*A.$$

Moreover, for any complex matrix A,

$$A = U|A| \tag{8.19}$$

is a polar decomposition of A, where U is some unitary matrix. Because of such a relation between $|A|$ and A, matrix $|A|$ has drawn much attention, and many interesting results have been obtained.

Note that the eigenvalues of $|A|$ are the square roots of the eigenvalues of A^*A; namely, the singular values of A. In symbols,

$$\lambda(|A|) = (\lambda(A^*A))^{1/2} = \sigma(A).$$

In addition, if $A = UDV$ is a singular value decomposition of A, then

$$|A| = V^*DV \quad \text{and} \quad |A^*| = UDU^*.$$

To prove Thompson's matrix triangle inequality, two theorems are needed. They are of interest in their own right.

Theorem 8.21 *Let A be an n-square complex matrix. Then*

$$\lambda_i\left(\frac{A^* + A}{2}\right) \leq \lambda_i(|A|), \quad i = 1, 2, \ldots, n.$$

Proof. Take $v_1, v_2, \ldots, v_n$ and $w_1, w_2, \ldots, w_n$ to be orthonormal sets of eigenvectors of $\frac{A^*+A}{2}$ and A^*A, respectively. Then for each i,

$$\left(\frac{A^* + A}{2}\right)v_i = \lambda_i\left(\frac{A^* + A}{2}\right)v_i, \quad (A^*A)w_i = (\lambda_i(|A|))^2 w_i.$$

For each fixed positive integer k with $1 \le k \le n$, let

$$S_1 = \mathrm{Span}\{v_1, v_2, \ldots, v_k\}, \quad S_2 = \mathrm{Span}\{w_k, \ldots, w_n\}.$$

Then for some unit vector $x \in S_1 \cap S_2$, by Theorem 8.7,

$$x^* \left(\frac{A^* + A}{2}\right) x \ge \lambda_k \left(\frac{A^* + A}{2}\right)$$

and

$$x^*(A^*A)x \le (\lambda_k(|A|))^2.$$

By the Cauchy–Schwarz inequality, we have

$$\begin{aligned} \lambda_k \left(\frac{A^* + A}{2}\right) &\le x^* \left(\frac{A^* + A}{2}\right) x \\ &= \mathrm{Re}(x^* A x) \\ &\le |x^* A x| \\ &\le \sqrt{x^* A^* A x} \\ &\le \lambda_k(|A|). \quad \blacksquare \end{aligned}$$

Combining Theorems 8.20 and 8.21, we see that for any n-square complex matrix A there exists a unitary matrix U such that

$$\frac{A^* + A}{2} \le U^* |A| U. \tag{8.20}$$

We are now ready to present the matrix triangle inequality.

Theorem 8.22 (Thompson) *For any square complex matrices A and B of the same size, unitary matrices U and V exist such that*

$$|A + B| \le U^* |A| U + V^* |B| V.$$

Proof. By the polar decomposition (8.19), we may write

$$A + B = W|A + B|,$$

where W is a unitary matrix. By (8.20), for some unitary U, V,

$$\begin{aligned} |A+B| &= W^*(A+B) \\ &= \frac{1}{2}\Big(W^*(A+B)+(A+B)^*W\Big) \\ &= \frac{1}{2}(A^*W+W^*A)+\frac{1}{2}(B^*W+W^*B) \\ &\le U^*|W^*A|U+V^*|W^*B|V \\ &= U^*|A|U+V^*|B|V. \ \blacksquare \end{aligned}$$

Note that Theorem 8.22 would be false without the presence of the unitary matrices U and V (Problem 16).

Problems

1. Show that $\frac{A^*+A}{2}$ is Hermitian for any $n\times n$ matrix A and that
$$\frac{A^*+A}{2}\ge 0 \quad\Leftrightarrow\quad \operatorname{Re}(x^*Ax)\ge 0 \quad \text{for all } x\in\mathbb{C}^n.$$

2. Show that $A^*A\ge 0$ for any matrix A. What is the rank of $(A^*A)^{1/2}$?

3. Let A be a square complex matrix and $|A|=(A^*A)^{1/2}$. Show that
   (a) A is positive semidefinite if and only if $|A|=A$.
   (b) A is normal if and only if $|A^*|=|A|$.
   (c) $|A|$ and $|A^*|$ are similar.
   (d) If $A=PU$ is a polar decomposition of A, where $P\ge 0$ and U is unitary, then $|A|=U^*PU$ and $|A^*|=P$.

4. Find $|A|$ and $|A^*|$ for each of the following matrices A.
$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},\quad \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix},\quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},\quad \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}.$$

5. Find $|A|$ and $|A^*|$ for $A=\left(\begin{smallmatrix} 1 & a \\ 0 & 0 \end{smallmatrix}\right)$, where $a>0$.

6. Let $M=\left(\begin{smallmatrix} 0 & A^* \\ A & 0 \end{smallmatrix}\right)$. Show that $|M|=\left(\begin{smallmatrix} |A| & 0 \\ 0 & |A^*| \end{smallmatrix}\right)$.

7. Find $|A|$ for the normal matrix $A\in\mathbb{M}_n$ with spectral decomposition
$$A=U\operatorname{diag}(\lambda_1,\ldots,\lambda_n)U^*=\sum_{i=1}^n \lambda_i u_i u_i^*, \quad \text{where } U=(u_1,\ldots,u_n).$$

8. Let A be an $m \times n$ complex matrix. Show that $|UAV| = V^*|A|V$ for any unitary matrices $U \in \mathbb{M}_m$ and $V \in \mathbb{M}_n$.

9. Let $A \in \mathbb{M}_n$. Then A and U^*AU have the same eigenvalues for any unitary $U \in \mathbb{M}_n$. Show that A and UAV have the same singular values for any unitary $U, V \in \mathbb{M}_n$. Do they have the same eigenvalues?

10. Let A be a Hermitian matrix. If X is a Hermitian matrix commuting with A such that $A \le X$ and $-A \le X$, show that $|A| \le X$.

11. Show that for any matrix A there exist matrices X and Y such that

$$A = (AA^*)X \quad \text{and} \quad A = (AA^*)^{1/2}Y.$$

12. Let A be an $m \times n$ complex matrix, $m \ge n$. Show that there exists an m-square unitary matrix U such that

$$AA^* = U^* \begin{pmatrix} A^*A & 0 \\ 0 & 0 \end{pmatrix} U.$$

13. Show that for any complex matrices A and B (of any sizes)

$$|A \otimes B| = |A| \otimes |B|.$$

14. Show that for any unit column vector $x \in \mathbb{C}^n$ and $A \in \mathbb{M}_n$

$$|x^*Ax|^2 \le x^*|A|^2x.$$

15. Show that $|A + A^*| \le |A| + |A^*$ for all normal matrices A.

16. Construct an example showing it is not true that for A, $B \in \mathbb{M}_n$

$$|A + B| \le |A| + |B|.$$

Show that the trace inequality, however, holds:

$$\operatorname{tr}|A + B| \le \operatorname{tr}(|A| + |B|).$$

17. Let A, $B \in \mathbb{M}_n$, p, $q > 1$, $1/p + 1/q = 1$. Show that

$$|A - B|^2 + |\sqrt{p/q}\,A + \sqrt{q/p}\,B|^2 = p|A|^2 + q|B|^2.$$

18. Let A, $B \in \mathbb{M}_n$, a, $b > 0$, $c \in \mathbb{R}$, $ab \ge c^2$. Show that

$$a|A|^2 + b|B|^2 + c(A^*B + B^*A) \ge 0.$$

19. Let $A_1, \cdots, A_k \in \mathbb{M}_n$, $a_1, \ldots, a_k \geq 0$, $a_1 + \cdots + a_k = 1$. Show that

$$|a_1 A_1 + \cdots + a_k A_k|^2 \leq a_1 |A_1^2 + \cdots + a_k |A_k|^2.$$

20. Let $A,\ B,\ C,\ D \in \mathbb{M}_n$. If $CC^* + DD^* \leq I_n$, show that

$$|CA + DB| \leq (|A|^2 + |B|^2)^{1/2}.$$

21. Let A and B be n-square Hermitian matrices. Show that

$$\begin{pmatrix} A^2 & B^2 \\ B^2 & I \end{pmatrix} \geq 0 \quad \Rightarrow \quad \begin{pmatrix} |A| & |B| \\ |B| & I \end{pmatrix} \geq 0.$$

22. Let A be an $n \times n$ Hermitian matrix with the largest eigenvalue $\lambda_{\max}$ and the smallest eigenvalue $\lambda_{\min}$. Show that

$$\lambda_{\max} - \lambda_{\min} = 2 \max\{|x^* A y| : \ x,\ y \in \mathbb{C}^n \text{ orthonormal}\}.$$

Derive that

$$\lambda_{\max} - \lambda_{\min} \geq 2 \max_{i,j} |a_{ij}|.$$

23. Let A be any $n \times n$ matrix. By Theorem 8.21, for $i = 1, 2, \ldots, n$,

$$\lambda_i\left(\frac{A + A^*}{2}\right) \leq \lambda_i(|A|).$$

Does it follow that

$$\sigma_i\left(\frac{A^* + A}{2}\right) \leq \sigma_i(A)?$$

24. Show that $\frac{A^2+B^2}{2} \geq (\frac{A+B}{2})^2$ for all $n \times n$ Hermitian matrices A and B. With $P = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $Q = \left(\begin{smallmatrix} 0 & x \\ x & 0 \end{smallmatrix}\right)$, show that for any given positive integer $k > 2$, one may choose a sufficiently small positive real number x such that $\frac{P^k+Q^k}{2} - (\frac{P+Q}{2})^k$ has a negative eigenvalue.

CHAPTER 9

Normal Matrices

Introduction: A great deal of elegant work has been done for normal matrices. The goal of this chapter is to present basic results and methods on normal matrices. Section 9.1 gives conditions equivalent to the normality of matrices, Section 9.2 focuses on a special type of normal matrix with entries consisting of zeros and ones, Section 9.3 studies the positive semidefinite matrix $(A^*A)^{1/2}$ associated with a matrix A, and finally Section 9.4 compares two normal matrices.

9.1 Equivalent Conditions

A square complex matrix A is said to be *normal* if it commutes with its conjugate transpose; in symbols,

$$A^*A = AA^*.$$

Matrix normality is one of the most interesting topics in linear algebra and matrix theory, since normal matrices have not only simple structures under unitary similarity but also many applications.

This section presents conditions equivalent to normality.

Theorem 9.1 *Let $A = (a_{ij})$ be an n-square complex matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. The following statements are equivalent.*

F. Zhang, *Matrix Theory: Basic Results and Techniques*, Universitext,
DOI 10.1007/978-1-4614-1099-7_9,

1. *A is normal; that is,* $A^*A = AA^*$.
2. *A is unitarily diagonalizable; namely, there exists an n-square unitary matrix U such that*
$$U^*AU = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n). \tag{9.1}$$
3. *There exists a polynomial* $p(x)$ *such that* $A^* = p(A)$.
4. *There exists a set of eigenvectors of A that form an orthonormal basis for* $\mathbb{C}^n$.
5. *Every eigenvector of A is an eigenvector of* A^*.
6. *Every eigenvector of A is an eigenvector of* $A + A^*$.
7. *Every eigenvector of A is an eigenvector of* $A - A^*$.
8. $A = B + iC$ *for some B and C Hermitian, and* $BC = CB$.
9. *If U is a unitary matrix such that* $U^*AU = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$, *where B and D are square, then B and D are normal and* $C = 0$.
10. *If* $W \subseteq \mathbb{C}^n$ *is an invariant subspace of A, then so is* $W^\perp$.
11. *If x is an eigenvector of A, then* $x^\perp$ *is invariant under A.*
12. *A can be written as* $A = \sum_{i=1}^n \lambda_i E_i$, *where* $\lambda_i \in \mathbb{C}$ *and* $E_i \in \mathbb{M}_n$ *satisfy* $E_i^2 = E_i = E_i^*$, $E_iE_j = 0$ *if* $i \neq j$, *and* $\sum_{i=1}^n E_i = I$.
13. $\operatorname{tr}(A^*A) = \sum_{i=1}^n |\lambda_i|^2$.
14. *The singular values of A are* $|\lambda_1|, |\lambda_2|, \ldots, |\lambda_n|$.
15. $\sum_{i=1}^n (\operatorname{Re}\lambda_i)^2 = \frac{1}{4}\operatorname{tr}(A + A^*)^2$.
16. $\sum_{i=1}^n (\operatorname{Im}\lambda_i)^2 = -\frac{1}{4}\operatorname{tr}(A - A^*)^2$.
17. *The eigenvalues of* $A + A^*$ *are* $\lambda_1 + \overline{\lambda_1}, \ldots, \lambda_n + \overline{\lambda_n}$.
18. *The eigenvalues of* AA^* *are* $\lambda_1\overline{\lambda_{\pi(1)}}, \ldots, \lambda_n\overline{\lambda_{\pi(n)}}$ *for some permutation* π *on* $\{1, 2, \ldots, n\}$.
19. $\operatorname{tr}(A^*A)^2 = \operatorname{tr}\left((A^*)^2A^2\right)$.
20. $(A^*A)^2 = (A^*)^2A^2$.
21. $\|Ax\| = \|A^*x\|$ *for all* $x \in \mathbb{C}^n$.
22. $(Ax, Ay) = (A^*x, A^*y)$ *for all* $x, y \in \mathbb{C}^n$.
23. $|A| = |A^*|$, *where* $|A| = (A^*A)^{1/2}$.

24. $A^* = AU$ *for some unitary* U.

25. $A^* = VA$ *for some unitary* V.

26. $UP = PU$ *if* $A = UP$, *a polar decomposition of* A.

27. $AU = UA$ *if* $A = UP$, *a polar decomposition of* A.

28. $AP = PA$ *if* $A = UP$, *a polar decomposition of* A.

29. A *commutes with a normal matrix of no duplicate eigenvalues.*

30. A *commutes with* $A + A^*$.

31. A *commutes with* $A - A^*$.

32. $A + A^*$ *and* $A - A^*$ *commute.*

33. A *commutes with* A^*A.

34. A *commutes with* $AA^* - A^*A$.

35. $A^*B = BA^*$ *whenever* $AB = BA$.

36. $A^*A - AA^*$ *is a positive semidefinite matrix.*

37. $|(Ax, x)| \leq (|A|x, x)$ *for all* $x \in \mathbb{C}^n$, *where* $|A| = (A^*A)^{1/2}$.

Proof. (2)⇔(1): We show that (1) implies (2). The other direction is obvious. Let $A = U^*TU$ be a Schur decomposition of A. It suffices to show that the upper-triangular matrix $T = (t_{ij})$ is diagonal.

Note that $A^*A = AA^*$ yields $T^*T = TT^*$. Computing and equating the $(1,1)$-entries of T^*T and TT^*, we have $|t_{11}|^2 = |t_{11}|^2 + \sum_{j=2}^n |t_{1j}|^2$. It follows that $t_{1j} = 0$ if $j > 1$. Inductively, we have $t_{ij} = 0$ whenever $i < j$. Thus T is diagonal.

(3)⇔(2): To show that (2) implies (3), we choose a polynomial $p(x)$ of degree at most $n-1$ (by interpolation) such that

$$p(\lambda_i) = \overline{\lambda_i}, \quad i = 1, \ldots, n.$$

Thus, if $A = U^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n)U$ for some unitary matrix U, then

$$\begin{aligned} A^* &= U^* \operatorname{diag}(\overline{\lambda_1}, \ldots, \overline{\lambda_n})U \\ &= U^* \operatorname{diag}(p(\lambda_1), \ldots, p(\lambda_n))U \\ &= U^* p(\operatorname{diag}(\lambda_1, \ldots, \lambda_n))U \\ &= p(U^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n)U) \\ &= p(A). \end{aligned}$$

For the other direction, if $A^* = p(A)$ for some polynomial p, then

$$A^*A = p(A)A = Ap(A) = AA^*.$$

(4)⇔(2): If (9.1) holds, then multiplying by U from the left gives

$$AU = U\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

or

$$Au_i = \lambda_i u_i, \quad i = 1, \ldots, n,$$

where u_i is the ith column of U, $i = 1, \ldots, n$. Thus, the column vectors of U are eigenvectors of A and they form an orthonormal basis of $\mathbb{C}^n$ because U is a unitary matrix.

Conversely, if A has a set of eigenvectors that form an orthonormal basis for $\mathbb{C}^n$, then the matrix U consisting of these vectors as columns is unitary and satisfies (9.1).

(5)⇔(1): Assume that A is normal and let u be a unit eigenvector of A corresponding to eigenvalue λ. Extend u to a unitary matrix with u as the first column. Then

$$U^*AU = \begin{pmatrix} \lambda & \alpha \\ 0 & A_1 \end{pmatrix}. \tag{9.2}$$

The normality of A forces $\alpha = 0$.

Taking the conjugate transpose and by a simple computation, u is an eigenvector of A^* corresponding to the eigenvalue $\overline{\lambda}$ of A^*.

To see the other way around, we use induction on n. Note that

$$Ax = \lambda x \quad \Leftrightarrow \quad (U^*AU)(U^*x) = \lambda(U^*x)$$

for any n-square unitary matrix U. Thus, when considering $Ax = \lambda x$, we may assume that A is upper-triangular by Schur decomposition.

Take $e_1 = (1, 0, \ldots, 0)^T$. Then e_1 is an eigenvector of A. Hence, by assumption, e_1 is an eigenvector of A^*. A direct computation of $A^*e_1 = \mu e_1$ (for some scalar μ) yields that the first column of A^* must consist of zeros except the first component. Thus, if we write

$$A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & B \end{pmatrix}, \quad \text{then} \quad A^* = \begin{pmatrix} \overline{\lambda_1} & 0 \\ 0 & B^* \end{pmatrix}.$$

Every eigenvector of A is an eigenvector of A^*, thus this property is inherited by B and B^*. An induction hypothesis on B shows that A is diagonal. It follows that A is normal.

(6)⇔(5): Let $(A + A^*)u = \lambda u$ and $Au = \mu u$, $u \neq 0$. Then $A^*u = \lambda u - Au = (\lambda - \mu)u$; that is, u is an eigenvector of A^*. Conversely, let $Au = \lambda u$ and $A^*u = \mu u$. Then $(A+A^*)u = (\lambda+\mu)u$.

(7)⇔(5) is similarly proven.

(8)⇔(1): It is sufficient to notice that $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$.

(9)⇔(1): We show that (1) implies (9). The other direction is easy. Upon computation, we have that $A^*A = AA^*$ implies

$$\begin{pmatrix} B^*B & B^*C \\ C^*B & C^*C + D^*D \end{pmatrix} = \begin{pmatrix} BB^* + CC^* & CD^* \\ DC^* & DD^* \end{pmatrix}.$$

Therefore,

$$B^*B = BB^* + CC^* \quad \text{and} \quad C^*C + D^*D = DD^*.$$

By taking the trace for both sides of the first identity and noticing that $\operatorname{tr}(BB^*) = \operatorname{tr}(B^*B)$, we obtain $\operatorname{tr}(CC^*) = 0$. This forces $C = 0$. Thus B is normal, and so is D by the second identity.

We have shown that the first nine conditions are all equivalent.

(9)⇒(10): It suffices to note that $\mathbb{C}^n = W \oplus W^\perp$ and that a basis of W and a basis of $W^\perp$ form a basis of $\mathbb{C}^n$.

(10)⇒(11)⇒(4): (11) is a restatement of (10) with W consisting of an eigenvector of A. For (11)⇒(4), if $Ax = \lambda x$, where $x \neq 0$, we may assume that x is a unit vector. By (11), $x^\perp$ is invariant under A. Consider the restriction of A on $x^\perp$. Inductively, we obtain a set of eigenvectors of A that form an orthonormal basis of $\mathbb{C}^n$.

(12)⇒(1): It is by a direct computation. To show (2)⇒(12), we write $U = (u_1, \ldots, u_n)$, where u_i is the ith column of U. Then

$$A = U \operatorname{diag}(\lambda_1, \ldots, \lambda_n) U^* = \lambda_1 u_1 u_1^* + \cdots + \lambda_n u_n u_n^*.$$

Take $E_i = u_i u_i^*$, $i = 1, \ldots, n$. (12) then follows.

(13)⇔(2): Let $A = U^*TU$ be a Schur decomposition of A, where U is unitary and $T = (t_{ij})$ is upper-triangular. Then $A^*A = U^*T^*TU$.

Hence, $\operatorname{tr}(A^*A) = \operatorname{tr}(T^*T)$. On the other hand, upon computation,

$$\operatorname{tr}(A^*A) = \sum_{i,\, j=1}^{n} |a_{ij}|^2, \quad \operatorname{tr}(T^*T) = \sum_{i=1}^{n} |\lambda_i|^2 + \sum_{i<j} |t_{ij}|^2.$$

Thus, $\operatorname{tr}(A^*A) = \sum_{i=1}^{n} |\lambda_i|^2$ if and only if $t_{ij} = 0$ for all $i < j$; that is, T is diagonal and A is unitarily diagonalizable.

(14)⇒(13): If the singular values of A are $\sigma_1, \dots, \sigma_n$, then

$$\begin{aligned} \operatorname{tr}(A^*A) &= \lambda_1(A^*A) + \cdots + \lambda_n(A^*A) \\ &= \sigma_1^2 + \cdots + \sigma_n^2 \\ &= |\lambda_1|^2 + \cdots + |\lambda_n|^2, \end{aligned}$$

which is (13). For the other direction, obviously (13)⇒(2)⇒(14).

(15)⇒(13): We may assume that A is an upper-triangular matrix, because the identity holds when A is replaced by U^*AU, where U is any unitary matrix. Notice that

$$\operatorname{tr}(A + A^*)^2 = \operatorname{tr} A^2 + 2\operatorname{tr}(A^*A) + \operatorname{tr}(A^*)^2.$$

It follows that

$$\operatorname{tr}(A^*A) = \frac{1}{2}\Big(\operatorname{tr}(A + A^*)^2 - \operatorname{tr} A^2 - \operatorname{tr}(A^*)^2\Big).$$

Since $4(\operatorname{Re}\lambda_i)^2 = (\lambda_i + \bar{\lambda}_i)^2$, (15) implies (13). (15) follows from (2).

Similarly, (16)⇒(13) and (2)⇒(16).

(17)⇒(15): If the eigenvalues of $A + A^*$ are $\lambda_1 + \overline{\lambda_1}, \dots, \lambda_n + \overline{\lambda_n}$, then their squares are the eigenvalues of $(A + A^*)^2$. Thus,

$$\operatorname{tr}(A + A^*)^2 = \sum_{i=1}^{n} (\lambda_i + \overline{\lambda_i})^2 = 4\sum_{i=1}^{n} (\operatorname{Re}\lambda_i)^2.$$

(17) follows from (2) at once.

(18)⇔(14): Obviously (14) implies (18). For the converse, suppose, without loss of generality, that $\lambda_1\overline{\lambda_{\pi(1)}} > 0$ is the largest eigenvalue of AA^*. As is known, all $|\lambda_i| \le \sigma_{\max}(A)$. If $|\lambda_{\pi(1)}| \ne |\lambda_1|$, then $|\lambda_1\overline{\lambda_{\pi(1)}}| < \sigma_{\max}^2(A) = \lambda_{\max}(AA^*)$, a contradiction. Hence,

$|\lambda_{\pi(1)}| = |\lambda_1|$. On the other hand, $\lambda_1\overline{\lambda_{\pi(1)}}$ is a positive number. Thus, $\lambda_1 = \lambda_{\pi(1)}$. The rest follows by induction.

(19) is immediate from (1). To see the other way, we make use of the facts that for any square matrices X and Y of the same size,

$$\operatorname{tr}(XY) = \operatorname{tr}(YX)$$

and

$$\operatorname{tr}(X^*X) = 0 \quad \Leftrightarrow \quad X = 0.$$

Upon computation and noting that $\operatorname{tr}(AA^*)^2 = \operatorname{tr}(A^*A)^2$, we have

$$\begin{aligned} &\operatorname{tr}\Big((A^*A - AA^*)^*(A^*A - AA^*)\Big) \\ &\quad= \operatorname{tr}(A^*A - AA^*)^2 \\ &\quad= \operatorname{tr}(A^*A)^2 - \operatorname{tr}\left((A^*)^2A^2\right) - \operatorname{tr}\left(A^2(A^*)^2\right) + \operatorname{tr}(AA^*)^2, \end{aligned}$$

which equals 0 by assumption. Thus, $A^*A - AA^* = 0$.

(20)$\Rightarrow$(19)$\Rightarrow$(1)$\Rightarrow$(20).

(21)$\Leftrightarrow$(1): By squaring both sides, the norm identity in (21) is rewritten as the inner product identity

$$(Ax, Ax) = (A^*x, A^*x),$$

which is equivalent to

$$(x, A^*Ax) = (x, AA^*x),$$

or

$$\Big(x, (A^*A - AA^*)x\Big) = 0.$$

This holds for all $x \in \mathbb{C}^n$ if and only if $A^*A - AA^* = 0$.

(22)$\Rightarrow$(21) by setting $x = y$; (21)$\Rightarrow$(1)$\Rightarrow$(22).

(23)$\Leftrightarrow$(1): This is by the uniqueness of the square root.

(24)$\Leftrightarrow$(1): If $A^* = AU$ for some unitary U, then

$$A^*A = A^*(A^*)^* = (AU)(AU)^* = AA^*,$$

and A is normal. For the converse, we show (2)$\Rightarrow$(24). Let

$$A = V^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n)V,$$

where V is unitary. Take

$$U = V^* \operatorname{diag}(l_1, \ldots, l_n)V,$$

where $l_i = \frac{\overline{\lambda_i}}{\lambda_i}$ if $\lambda_i \neq 0$, and $l_i = 1$ otherwise, for $i = 1, \ldots, n$. Then

$$\begin{aligned} A^* &= V^* \operatorname{diag}(\,\overline{\lambda_1}, \ldots, \overline{\lambda_n}\,)V \\ &= V^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n)V\, V^* \operatorname{diag}(l_1, \ldots, l_n)V \\ &= AU. \end{aligned}$$

Similarly, (25) is equivalent to (1).

(26)⇔(1): If $A = UP$, where U is unitary and P is positive semidefinite, then $A^*A = AA^*$ implies

$$P^*P = UPP^*U^* \quad \text{or} \quad P^2 = UP^2U^*.$$

By taking square roots, we have

$$P = UPU^* \quad \text{or} \quad PU = UP.$$

The other direction is easy to check: $A^*A = P^2 = AA^*$.

(27)⇔(26): Note that U is invertible.

(28)⇔(26): We show that (28) implies (26). The other direction is immediate by multiplying P from the right-hand side.

Suppose $AP = PA$; that is, $UP^2 = PUP$. If P is nonsingular, then obviously $UP = PU$. Let $r = \operatorname{rank}(A) = \operatorname{rank}(P)$ and write

$$P = V^* \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} V,$$

where V is unitary and D is $r \times r$ positive definite diagonal, $r < n$. Then $UP^2 = PUP$ gives, with $W = VUV^*$,

$$W \begin{pmatrix} D^2 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} W \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}.$$

Partition W as

$$\begin{pmatrix} W_1 & W_2 \\ W_3 & W_4 \end{pmatrix}, \quad \text{where } W_1 \text{ is } r \times r.$$

Then

$$W_1 D^2 = DW_1 D \quad \text{and} \quad W_3 D^2 = 0,$$

which imply, for D is nonsingular,

$$W_1 D = DW_1 \quad \text{and} \quad W_3 = 0.$$

It follows that $W_2 = 0$ because W is unitary and that

$$W \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} W.$$

This results in $UP = PU$ at once.

(29)$\Leftrightarrow$(2): Let A commute with B, where B is normal and all the eigenvalues of B are distinct. Write $B = V^*CV$, where V is unitary and $C = \operatorname{diag}(c_1, \ldots, c_n)$ is diagonal and all c_i are distinct. Then $AB = BA$ implies $WC = CW$, where $W = (w_{ij}) = VAV^*$. It follows that $w_{ij}c_i = w_{ij}c_j$, i.e., $w_{ij}(c_i - c_j) = 0$ for all i and j. Since $c_i \neq c_j$ whenever $i \neq j$, we have $w_{ij} = 0$ whenever $i \neq j$. Thus, VAV^* is diagonal; that is, A is unitarily diagonalizable, and it is normal. Conversely, if (2) holds, then we take $B = U \operatorname{diag}(1, 2, \ldots, n)U^*$. It is easy to check that B is normal and $AB = BA$.

(30), (31), and (32) are equivalent to (1) by direct computation.

(33)$\Rightarrow$(20): If A commutes with A^*A, then

$$AA^*A = A^*A^2.$$

Multiply both sides by A^* from the left to get

$$(A^*A)^2 = (A^*)^2A^2.$$

(34)$\Rightarrow$(19) is similarly proven. (1) easily implies (33) and (34).

(35)$\Leftrightarrow$(1): Take $B = A$ for the normality. For the converse, suppose that A is normal and that A and B commute, and let $A = U^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n)U$, where U is unitary. Then $AB = BA$ implies

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n)(UBU^*) = (UBU^*) \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

Denote $T = UBU^* = (t_{ij})$. Then

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n)T = T \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

which gives $(\lambda_i - \lambda_j)t_{ij} = 0$; thus, $(\overline{\lambda_i} - \overline{\lambda_j})t_{ij} = 0$, for all i and j, which, in return, implies that $A^*B = BA^*$.

(36)$\Leftrightarrow$(1): This is a combination of two facts: $\operatorname{tr}(XY - YX) = 0$ for all square matrices X and Y of the same size; and if matrix X is positive semidefinite, then $\operatorname{tr} X = 0$ if and only if $X = 0$.

For (37), notice that $|(Ax, x)| \le (|A|x, x)$ is unitarily invariant; that is, it holds if and only if $|(U^*AUx, x)| \le (|U^*AU|x, x)$, where U is any unitary matrix. (Bear in mind that $|U^*AU| = U^*|A|U$.) By the Schur decomposition, we may assume that A is upper-triangular.

If A is normal, then A is unitarily diagonalizable. We may assume that $A = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. Then the inequality is the same as saying that $|\sum_i \lambda_i |x_i|^2| \le \sum_i |\lambda_i||x_i|^2$, which is obvious.

For the converse, if $|(Ax, x)| \le (|A|x, x)$ for all $x \in \mathbb{C}^n$, where A is upper-triangular, we show that A is in fact a diagonal matrix. We demonstrate the proof for the case of $n = 2$; the general case is similarly proven by induction. Let $A = \begin{pmatrix} \lambda_1 & \alpha \\ 0 & \lambda_2 \end{pmatrix}$. We show $\alpha = 0$.

If $\lambda_1 = \lambda_2 = 0$ and $\alpha \ne 0$, then take positive numbers s, t, $s > t$; set $x = (s, t)^T$. Then $|x^*Ax| = st|\alpha|$ and $x^*|A|x = t^2|\alpha|$. That $|x^*Ax| \le x^*|A|x$ implies $s \le t$, a contradiction.

If λ_1 (or λ_2) is not 0. Let $|A| = \begin{pmatrix} a & b \\ \bar{b} & c \end{pmatrix}$. Putting $x = (1, 0)^T$ in $|x^*Ax| \le x^*|A|x$ gives $|\lambda_1| \le a$. Computing $|A|^2 = A^*A$ and comparing the entries in the upper-left corners, we have $a^2 + |b|^2 = |\lambda_1|^2$. So $b = 0$. Inspecting the entries in the $(1, 2)$ positions of $|A|^2$ and A^*A, we get $\bar{\lambda}_1\alpha = (a + c)b = 0$. Thus $\alpha = 0$. ∎

As many as 90 equivalent conditions of normal matrices have been observed in the literature. More are shown in the following exercises and in the later sections.

Problems

1. Is matrix $\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$ normal, Hermitian, or symmetric? How about $\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$?

2. Let A be an n-square matrix. Show that for all unit vector $x \in \mathbb{C}^n$

$$|(Ax, x)| \le (|A|^2x, x).$$

[Note: If the square is dropped, then A has to be normal. See (37).]

3. Show that each of the following conditions is equivalent to the normality of matrix $A \in \mathbb{M}_n$.

   (a) A has a linearly independent set of n eigenvectors, and any two corresponding to distinct eigenvalues are orthogonal.

   (b) $(Ax, Ay) = (A^*x, A^*y)$ for all $x, y \in \mathbb{C}^n$.

   (c) $(Ax, Ax) = (A^*x, A^*x)$ for all $x \in \mathbb{C}^n$.

4. Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the eigenvalues of matrix $A \in \mathbb{M}_n$. Show that

$$|\lambda_i \lambda_j| \le \lambda_1(A^*A)$$

   for any pair λ_i, λ_j, where $\lambda_1(A^*A)$ is the largest eigenvalue of A^*A.

5. Let A be an $n \times n$ complex matrix with eigenvalues λ_i and singular values σ_i arranged as $|\lambda_1| \ge \cdots \ge |\lambda_n|$ and $\sigma_1 \ge \cdots \ge \sigma_n$. Show that

$$\sigma_1\sigma_2\cdots\sigma_k = |\lambda_1\lambda_2\cdots\lambda_k|, \quad k = 1, 2, \dots, n, \quad \Leftrightarrow \quad A^*A = AA^*.$$

6. Let A be a normal matrix. Show that $Ax = 0$ if and only if $A^*x = 0$.

7. Let A be a normal matrix. Show that if x is an eigenvector of A, then A^*x is also an eigenvector of A for the same eigenvalue.

8. If matrix A commutes with some normal matrix with distinct eigenvalues, show that A is normal. Is the converse true?

9. Show that unitary matrices, Hermitian matrices, skew-Hermitian matrices, real orthogonal matrices, and permutation matrices are all normal. Is a complex orthogonal matrix normal?

10. When is a normal matrix Hermitian? Positive semidefinite? Skew-Hermitian? Unitary? Nilpotent? Idempotent?

11. When is a triangular matrix normal?

12. Let A be a square matrix. Show that if A is a normal matrix, then $f(A)$ is normal for any polynomial f. If $f(A)$ is normal for some nonzero polynomial f, does it follow that A is normal?

13. Show that (33) is equivalent to (28) using Problem 31, Section 7.1.

14. Show that two normal matrices are similar if and only if they have the same set of eigenvalues and if and only if they are unitarily similar.

15. Let A and B be normal matrices of the same size. If $AB = BA$, show that AB is normal and that there exists a unitary matrix U that diagonalizes both A and B.

16. Let A be a nonsingular matrix and let $M = A^{-1}A^*$. Show that A is normal if and only if M is unitary.

17. Let $A \in \mathbb{M}_n$ be a normal matrix. Show that for any unitary $U \in \mathbb{M}_n$,
$$\min_i\{|\lambda_i(A)|\} \le |\lambda_i(AU)| \le \max_i\{|\lambda_i(A)|\}.$$

18. Let A be a normal matrix. If $A^k = I$ for some positive integer k, show that A is unitary.

19. Let B be an n-square matrix and let A be the block matrix
$$A = \begin{pmatrix} B & B^* \\ B^* & B \end{pmatrix}.$$
Show that A is normal and that if B is normal with eigenvalues $\lambda_t = x_t + y_t i$, $x_t,\ y_t \in \mathbb{R}$, $t = 1, 2, \dots, n$, then
$$\det A = 4i \prod_{t=1}^{n} x_t y_t.$$

20. Show that $\mathbb{C}^n = \operatorname{Ker} A \oplus \operatorname{Im} A$ for any n-square normal matrix A.

21. Let A and B be $n \times n$ normal matrices. If $\operatorname{Im} A \perp \operatorname{Im} B$; that is, $(x, y) = 0$ for all $x \in \operatorname{Im} A$, $y \in \operatorname{Im} B$, show that $A + B$ is normal.

22. Let A be a normal matrix. Show that $A\bar{A} = 0 \Leftrightarrow AA^T = A^T A = 0$.

23. Let A be Hermitian, B be skew-Hermitian, and $C = A + B$. Show that the following statements are equivalent.

(a) C is normal. (b) $AB = BA$. (c) AB is skew-Hermitian.

24. Show that for any n-square complex matrix A
$$\operatorname{tr}(A^*A)^2 \ge \operatorname{tr}\Big((A^*)^2 A^2\Big).$$
Equality holds if and only if A is normal. Is it true that
$$(A^*A)^2 \ge (A^*)^2 A^2?$$

25. Verify with the following matrix A that $A^*A - AA^*$ is an entrywise nonnegative matrix, but A is not normal (i.e., $A^*A - AA^* \ne 0$):
$$A = \begin{pmatrix} 0 & 1 & 3 \\ -\sqrt{2} & 0 & -\sqrt{3} \\ 2\sqrt{2} & 2 & 0 \end{pmatrix}.$$

26. Let A and B be $n \times n$ matrices. The matrix $AB - BA$ is called the *commutator* of A and B, and it is denoted by $[A, B]$. Show that

    (a) $\operatorname{tr}[A, B] = 0$.

    (b) $[A, B]^* = [B^*, A^*]$.

    (c) $[A, B + C] = [A, B] + [A, C]$.

    (d) $[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0$.

    (e) $[A, B]$ is never similar to the identity matrix.

    (f) If A and B are both Hermitian or skew-Hermitian, then $[A, B]$ is skew-Hermitian.

    (g) If A and B are Hermitian, then the real part of every eigenvalue of $[A, B]$ is zero.

    (h) A is normal if and only if $[A, A^*] = 0$.

    (i) A is normal if and only if $[A, [A, A^*]] = 0$.

27. Complete the proof of Condition (37) for the case $n > 2$.

28. Let A and B be normal matrices of the same size. Show that

    (a) $AM = MA \Rightarrow A^*M = MA^*$.

    (b) $AM = MB \Rightarrow A^*M = MB^*$.

29. Let A and B be n-square matrices.

    (a) If A and B are Hermitian, show that AB is Hermitian if and only if A and B commute.

    (b) Give an example showing that the generalization of (a) to normal matrices is not valid.

    (c) Show (a) does hold for normal matrices if A (or B) is such a matrix that its different eigenvalues have different moduli.

    (d) If A is positive semidefinite and B is normal, show that AB is normal if and only if A and B commute.

9.2 Normal Matrices with Zero and One Entries

Matrices of zeros and ones are referred to as $(0,1)$-*matrices*. $(0,1)$-matrices have applications in graph theory and combinatorics. This section presents three theorems on matrices with zero and one entries: the first one shows how to construct a symmetric (normal) $(0,1)$-matrix from a given $(0,1)$-matrix, the second one gives a sufficient condition on normality, and the last one is on commutativity.

Given an $m \times n$ $(0,1)$-matrix, say A, we add the 1s in each row to get row sums $r_1, r_2, \ldots, r_m$. We call $R = (r_1, r_2, \ldots, r_m)$ the *row sum vector* of A and denote it by $R(A)$. Similarly, we can define the *column sum vector* of A and denote it by $S(A)$. For example,

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \qquad \begin{array}{l} R(A) = (r_1, r_2, r_3, r_4) = (1, 2, 3, 2), \\ S(A) = (s_1, s_2, s_3, s_4) = (2, 3, 1, 2). \end{array}$$

The sum of the components of $R(A)$ is equal to the total number of 1s in A. The same is true of $S(A)$. Given vectors R and S of nonnegative integers, there may not exist a $(0,1)$-matrix that has R as its row sum vector and S as its column sum vector. For instance, no 3×3 $(0,1)$-matrix A satisfies $R(A) = (3, 1, 1)$ and $S(A) = (3, 2, 0)$. For what R and S does there exist a $(0,1)$-matrix that has R and S as its row and column sum vectors? This is an intriguing problem and has been well studied in combinatorial matrix theory.

Apparently, for a symmetric $(0,1)$-matrix, the row sum vector and column sum vector coincide. The following result says the converse is also true in some sense (via reconstruction). Thus there exists a $(0,1)$-matrix that has the given vector R as its row and column sum vectors if and only if there exists such a normal $(0,1)$-matrix and if and only if there exists such a symmetric $(0,1)$-matrix.

Theorem 9.2 *If there exists a (0, 1)-matrix A with $R(A) = S(A) = R$, then there exists a symmetric (0, 1)-matrix B that has the same row and column vector as A; that is, $R(B) = S(B) = R$.*

Proof. We use induction on n. If $n = 1$ or 2, it is obvious. Let $n > 2$. We may also assume that A contains no zero row (or column). Suppose that the assertion is true for matrices of order less than n. If A is symmetric, there is nothing to prove. So we assume that $A = (a_{ij})$ is not symmetric.

If the first column is the transpose of the first row, i.e., they are identical when regarded as n-tuples, then by induction hypothesis on the $(n-1)$-square submatrix in the lower-right corner, we are done. Otherwise, because A is not symmetric and $R(A) = S(A)$, we have $a_{1p} = 1$, $a_{1q} = 0$, $a_{p1} = 0$, $a_{q1} = 1$ for some p and q.

Now consider the rows p and q. With $a_{p1} = 0, a_{q1} = 1$, if $a_{pt} \leq a_{qt}, t = 2, \ldots, n$, then $r_p < r_q$. For the columns p and q, with $a_{1p} = 1, a_{1q} = 0$, if $a_{tp} \geq a_{tq}, t = 2, \ldots, n$, then $s_p > s_q$. Both cases cannot happen at the same time, or they would lead to $r_p < r_q = s_q < s_p = r_p$, a contradiction. Therefore, there must exist a t such that either $a_{pt} = 1$ and $a_{qt} = 0$ or $a_{tp} = 0$ and $a_{tq} = 1$.

Now interchange the 0s and 1s in the intersections of rows p, q and columns $1, t$, or columns p, q and rows $1, t$. Notice that such an interchange reduces the number of different entries in the first row and first column of A without affecting the row and column sum vectors of A. Inductively, we can have a matrix in which the first column is the transpose of the first row. By induction on the size of matrices, a symmetric matrix is obtained from A. ∎

Take the following matrix A as an example. In rows 2 and 4, replacing the 0s by 1s and 1s by 0s (all $\times$s remain unchanged) results in the matrix A_1 that has the same row and column sum vector as A, whereas the first column of A_1 is "closer" to the transpose of the first row of A_1 (or A).

$$A = \begin{pmatrix} \times & 1 & \times & 0 \\ 0 & \times & 1 & \times \\ \times & \times & \times & \times \\ 1 & \times & 0 & \times \end{pmatrix}, \quad A_1 = \begin{pmatrix} \times & 1 & \times & 0 \\ 1 & \times & 0 & \times \\ \times & \times & \times & \times \\ 0 & \times & 1 & \times \end{pmatrix}.$$

In what follows, J_n, or simply J, denotes the n-square matrix all of whose entries are 1. As usual, I is the identity matrix.

The following theorem often appears in combinatorics when a configuration of subsets is under investigation.

Theorem 9.3 *Let A be an n-square (0, 1)-matrix. If*

$$AA^T = tI + J \tag{9.3}$$

for some positive integer t, then A is a normal matrix.

Proof. By considering the diagonal entries, we see that (9.3) implies that each row sum of A equals $t+1$; that is,

$$AJ = (t+1)J. \tag{9.4}$$

Matrix A is nonsingular, since the determinant

$$(\det A)^2 = \det(AA^T) = \det(tI + J) = (t+n)t^{n-1}$$

is nonzero. Thus by (9.4), we have

$$A^{-1}J = (t+1)^{-1}J.$$

Multiplying both sides of (9.3) by J from the right reveals

$$AA^TJ = tJ + J^2 = (t+n)J.$$

It follows by multiplying A^{-1} from the left that

$$A^TJ = (t+1)^{-1}(t+n)J.$$

By taking the transpose, we have

$$JA = (t+1)^{-1}(t+n)J. \tag{9.5}$$

Multiply both sides by J from the right to get

$$JAJ = n(t+1)^{-1}(t+n)J.$$

Multiply both sides of (9.4) by J from the left to get

$$JAJ = n(t+1)J.$$

It follows by comparison of the right-hand sides that

$$(t+1)^2 = t+n \quad \text{or} \quad n = t^2 + t + 1.$$

Substituting this into (9.5), one gets

$$JA = (t+1)J.$$

From (9.4),

$$AJ = JA \quad \text{or} \quad A^{-1}JA = J.$$

We then have

$$\begin{aligned} A^T A &= A^{-1}(AA^T)A \\ &= A^{-1}(tI + J)A \\ &= tI + A^{-1}JA \\ &= tI + J \\ &= AA^T. \quad \blacksquare \end{aligned}$$

Our next result asserts that if the product of two $(0,1)$-matrices is a matrix whose diagonal entries are all 0s and whose off-diagonal entries are all 1s, then these two matrices commute. The proof of this theorem uses the determinant identity (Problem 5, Section 2.2)

$$\det(I_m + AB) = \det(I_n + BA),$$

where A and B are $m \times n$ and $n \times m$ matrices, respectively.

Theorem 9.4 *Let A and B be n-square (0, 1)-matrices such that*

$$AB = J_n - I_n. \tag{9.6}$$

Then

$$AB = BA.$$

Proof. Let a_i and b_j be the columns of A and B^T, respectively:

$$A = (a_1, \ldots, a_n), \quad B^T = (b_1, \ldots, b_n).$$

Then, by computation,

$$0 = \operatorname{tr}(AB) = \operatorname{tr}(BA) = \sum_{i=1}^{n} b_i^T a_i.$$

Thus $b_i^T a_i = 0$ for each i, since A and B have nonnegative entries.

Rewrite equation (9.6) as

$$I_n = J_n - AB \quad \text{or} \quad I_n = J_n - \sum_s a_s b_s^T.$$

Then

$$I_n + a_i b_i^T + a_j b_j^T = J_n - \sum_{s \neq i,\, j} a_s b_s^T.$$

Notice that the right-hand side contains $n-1$ matrices, each of which is of rank one. Thus, by the rank formula for sum (Problem 6), the matrix on the right-hand side has rank at most $n-1$. Hence the matrix on the left-hand side is singular. We have

$$\begin{aligned} 0 &= \det(I_n + a_i b_i^T + a_j b_j^T) \\ &= \det\Big(I_n + (a_i, a_j)(b_i, b_j)^T\Big) \\ &= \det\Big(I_2 + (b_i, b_j)^T (a_i, a_j)\Big) \\ &= \det\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & b_i^T a_j \\ b_j^T a_i & 0 \end{pmatrix}\right) \\ &= 1 - (b_i^T a_j)(b_j^T a_i). \end{aligned}$$

This forces $b_i^T a_j = 1$ for each pair of i and j, $i \neq j$, because A and B are $(0,1)$-matrices. It follows, by combining with $b_i^T a_i = 0$, that

$$BA = (b_i^T a_j) = J_n - I_n = AB. \quad \blacksquare$$

Problems

1. Show that no 3×3 $(0,1)$-matrix A satisfies $R(A) = (3,1,1)$ and $S(A) = (3,2,0)$. But there does exist a 3×3 $(0,1)$-matrix B that satisfies $S(B) = (3,1,1)$ and $R(B) = (3,2,0)$.
2. Show that $A \in \mathbb{M}_n$ is normal if and only if $I - A$ is normal.
3. Let A be an n-square $(0,1)$-matrix. Denote the number of 1s in row i by r_i and in column j by c_j. Show that

$$A \text{ is normal} \quad \Rightarrow \quad r_i = c_i \text{ for each } i \text{ and } J_n - A \text{ is normal.}$$

4. Construct a nonsymmetric 4×4 normal matrix of zeros and ones such that each row and each column sum equal 2.

5. Let A be a 3×3 $(0,1)$-matrix. Show that $\det A$ equals 0, ± 1, or ± 2.

6. Let $A_1, \ldots, A_n$ be $m \times m$ matrices each with rank 1. Show that

$$\operatorname{rank}(A_1 + \cdots + A_n) \le n.$$

7. Let A be a $(0,1)$-matrix with row sum vector $R = (r_1, r_2, \ldots, r_m)$, where $r_1 \ge r_2 \ge \cdots \ge r_m$. Show that

$$\sum_{i=1}^{k} r_i \le kt + \sum_{j=t+1}^{n} r_j.$$

8. Does there exist a normal matrix with real entries of the sign pattern

$$\begin{pmatrix} + & + & 0 \\ + & 0 & + \\ + & + & + \end{pmatrix}?$$

9. If A is an $n \times n$ matrix with integer entries, show that $2Ax = x$ has no nontrivial solutions; that is, the only solution is $x = 0$.

10. Let A be an n-square $(0,1)$-matrix such that $AA^T = tI + J$ for some positive integer t, and let $C = J - A$. Show that A commutes with C and C^T. Compute CC^T.

11. Let A be an n-square $(0,1)$-matrix such that $AA^T = tI + J$ for some positive integer t. Find the singular values of A in terms of n and t and conclude that all A satisfying the equation have the same singular values. Do they have the same eigenvalues? When is A nonsingular?

12. Let A be a $v \times v$ matrix with zero and one entries such that

$$AA^T = (k - h)I + hJ,$$

where v, k, and h are positive integers satisfying $0 < h < k < v$.

   (a) Show that A is normal.
   (b) Show that $h = \frac{1}{v-1} k(k-1)$.
   (c) Show that $A^{-1} = \frac{1}{k(k-h)}(kA^T - hJ)$.
   (d) Find the eigenvalues of AA^T.

9.3 Normality and Cauchy–Schwarz–Type Inequalities

We consider in this section the inequalities involving diagonal entries, eigenvalues, and singular values of matrices. The equality cases of these inequalities will result in the normality of the matrices.

Theorem 9.5 (Schur Inequality) *Let $A = (a_{ij})$ be an n-square complex matrix having eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then*

$$\sum_{i=1}^{n} |\lambda_i|^2 \le \sum_{i,\, j=1}^{n} |a_{ij}|^2.$$

Equality occurs if and only if A is normal.

Proof. Let $A = U^*TU$ be a Schur decomposition of A, where U is unitary and T is upper-triangular. Then $A^*A = U^*T^*TU$; consequently, $\operatorname{tr}(A^*A) = \operatorname{tr}(T^*T)$. Upon computation, we have

$$\operatorname{tr}(A^*A) = \sum_{i,\, j=1}^{n} |a_{ij}|^2$$

and

$$\operatorname{tr}(T^*T) = \sum_{i=1}^{n} |\lambda_i|^2 + \sum_{i<j} |t_{ij}|^2.$$

The inequality is immediate. For the equality case, notice that each $t_{ij} = 0,\ i < j$; that is, T is diagonal. Hence A is unitarily diagonalizable, thus normal. The other direction is obvious. ∎

An interesting application of this result is to show that if matrices A, B, and AB are normal, then so is BA (Problem 11).

Theorem 9.6 *Let $A = (a_{ij})$ be an n-square complex matrix having singular values $\sigma_1, \sigma_2, \ldots, \sigma_n$. Then*

$$|\operatorname{tr} A| \le \sigma_1 + \cdots + \sigma_n. \tag{9.7}$$

Equality holds if and only if $A = uP$ for some $P \ge 0$ and some $u \in \mathbb{C}$ with $|u| = 1$; consequently, A is normal (but not conversely).

Proof. Let $A = UDV$ be a singular value decomposition of A, where $D = \operatorname{diag}(\sigma_1, \dots, \sigma_n)$ with $\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = 0 = \cdots = 0$, $r = \operatorname{rank}(A)$, and U and V are unitary. By computation, we have,

$$\begin{aligned}
|\operatorname{tr} A| &= \left| \sum_{i=1}^{n} \sum_{j=1}^{n} u_{ij} \sigma_j v_{ji} \right| \\
&= \left| \sum_{j=1}^{n} \sum_{i=1}^{n} u_{ij} v_{ji} \sigma_j \right| \\
&\leq \sum_{j=1}^{n} \left| \sum_{i=1}^{n} u_{ij} v_{ji} \right| \sigma_j \\
&\leq \sum_{j=1}^{n} \left(\sum_{i=1}^{n} |u_{ij} v_{ji}| \right) \sigma_j \\
&\leq \sum_{j=1}^{n} \sigma_j.
\end{aligned}$$

The last inequality was due to the Cauchy–Schwarz inequality:

$$\sum_{i=1}^{n} |u_{ij} v_{ji}| \leq \sqrt{\sum_{i=1}^{n} |u_{ij}|^2 \, \sum_{i=1}^{n} |v_{ji}|^2} = 1.$$

If equality holds for the overall inequality, then

$$\left| \sum_{i=1}^{n} u_{ij} v_{ji} \right| = \sum_{i=1}^{n} |u_{ij} v_{ji}| = 1, \quad \text{for each } j \leq r.$$

Rewrite $\sum_{i=1}^{n} u_{ij} v_{ji}$ as (u_j, v_j^*), where u_j is the jth column of U and v_j is the jth row of V. By the equality case of the Cauchy–Schwarz inequality, it follows that $u_j = c_j v_j^*$, for each $j \leq r$, where c_j is a constant with $|c_j| = 1$. Thus, by switching v_i and v_i^*, $|\operatorname{tr} A|$ equals

$$|\operatorname{tr}(c_1 \sigma_1 v_1^* v_1 + \cdots + c_r \sigma_r v_r^* v_r)| = |c_1 \sigma_1 v_1 v_1^* + \cdots + c_r \sigma_r v_r v_r^*|.$$

Notice that $v_i v_i^* = 1$ for each i. By Problem 6 of Section 5.7, we have $A = c_1 V^* D V$. The other direction is easy to verify. ■

There is a variety of Cauchy–Schwarz inequalities of different types. The matrix version of the Cauchy–Schwarz inequality has been obtained by different methods and techniques. Below we give some Cauchy–Schwarz matrix inequalities involving the matrix

$$|A| = (A^*A)^{1/2}, \ \text{ the } \textit{modulus} \text{ of } A.$$

Obviously, if $A = UDV$ is a singular value decomposition of A, where U and V are unitary and D is nonnegative diagonal, then

$$|A| = V^*DV \quad \text{and} \quad |A^*| = UDU^*.$$

Theorem 9.7 *Let A be an n-square matrix. Then for any $u, v \in \mathbb{C}^n$,*

$$|(Au, v)|^2 \leq (|A|u, u)(|A^*|v, v) \tag{9.8}$$

and

$$|((A \circ A^*)u, u)| \leq ((|A| \circ |A^*|)u, u). \tag{9.9}$$

Proof. It is sufficient, by Theorem 7.29, to observe that

$$\begin{pmatrix} |A| & A^* \\ A & |A^*| \end{pmatrix} = \begin{pmatrix} V^* & 0 \\ 0 & U \end{pmatrix} \begin{pmatrix} D & D \\ D & D \end{pmatrix} \begin{pmatrix} V & 0 \\ 0 & U^* \end{pmatrix} \geq 0.$$

and, by Theorem 7.21,

$$\begin{pmatrix} |A| \circ |A^*| & A^* \circ A \\ A \circ A^* & |A^*| \circ |A| \end{pmatrix} = \begin{pmatrix} |A| & A^* \\ A & |A^*| \end{pmatrix} \circ \begin{pmatrix} |A^*| & A \\ A^* & |A| \end{pmatrix} \geq 0.$$

Note that $|A| \circ |A^*| = |A^*| \circ |A|$, with $u = v$, gives (9.9). ∎

For more results, consider Hermitian matrices A decomposed as

$$A = U^* \operatorname{diag}(\lambda_1, \dots, \lambda_n)U,$$

where U is unitary. Define A^α for $\alpha \in \mathbb{R}$, if each λ_i^α makes sense, as

$$A^\alpha = U^* \operatorname{diag}(\lambda_1^\alpha, \dots, \lambda_n^\alpha)U.$$

Theorem 9.8 *Let $A \in \mathbb{M}_n$. Then for any $\alpha \in (0, 1)$,*

$$|(Au, v)| \leq \||A|^\alpha u\| \, \||A^*|^{1-\alpha} v\|, \quad u, \ v \in \mathbb{C}^n. \tag{9.10}$$

Proof. Let $A = U|A|$ be a polar decomposition of A. Then

$$A = U|A|^{1-\alpha}U^*U|A|^\alpha = |A^*|^{1-\alpha}U|A|^\alpha.$$

By the Cauchy–Schwarz inequality, we have

$$|(Au, v)| = |(U|A|^\alpha u, |A^*|^{1-\alpha}v)| \le \||A|^\alpha u\| \, \||A^*|^{1-\alpha}v\|. \quad \blacksquare$$

Note that (9.8) is a special case of (9.10) by taking $\alpha = \frac{1}{2}$.

Theorem 9.9 *Let $A \in \mathbb{M}_n$ and $\alpha \in \mathbb{R}$ be different from $\frac{1}{2}$. If*

$$|(Au, u)| \le (|A|u, u)^\alpha (|A^*|u, u)^{1-\alpha}, \quad \textit{for all } u \in \mathbb{C}^n, \tag{9.11}$$

then A is normal. The converse is obviously true.

Proof. We first consider the case where A is nonsingular.

Let $A = UDV$ be a singular value decomposition of A, where D is diagonal and invertible, and U and V are unitary. With $|A| = V^*DV$ and $|A^*| = UDU^*$, the inequality in (9.11) becomes

$$|(UDVu, u)| \le (V^*DVu, u)^\alpha (UDU^*u, u)^{1-\alpha}$$

or

$$|(D^{1/2}Vu, D^{1/2}U^*u)| \le (D^{1/2}Vu, D^{1/2}Vu)^\alpha (D^{1/2}U^*u, D^{1/2}U^*u)^{1-\alpha}.$$

For any nonzero $u \in \mathbb{C}^n$, set

$$y = \frac{1}{||D^{1/2}U^*u||} D^{1/2}U^*u.$$

Then $||y|| = 1$, and y ranges over all unit vectors as u runs over all nonzero vectors. By putting $\hat{A} = D^{1/2}VUD^{-1/2}$, we have

$$|(\hat{A}y, y)| \le (\hat{A}y, \hat{A}y)^\alpha, \quad \text{for all unit } y \in \mathbb{C}^n.$$

Applying Problem 19 of Section 6.1 to $\hat{A}$, we see $\hat{A}$ is unitary. Thus,

$$D^{-1/2}U^*V^*DVUD^{-1/2} = I.$$

It follows that

$$UDU^* = V^*DV.$$

By squaring both sides, we have $AA^* = A^*A$, or A is normal.

We next deal with the case where A is singular using mathematical induction on n. If $n = 1$, we have nothing to prove. Suppose that the assertion is true for $(n-1)$-square matrices.

Noting that (9.11) still holds when A is replaced by U^*AU for any unitary matrix U, we assume, without loss of generality, that

$$A = \begin{pmatrix} A_1 & b \\ 0 & 0 \end{pmatrix},$$

where $A_1 \in \mathbb{M}_{n-1}$ and b is an $(n-1)$-column vector.

If $b = 0$, then A is normal by induction on A_1. If $b \neq 0$, we take $u_1 \in \mathbb{C}^{n-1}$ such that $(b, u_1) \neq 0$. Let $u = \binom{u_1}{u_2}$ with $u_2 > 0$. Then

$$|(Au, u)| = |(A_1u_1, u_1) + (b, u_1)u_2|$$

and

$$(|A^*|u, u) = ((A_1A_1^* + bb^*)^{1/2}u_1, u_1)$$

which is independent of u_2. To compute $(|A|u, u)$, we write

$$|A| = \begin{pmatrix} C & d \\ d^* & \beta \end{pmatrix}.$$

Then

$$b^*b = d^*d + \beta^2.$$

Hence, $\beta \neq 0$; otherwise $d = 0$ and thus $b = 0$. Therefore, $\beta > 0$ and

$$(|A|u, u) = (Cu_1, u_1) + u_2((d, u_1) + (u_1, d)) + \beta u_2^2.$$

Letting $u_2 \to \infty$ in (9.11) implies $2\alpha \geq 1$ or $\alpha \geq \frac{1}{2}$.

With A replaced by A^* and α by $1 - \alpha$, we can rewrite (9.11) as

$$|(A^*u, u)| \leq (|A^*|u, u)^{1-\alpha}(|A|u, u)^{1-(1-\alpha)}.$$

Applying the same argument to A^*, one obtains $\alpha \leq \frac{1}{2}$. By the induction hypothesis, we see that A is normal if $\alpha \neq \frac{1}{2}$. ∎

Problems

1. For what value(s) of $x \in \mathbb{C}$ is the matrix $A = \left(\begin{smallmatrix} 0 & 1 \\ x & 0 \end{smallmatrix}\right)$ normal? Hermitian? Diagonalizable? Show that A^n is always normal if n is even.

2. Let A be a square complex matrix. Show that

$$\operatorname{tr}|A| = \operatorname{tr}|A^*| \geq |\operatorname{tr} A| \quad \text{and} \quad \det|A| = \det|A^*| = |\det A|.$$

3. Give an example showing that the unitary matrix U in the decomposition $A = U^* \operatorname{diag}(\lambda_1, \ldots, \lambda_n)U$ is not unique. Show that the definition $A^\alpha = U^* \operatorname{diag}(\lambda_1^\alpha, \ldots, \lambda_n^\alpha)U$ for the Hermitian A and real α is independent of choices of unitary matrices U.

4. Let $A \in \mathbb{M}_n$, $H = \frac{1}{2}(A + A^*)$, and $S = \frac{1}{2}(A - A^*)$. Let $\lambda_t = a_t + b_t i$ be the eigenvalues of A, where a_t, b_t are real, $t = 1, \ldots, n$. Show that

$$\sum_{t=1}^n |a_t|^2 \leq \|H\|_2^2, \qquad \sum_{t=1}^n |b_t|^2 \leq \|S\|_2^2.$$

5. Show that for any n-square complex matrix A

$$\begin{pmatrix} |A| & A^* \\ A & |A^*| \end{pmatrix} \geq 0, \quad \text{but} \quad \begin{pmatrix} |A| & A \\ A^* & |A^*| \end{pmatrix} \not\geq 0$$

in general. Conclude that

$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geq 0 \quad \not\Rightarrow \quad \begin{pmatrix} A & B^* \\ B & C \end{pmatrix} \geq 0.$$

Show that it is always true that

$$\begin{pmatrix} A & B \\ B^* & A \end{pmatrix} \geq 0 \quad \Rightarrow \quad \begin{pmatrix} A & B^* \\ B & A \end{pmatrix} \geq 0.$$

6. For any n-square complex matrices A and B, show that

$$\begin{pmatrix} |A| + |B| & A^* + B^* \\ A + B & |A^*| + |B^*| \end{pmatrix} \geq 0.$$

Derive the determinant inequality

$$(\det|A + B|)^2 \leq \det(|A| + |B|)\det(|A^*| + |B^*|).$$

In particular,

$$\det|A + A^*| \leq \det(|A| + |A^*|).$$

Discuss the analogue for the Hadamard product.

7. Show that an n-square complex matrix A is normal if and only if

$$|u^*Au| \le u^*|A|u \quad \text{for all } \ u \in \mathbb{C}^n.$$

8. Let A be an n-square complex matrix and $\alpha \in [0,1]$. Show that

$$\begin{pmatrix} |A|^{2\alpha} & A^* \\ A & |A^*|^{2(1-\alpha)} \end{pmatrix} \ge 0.$$

9. Let A and B be n-square complex matrices. Show that

$$A^*A = B^*B \quad \Leftrightarrow \quad |A| = |B|.$$

Is it true that

$$A^*A \ge B^*B \quad \Leftrightarrow \quad |A| \ge |B|?$$

Prove or disprove

$$|A| \ge |B| \quad \Leftrightarrow \quad |A^*| \ge |B^*|.$$

10. Let $[A]$ denote a principal submatrix of a square matrix A. Show by example that $|[A]|$ and $[|A|]$ are not comparable; that is,

$$|[A]| \not\ge [|A|] \quad \text{and} \quad [|A|] \not\ge |[A]|.$$

But the inequalities below hold, assuming the inverses involved exist:

(a) $[|A|^2] \ge [|A|]^2$,
(b) $[|A|^2] \ge |[A]|^2$,
(c) $[|A|^{1/2}] \le [|A|]^{1/2}$,
(d) $[|A|^{-1/2}] \ge [|A|]^{-1/2}$,
(e) $[|A|^{-1}] \le [|A|^{-2}]^{1/2}$.

11. If A and B are normal matrices such that AB is normal, show that BA is normal. Construct an example showing that matrices A, B, AB, BA are all normal, but $AB \ne BA$.

9.4 Normal Matrix Perturbation

Given two square matrices, how "close" are the matrices in terms of their eigenvalues? More interestingly, if a matrix is "perturbed" a little bit, how would the eigenvalues of the matrix change? In this section we present three results on normal matrix perturbations. The first one is on comparison of $|A|-|B|$ and $A-B$ in terms of a norm, and the second one is on the difference (closeness) of the eigenvalues of the normal matrices in certain order. The third result is of the same kind except that only one matrix is required to be normal.

Theorem 9.10 (Kittaneh) *If A, B are $n\times n$ normal matrices, then*

$$\|\,|A|-|B|\,\|_2 \le \|A-B\|_2.$$

Proof. By the spectral theorem, let $A = U^* \operatorname{diag}(\lambda_1,\dots,\lambda_n)U$ and $B = V^* \operatorname{diag}(\mu_1,\dots,\mu_n)V$, where U and V are unitary matrices. Then $|A| = U^* \operatorname{diag}(|\lambda_1|,\dots,|\lambda_n|)U$, $|B| = V^* \operatorname{diag}(|\mu_1|,\dots,|\mu_n|)V$. For simplicity, let $C = \operatorname{diag}(|\lambda_1|,\dots,|\lambda_n|)$, $D = \operatorname{diag}(|\mu_1|,\dots,|\mu_n|)$ and $W = (w_{ij}) = UV^*$. Then, upon computation, we have

$$\begin{aligned}
\|\,|A|-|B|\,\|_2 &= \|U^*CU - V^*DV\|_2 \\
&= \|CUV^* - UV^*D\|_2 \\
&= \|CW - WD\|_2 \\
&= \Big(\sum_{i,j=1}^{n} \big(\,|\lambda_i| - |\mu_j|\,\big)^2 \cdot |w_{ij}|^2\Big)^{1/2} \\
&\le \Big(\sum_{i,j=1}^{n} |\lambda_i - \mu_j|^2 \cdot |w_{ij}|^2\Big)^{1/2} \\
&= \Big(\sum_{i,j=1}^{n} |(\lambda_i - \mu_j)w_{ij}|^2\Big)^{1/2} \\
&= \|\operatorname{diag}(\lambda_1,\dots,\lambda_n)W - W\operatorname{diag}(\mu_1,\dots,\mu_n)\|_2 \\
&= \|A-B\|_2. \quad \blacksquare
\end{aligned}$$

Corollary 9.1 *Let A and B be $n \times n$ complex matrices. Then*

$$\| |A| - |B| \|_2 \leq \sqrt{2}\, \|A - B\|_2.$$

Proof. Let $\hat{A} = \left(\begin{smallmatrix} 0 & A \\ A^* & 0 \end{smallmatrix}\right)$ and $\hat{B} = \left(\begin{smallmatrix} 0 & B \\ B^* & 0 \end{smallmatrix}\right)$. Then $\hat{A}$ and $\hat{B}$ are Hermitian (normal). Applying Theorem 9.10 to $\hat{A}$ and $\hat{B}$, we have

$$\| |\hat{A}| - |\hat{B}| \|_2^2 = \| |A| - |B| \|_2^2 + \| |A^*| - |B^*| \|_2^2 \leq 2\|A - B\|_2^2.$$

The desired inequality follows immediately. ∎

Theorem 9.11 (Hoffman–Wielandt) *Let A and B be $n \times n$ normal matrices having eigenvalues $\lambda_1, \dots, \lambda_n$ and $\mu_1, \dots, \mu_n$, respectively. Then there exists a permutation p on $\{1, 2, \dots, n\}$ such that*

$$\left(\sum_{i=1}^{n} |\lambda_i - \mu_{p(i)}|^2 \right)^{1/2} \leq \|A - B\|_2.$$

Proof. Let $A = U^* \operatorname{diag}(\lambda_1, \dots, \lambda_n) U$ and $B = V^* \operatorname{diag}(\mu_1, \dots, \mu_n) V$ be spectral decompositions of A and B, respectively, where U and V are unitary matrices. For simplicity, denote $E = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, $F = \operatorname{diag}(\mu_1, \dots, \mu_n)$, and $W = (w_{ij}) = UV^*$. Then

$$\begin{aligned} \|A - B\|_2^2 &= \|U^*(EUV^* - UV^*F)V\|_2^2 \\ &= \|EW - WF\|_2^2 \\ &= \sum_{i,j=1}^{n} |\lambda_i - \mu_j|^2 |w_{ij}|^2. \end{aligned} \tag{9.12}$$

Set $G = (|\lambda_i - \mu_j|^2)$ and $S = (|w_{ij}|^2)$. Then (9.12) is rewritten as

$$\|A - B\|_2^2 = \sum_{i,j=1}^{n} |\lambda_i - \mu_j|^2 |w_{ij}|^2 = e^T (G \circ S) e,$$

where $G \circ S$ is the Hadamard (entrywise) product of G and S, and e is the n-column vector all of whose components are 1. Note that S is a doubly stochastic matrix. By the Birkhoff theorem (Theorem 5.21), S is a convex combination of permutation matrices: $S = \sum_{i=1}^{m} t_i P_i$, where all t_i are nonnegative and add up to 1, P_i are permutation matrices. Among all $e^T(G \circ P_i)e$, $i = 1, \dots, m$, suppose $e^T(G \circ P_k)e$ is the smallest for some k. Consider this P_k as a permutation p on the set $\{1, 2, \dots, n\}$. Then

$$\begin{aligned}
\|A - B\|_2^2 &= e^T(G \circ S)e = \sum_{i=1}^{m} t_i e^T(G \circ P_i)e \\
&\geq \sum_{i=1}^{m} t_i e^T(G \circ P_k)e \\
&= e^T(G \circ P_k)e = \sum_{i=1}^{n} |\lambda_i - \mu_{p(i)}|^2. \quad \blacksquare
\end{aligned}$$

The Hoffman–Wielandt theorem requires both matrices be normal. In what follows we present a result in which one matrix is normal and the other is arbitrary. For this, we need a lemma. For convenience, if A is a square matrix, we write $A = U_A + D_A + L_A$, where U_A, D_A, and L_A are the upper part, diagonal, and lower part of A, respectively. For instance, $A = \left(\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 2 \\ 0 & 0 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 & 0 \\ 0 & 4 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 0 & 0 \\ 3 & 0 \end{smallmatrix}\right)$.

Lemma 9.1 *Let A be an $n \times n$ normal matrix. Then*

$$\|U_A\|_2 \leq \sqrt{n-1}\, \|L_A\|_2, \quad \|L_A\|_2 \leq \sqrt{n-1}\, \|U_A\|_2.$$

Proof. Upon computation, we have

$$\begin{aligned}
\|U_A\|_2^2 &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |a_{ij}|^2 \\
&\leq \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (j-i)|a_{ij}|^2 \\
&= \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} (i-j)|a_{ij}|^2 \quad \text{(Problem 12)} \\
&\leq (n-1) \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} |a_{ij}|^2 = (n-1)\|L_A\|_2^2.
\end{aligned}$$

The second inequality follows by applying the argument to A^T. $\blacksquare$

Theorem 9.12 (Sun) *Let A be an $n \times n$ normal matrix having eigenvalues $\lambda_1, \ldots, \lambda_n$ and B be any $n \times n$ matrix having eigenvalues $\mu_1, \ldots, \mu_n$. Then there is a permutation p on $\{1, 2, \ldots, n\}$ such that*

$$\left(\sum_{i=1}^{n} |\lambda_i - \mu_{p(i)}|^2\right)^{1/2} \le \sqrt{n}\, \|A - B\|_2.$$

Proof. By the Schur triangularization theorem (Theorem 3.3), there exists a unitary matrix U such that U^*BU is upper-triangular. Without loss of generality, we may assume that B is already upper-triangular. Thus $D_B = \operatorname{diag}(\mu_1, \ldots, \mu_n)$. Let $C = A - B$. Then

$$A - D_B = C + U_B, \quad U_B = U_A - U_C, \quad L_A = L_C.$$

Because A and D_B are both normal, the Hoffman–Wielandt theorem ensures a permutation p on the index set $\{1, 2, \ldots, n\}$ such that

$$\left(\sum_{i=1}^{n} |\lambda_i - \mu_{p(i)}|^2\right)^{1/2} \le \|A - D_B\|_2 = \|C + U_B\|_2.$$

Now we apply the lemma to get

$$\begin{aligned}
\|C + U_B\|_2^2 &= \|C + U_A - U_C\|_2^2 \\
&= \|L_C + D_C + U_A\|_2^2 \\
&= \|L_C\|_2^2 + \|D_C\|_2^2 + \|U_A\|_2^2 \\
&\le \|L_C\|_2^2 + \|D_C\|_2^2 + (n-1)\|L_A\|_2^2 \\
&= \|L_C\|_2^2 + \|D_C\|_2^2 + (n-1)\|L_C\|_2^2 \\
&\le n\|C\|_2^2 = n\|A - B\|_2^2.
\end{aligned}$$

Taking square roots of both sides yields the desired inequality. ∎

Problems

1. If A is a normal matrix such that $A^2 = A$, show that A is Hermitian.
2. If A is a normal matrix such that $A^3 = A^2$, show that $A^2 = A$.
3. Let $A = (a_{ij})$ be a normal matrix. Show that $|\lambda| \ge \max_{i,j} |a_{ij}|$ for at least one eigenvalue λ of A.

4. Let $A = (a_{ij})$ be a normal matrix. If K is a $k \times k$ principal submatrix of A, we denote by $\Sigma(K)$ the sum of all entries of K. Show that $k|\lambda| \geq \max_K |\Sigma(K)|$ for at least one eigenvalue λ of A.

5. Let $A = (a_{ij})$ be an n-square complex matrix and denote
$$m_1 = \max_{i,j} |a_{ij}|, \quad m_2 = \max_{i,j} |a_{ij} + \bar{a}_{ji}|/2, \quad m_3 = \max_{i,j} |a_{ij} - \bar{a}_{ji}|/2.$$
Show that for any eigenvalue $\lambda = a + bi$ of A, where a and b are real,
$$|\lambda| \leq nm_1, \quad |a| \leq nm_2, \quad |b| \leq nm_3.$$

6. Let $A = (a_{ij}) \in \mathbb{M}_n$ and denote $m = \max_{i,j} |a_{ij}|$. Show that
$$|\det A| \leq (m\sqrt{n}\,)^n.$$

7. Let $A = (a_{ij})$ be an $n \times n$ real matrix and $d = \max_{i,j}\{|a_{ij} - a_{ji}|/2\}$. Let $\lambda = a + bi$ be any eigenvalue of A, where a, b are real. Show that
$$|b| \leq \sqrt{\frac{n(n-1)}{2}}\, d.$$

8. Let $A = (a_{ij})$ be an $n \times n$ normal matrix having eigenvalues $\lambda_t = a_t + b_t i$, where a_t and b_t are real, $t = 1, \ldots, n$. Show that
$$\max_t |a_t| \geq \max_{i,j} \Big|\frac{1}{2}(a_{ij} + \bar{a}_{ij}\Big|, \quad \max_t |b_t| \geq \max_{i,j} \Big|\frac{1}{2}(a_{ij} - \bar{a}_{ij}\Big|.$$

9. Show that Theorem 9.12 is false if not both A and B are normal by the example $A = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 4 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} -1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$.

10. Show that the scalar $\sqrt{2}$ in Corollary 9.1 is best possible by considering $\frac{\| |A| - |B| \|_2}{\|A - B\|_2}$ with $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 1 & x \\ 0 & 0 \end{smallmatrix}\right)$ as $x \to 0$.

11. Let A be a normal matrix partitioned as $\left(\begin{smallmatrix} E & F \\ G & H \end{smallmatrix}\right)$, where E and H are square matrices (of possibly different sizes). Show that $\|F\|_2 = \|G\|_2$.

12. Use Problem 11 to show that for any $n \times n$ normal matrix $X = (x_{ij})$,
$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (j-i)|x_{ij}|^2 = \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} (i-j)|x_{ij}|^2.$$

13. Show that the scalar $\sqrt{n}$ in Theorem 9.12 is best possible by considering the $n \times n$ matrices

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

14. Let A be a Hermitian matrix. If the diagonal entries of A are also the eigenvalues of A, show that A has to be diagonal, i.e., all off-diagonal entries are 0. Prove or disprove such a statement for normal matrices.

15. Let A and B be $n \times n$ normal matrices having eigenvalues $\lambda_1, \dots, \lambda_n$ and $\mu_1, \dots, \mu_n$, respectively. Show that

$$\min_p \max_i |\lambda_i - \mu_{p(i)}| \le n \|A - B\|_\infty,$$

where p represents permutations on $\{1, 2, \dots, n\}$ and $\|A - B\|_\infty$ is the spectral norm of $A - B$, i.e., the largest singular value of $A - B$.

16. Let $A \in \mathbb{M}_n$ and let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of A. The *spread* of A, written $s(A)$, is defined by $s(A) = \max_{i,j} |\lambda_i - \lambda_j|$. Show that

    (a) $s(A) \le \sqrt{2}\|A\|_2$ for any $A \in \mathbb{M}_n$.

    (b) $s(A) \ge \sqrt{3} \max_{i \ne j} |a_{ij}|$ if A is normal.

    (c) $s(A) \ge 2 \max_{i \ne j} |a_{ij}|$ if A is Hermitian.

17. Let A, B, and C be $n \times n$ complex matrices. If all the eigenvalues of B are contained in a disc $\{z \in \mathbb{C} : |z| < r\}$, and all the eigenvalues of A lie outside the disc, i.e., in $\{z \in \mathbb{C} : |z| \ge r + d\}$ for some $d > 0$, show that the matrix equation $AX - XB = C$ has a solution

$$X = \sum_{i=0}^{\infty} A^{-i-1} C B^i.$$

Show further that if A and B are normal, then $\|X\|_{\text{op}} \le \frac{1}{d}\|C\|_{\text{op}}$, where $\|\cdot\|_{\text{op}}$ denotes the operator norm. [Hint: Use Theorem 4.4.]

CHAPTER 10

Majorization and Matrix Inequalities

Introduction: Majorization is an important tool in deriving matrix inequalities of eigenvalues, singular values, and matrix norms. In this chapter we introduce the concept of majorization, present its basic properties, and show a variety of matrix inequalities in majorization.

10.1 Basic Properties of Majorization

Two vectors in $\mathbb{R}^n$ can be compared in different ways. For instance, one vector may be longer than the other one when measured in terms of norm (length); one may dominate the other componentwise. In this section, we introduce the concept of *majorization*, with which we may compare two real vectors and see whose components are "less spread out" or if one vector "contains" or "controls" the other.

Let $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. We rearrange the components of x in decreasing order and obtain a vector $x^{\downarrow} = (x_1^{\downarrow}, x_2^{\downarrow}, \ldots, x_n^{\downarrow})$, where

$$x_1^{\downarrow} \geq x_2^{\downarrow} \geq \cdots \geq x_n^{\downarrow}.$$

Similarly, let $x_1^{\uparrow} \leq x_2^{\uparrow} \leq \cdots \leq x_n^{\uparrow}$ denote the components of x in increasing order and write $x^{\uparrow} = (x_1^{\uparrow}, x_2^{\uparrow}, \ldots, x_n^{\uparrow})$. For example,

$$x = (-1, -2, 3), \quad x^{\downarrow} = (3, -1, -2), \quad x^{\uparrow} = (-2, -1, 3).$$

F. Zhang, *Matrix Theory: Basic Results and Techniques*, Universitext,
DOI 10.1007/978-1-4614-1099-7_10,

For $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$, if

$$\sum_{i=1}^{k} x_i^{\downarrow} \leq \sum_{i=1}^{k} y_i^{\downarrow}, \quad k = 1, 2, \ldots, n-1, \tag{10.1}$$

and

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \tag{10.2}$$

we say that y *majorizes* x or x is *majorized* by y, written as $x \prec y$ or $y \succ x$. If the equality (10.2) is replaced with the inequality $\sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n} y_i$, we say that y *weakly majorizes* x or x is *weakly majorized* by y, denoted by $x \prec_w y$ or $y \succ_w x$. Obviously, $x \prec y \Rightarrow x \prec_w y$. As an example, take $x = (-1, 0, 1)$, $y = (3, -2, -1)$, and $z = (3, 0, 0)$. Then $x \prec y$ and $y \prec_w z$. Of course, $x \prec_w z$.

Note that the positions of the components in the vectors are unimportant for majorization; if a vector x is majorized by y, then any vector of reordering the components of x is also majorized by y. The inequalities in (10.1) may be rewritten in the equivalent form:

$$\max_{1 \leq i_1 < \cdots < i_k \leq n} \sum_{t=1}^{k} x_{i_t} \leq \max_{1 \leq i_1 < \cdots < i_k \leq n} \sum_{t=1}^{k} y_{i_t}, \quad k = 1, 2, \ldots, n-1.$$

For the case of $n = 2$, intuitively, the set $\{x \in \mathbb{R}^2 : x \prec y\}$ for a given $y \in \mathbb{R}^2$ is the line segment joining $y^{\downarrow}$ and $y^{\uparrow}$.

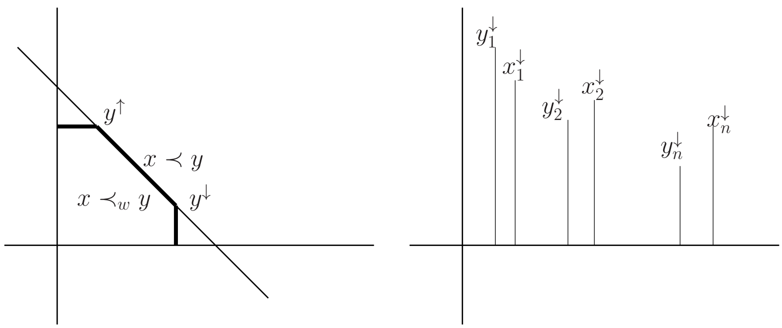

Figure 10.8: Majorization

For $x = (x_1, \ldots, x_n)$, let $|x| = (|x_1|, \ldots, |x_n|)$. For $x, y \in \mathbb{R}^n$, $x+y$ and $x \circ y$ are, respectively, the componentwise sum and product of x and y. Apparently, $x \le y$ (componentwise) implies $x \prec_w y$. For the sake of convenience, sometimes we simply write $x \in \mathbb{R}^n$ to mean that $x = (x_1, x_2, \ldots, x_n)$, where each $x_i \in \mathbb{R}$.

It is easily checked from the definitions that majorization $\prec$ and weak majorization $\prec_w$ are transitive binary relations on $\mathbb{R}^n$:

$$x \prec y,\ y \prec z \ \Rightarrow\ x \prec z; \qquad x \prec_w y,\ y \prec_w z \ \Rightarrow\ x \prec_w z.$$

Theorem 10.1 *Let x, y, $z \in \mathbb{R}^n$. Then*

1. $x \prec_w y \Rightarrow x_1^\downarrow \le y_1^\downarrow$ *and* $x \prec y \Rightarrow y_n^\downarrow \le x_i \le y_1^\downarrow$ *for all* x_i.
2. $x \prec z,\ y \prec z \ \Rightarrow\ px + qy \prec z$, *where* p, $q \ge 0$, $p + q = 1$.
3. $x \prec_w z,\ y \prec_w z \ \Rightarrow\ px + qy \prec_w z$, *where* p, $q \ge 0$, $p + q = 1$.
4. $x \prec y \ \Leftrightarrow\ x \prec_w y$ *and* $-x \prec_w -y$.
5. $x \prec y,\ y \prec x \ \Leftrightarrow\ x = yP$ *for some permutation matrix* P.
6. $x \prec_w y,\ y \prec_w x \ \Leftrightarrow\ x = yP$ *for some permutation matrix* P.

Proof. The first part of (1) is obvious from the definition by taking $k = 1$. For the second part of (1), we show $y_n^\downarrow \le x_n^\downarrow$. $x \prec y$ reveals

$$\sum_{i=1}^{n} x_i^\downarrow = \sum_{i=1}^{n} y_i^\downarrow, \quad \sum_{i=1}^{n-1} x_i^\downarrow \le \sum_{i=1}^{n-1} y_i^\downarrow.$$

Subtracting the inequality from the equality yields $x_n^\downarrow \ge y_n^\downarrow$.

(2) and (3) are similar. We show (3). Let $u = px + qy$. Then $u \prec_w z$ is equivalent to, for any $k = 1, 2, \ldots, n$,

$$\begin{aligned} \sum_{i=1}^{k} u_i^\downarrow &= \sum_{i=1}^{k} (px+qy)_i^\downarrow \le \sum_{i=1}^{k} (p\, x_i^\downarrow + q\, y_i^\downarrow) \\ &= p \sum_{i=1}^{k} x_i^\downarrow + q \sum_{i=1}^{k} y_i^\downarrow \le (p+q) \sum_{i=1}^{k} z_i^\downarrow = \sum_{i=1}^{k} z_i^\downarrow. \end{aligned}$$

We now show (4). If $x \prec y$, then $x \prec_w y$. Moreover, $\sum_{i=1}^n x_i^\downarrow = \sum_{i=1}^n y_i^\downarrow$ and $\sum_{i=1}^k x_i^\downarrow \le \sum_{i=1}^k y_i^\downarrow$ for all $k < n$. Subtracting the inequality from the equality yields $\sum_{i=1}^{n-k} x_{n-i+1}^{\downarrow} \geq \sum_{i=1}^{n-k} y_{n-i+1}^{\downarrow}$. Thus, $\sum_{i=1}^{n-k} -(x_{n-i+1}^{\downarrow}) \leq \sum_{i=1}^{n-k} -(y_{n-i+1}^{\downarrow})$. Noticing that

$$-(x_{n-i+1}^{\downarrow}) = (-x)_i^{\downarrow}, \quad -(y_{n-i+1}^{\downarrow}) = (-y)_i^{\downarrow},$$

we have $\sum_{i=1}^{n-k}(-x)_i^{\downarrow} \leq \sum_{i=1}^{n-k}(-y)_i^{\downarrow}$ for all $k = n-1, \ldots, 2, 1$. With $\sum_{i=1}^{n}(-x)_i^{\downarrow} = \sum_{i=1}^{n}(-y)_i^{\downarrow}$, we conclude $-x \prec_w -y$. For the converse, it is sufficient to show that $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$. We have $x \prec_w y$ implies $\sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n} y_i$. On the other hand, $-x \prec_w -y$ reveals the reversed inequality. Thus, the equality has to hold. So $x \prec y$.

(5) and (6) are similar. We show (5). If $x \prec y$ and $y \prec x$, then $x_1^{\downarrow} \leq y_1^{\downarrow}$ and $y_1^{\downarrow} \leq x_1^{\downarrow}$. Thus $x_1^{\downarrow} = y_1^{\downarrow}$. Let $\tilde{x}$ and $\tilde{y}$ be the vectors obtained from x and y by deleting $x_1^{\downarrow}$ and $y_1^{\downarrow}$, respectively. Then $\tilde{x} \prec \tilde{y}$ and $\tilde{y} \prec \tilde{x}$. From the above argument, we have $\tilde{x}_1^{\downarrow} = \tilde{y}_1^{\downarrow}$, i.e., $x_2^{\downarrow} = y_2^{\downarrow}$. Inductively, $x_i^{\downarrow} = y_i^{\downarrow}$ for all i. This says that the components of x are rearrangements of the components of y; that is, $x = yP$ for some permutation matrix P. The converse is trivial. ∎

Our next theorem best characterizes the relationship between the weak majorization $\prec_w$ and the majorization $\prec$ via the componentwise dominance $\leq$. This theorem is used repeatedly.

Theorem 10.2 *The following statements are equivalent.*

1. *$x \prec_w y$, where $x, y \in \mathbb{R}^n$.*
2. *$x \leq z$ and $z \prec y$ for some $z \in \mathbb{R}^n$.*
3. *$x \prec u$ and $u \leq y$ for some $u \in \mathbb{R}^n$.*

Proof. (1)⇔(2): It is easy to see that (2)⇒(1). We show the converse by induction on the number of components. If $n = 1$, it is obvious.

Let $n > 1$ and suppose it is true for vectors with less than n components. We may assume that the components of x and y are already in decreasing order. Let $\epsilon = \min_k\{\sum_{i=1}^{k}(y_i - x_i)\} \geq 0$ and $\tilde{x} = x + (\epsilon, 0, \ldots, 0)$. Then $\tilde{x} = \tilde{x}^{\downarrow}$, $x \leq \tilde{x}$, and $\tilde{x} \prec_w y$, as for each p,

$$\sum_{i=1}^{p} \tilde{x}_i = \sum_{i=1}^{p} x_i + \epsilon \leq \sum_{i=1}^{p} x_i + \sum_{i=1}^{p}(y_i - x_i) = \sum_{i=1}^{p} y_i.$$

If ϵ is attained when $k = m$, then $\sum_{i=1}^{m} \tilde{x}_i = \sum_{i=1}^{m} y_i$. This yields

$$(\tilde{x}_1, \ldots, \tilde{x}_m) \prec (y_1, \ldots, y_m), \quad (\tilde{x}_{m+1}, \ldots, \tilde{x}_n) \prec_w (y_{m+1}, \ldots, y_n).$$

By induction, we have a real vector $(z_{m+1}, \ldots, z_n)$ such that

$$(\tilde{x}_{m+1}, \ldots, \tilde{x}_n) \leq (z_{m+1}, \ldots, z_n) \prec (y_{m+1}, \ldots, y_n).$$

Set $z = (\tilde{x}_1, \ldots, \tilde{x}_m, z_{m+1}, \ldots, z_n)$. This z serves the purpose.

(1)⇔(3): One easily checks that (3)⇒(1). We show the converse. Let y_k be the smallest component of y. Let $\delta = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i$ and $u = y - \delta e_k$, where $e_k \in \mathbb{R}^n$ has component 1 in the kth position and 0 elsewhere. Then it is easy to verify that $x \prec u$ and $u \leq y$. ∎

Theorem 10.3 *Let x, $y \in \mathbb{R}^m$ and u, $v \in \mathbb{R}^n$. Then*

1. $x \prec y,\ u \prec v \ \Rightarrow\ (x, u) \prec (y, v)$.
2. $x \prec_w y,\ u \prec_w v \ \Rightarrow\ (x, u) \prec_w (y, v)$.
3. $x \prec y,\ u \prec v \ \Rightarrow\ x + u \prec y^{\downarrow} + v^{\downarrow}$ *(when $m = n$).*
4. $x \prec_w y,\ u \prec_w v \ \Rightarrow\ x + u \prec_w y^{\downarrow} + v^{\downarrow}$ *(when $m = n$).*

Proof. (1) is similar to (2) and (3) is similar to (4). We show (2) and (4). Let $\tilde{x} = (x, u)$ and $\tilde{y} = (y, v)$. For positive integer $k \leq n$, suppose that the first k largest components of $\tilde{x}$ consist of $x_1^{\downarrow}, \ldots, x_r^{\downarrow}$ and $u_1^{\downarrow}, \ldots, u_s^{\downarrow}$, $r + s = k$. Since $x \prec_w y$ and $u \prec_w v$, we have

$$\sum_{i=1}^{k} \tilde{x}_i^{\downarrow} = \sum_{i=1}^{r} x_i^{\downarrow} + \sum_{i=1}^{s} u_i^{\downarrow} \leq \sum_{i=1}^{r} y_i^{\downarrow} + \sum_{i=1}^{s} v_i^{\downarrow} \leq \sum_{i=1}^{k} \tilde{y}_i^{\downarrow}.$$

This says that $\tilde{x}$ is weakly majorized by $\tilde{y}$, i.e., $(x, u) \prec_w (y, v)$.

(4) is proven in a similar way by checking that

$$\begin{aligned} \sum_{i=1}^{k} (x+u)_i^{\downarrow} &\leq \sum_{i=1}^{k} x_i^{\downarrow} + \sum_{i=1}^{k} u_i^{\downarrow} \leq \sum_{i=1}^{k} y_i^{\downarrow} + \sum_{i=1}^{k} v_i^{\downarrow} \\ &= \sum_{i=1}^{k} (y_i^{\downarrow} + v_i^{\downarrow}) = \sum_{i=1}^{k} (y^{\downarrow} + v^{\downarrow})_i \\ &= \sum_{i=1}^{k} (y^{\downarrow} + v^{\downarrow})_i^{\downarrow}. \ \blacksquare \end{aligned}$$

Theorem 10.4 *Let* $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$. *Then*

1. $x \prec_w |x|$.
2. $|x + y| \prec_w |x|^{\downarrow} + |y|^{\downarrow}$.
3. $x^{\downarrow} + y^{\uparrow} \prec x + y \prec x^{\downarrow} + y^{\downarrow}$.
4. $\sum_{i=1}^n x_i^{\downarrow} y_i^{\uparrow} \le \sum_{i=1}^n x_i y_i \le \sum_{i=1}^n x_i^{\downarrow} y_i^{\downarrow}$.

Proof. (1) is due to the fact that for each $k = 1, 2, \ldots, n$

$$\sum_{i=1}^k x_i^{\downarrow} \le \sum_{i=1}^k |x_i^{\downarrow}| \le \sum_{i=1}^k |x|_i^{\downarrow}.$$

In a similar way, that $\sum_{i=1}^k |x + y|_i^{\downarrow} \le \sum_{i=1}^k (|x|_i^{\downarrow} + |y|_i^{\downarrow})$ implies (2).

The second inequality in (3) is a consequence of Theorem 10.3 (3). To show the first one, we may assume that $x = x^{\downarrow}$. Consider the case $n = 2$ first. If $y_1 \le y_2$, then $y = y^{\uparrow}$ and $x^{\downarrow} + y^{\uparrow} = x + y \prec x + y$. If $y_1 > y_2$, then $y = y^{\downarrow}$ and $x + y = x^{\downarrow} + y^{\downarrow}$. As $y^{\uparrow} \prec y^{\downarrow}$, by Theorem 10.3 (3), we have $x^{\downarrow} + y^{\uparrow} \prec x^{\downarrow} + y^{\downarrow} = x + y$.

Let $n > 2$. Our goal is to obtain $y^{\uparrow}$ by repeatedly exchanging components y so that they are in increasing order. If $y = y^{\uparrow}$, there is nothing to prove. Suppose $y_i > y_j$ for some i and j, $i < j$. Switch the components y_i and y_j in y and denote the resulting vector by $\tilde{y}$. Note that a pair of components y_i, y_j in y now are in increasing order y_j, y_i in $\tilde{y}$. Observe that $x + \tilde{y}$ and $x + y$ differ by two components. By Theorem 10.3 (1) and the above argument for the case of $n = 2$, we see $x + \tilde{y} \prec x + y$. If $\tilde{y} = y^{\uparrow}$, we are done. Otherwise, by the same argument, we have $\hat{y}$ so that $x + \hat{y} \prec x + \tilde{y}$ and $\hat{y}$ has two more components than $\tilde{y}$ that are in increasing order. Repeating the process reveals $x^{\downarrow} + y^{\uparrow} \prec \cdots \prec x + \hat{y} \prec x + \tilde{y} \prec x + y$.

For (4), again, we assume $x = x^{\downarrow}$ and show the first inequality; the proof for the second one is similar. If $y = y^{\uparrow}$, then we have nothing to prove. Otherwise, let $\tilde{y}$ be the vector as above. Compute

$$x_i^{\downarrow} y_j + x_j^{\downarrow} y_i - x_i^{\downarrow} y_i - x_j^{\downarrow} y_j = (x_i^{\downarrow} - x_j^{\downarrow})(y_j - y_i) \le 0;$$

that is,

$$x_i^{\downarrow} \tilde{y}_i + x_j^{\downarrow} \tilde{y}_j = x_i^{\downarrow} y_j + x_j^{\downarrow} y_i \le x_i^{\downarrow} y_i + x_j^{\downarrow} y_j.$$

This yields $\sum_{t=1}^n x_t^{\downarrow}\tilde{y}_t \le \sum_{t=1}^n x_t^{\downarrow} y_t$. Likewise, $\sum_{t=1}^n x_t^{\downarrow}\hat{y}_t \le \sum_{t=1}^n x_t^{\downarrow}\tilde{y}_t$. Repeat this process until $y^{\uparrow}$ is obtained. Thus (4) follows. ∎

The inequalities in Theorem 10.4 (4) can be generalized to majorization inequalities when the components of x and y are all nonnegative; in other words, the upper limit n for the summation can be replaced by $k = 1, 2, \ldots, n$. (See Theorem 10.16.)

Denote by $\mathbb{R}_+^n$ the set of all vectors in $\mathbb{R}^n$ with nonnegative components; that is, $u = (u_1, u_2, \ldots, u_n) \in \mathbb{R}_+^n$ means that all $u_i \ge 0$.

Theorem 10.5 *Let $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$. Then*

$$x \prec y \;\Leftrightarrow\; \sum_{i=1}^n x_i^{\downarrow} u_i^{\downarrow} \le \sum_{i=1}^n y_i^{\downarrow} u_i^{\downarrow} \text{ for all } u = (u_1, \ldots, u_n) \in \mathbb{R}^n$$

and

$$x \prec_w y \;\Leftrightarrow\; \sum_{i=1}^n x_i^{\downarrow} u_i^{\downarrow} \le \sum_{i=1}^n y_i^{\downarrow} u_i^{\downarrow} \text{ for all } u = (u_1, \ldots, u_n) \in \mathbb{R}_+^n.$$

Proof. We show the one for weak majorization. The other one is similar. "$\Leftarrow$" is immediate by setting $u = (1, \ldots, 1, 0, \ldots, 0)$ in which there are k 1s, $k = 1, 2, \ldots, n$. For "$\Rightarrow$", let $t_i = y_i^{\downarrow} - x_i^{\downarrow}$ for each i. Then $x \prec_w y$ implies $\sum_{i=1}^k t_i \ge 0$ for $k = 1, 2, \ldots, n$. Compute

$$\begin{aligned}
\sum_{i=1}^n y_i^{\downarrow} u_i^{\downarrow} - \sum_{i=1}^n x_i^{\downarrow} u_i^{\downarrow} = \sum_{i=1}^n t_i u_i^{\downarrow} & \\
= \;& t_1(u_1^{\downarrow} - u_2^{\downarrow}) + (t_1 + t_2)(u_2^{\downarrow} - u_3^{\downarrow}) + \cdots \\
+ \;& (t_1 + \cdots + t_{n-1})(u_{n-1}^{\downarrow} - u_n^{\downarrow}) + (t_1 + \cdots + t_n)u_n^{\downarrow} \ge 0.
\end{aligned}$$

Therefore, the desired inequality follows. ∎

Theorem 10.6 *Let x, y, u, $v \in \mathbb{R}_+^n$. Then*

1. $x \prec_w y \;\Rightarrow\; x \circ u \prec_w y^{\downarrow} \circ u^{\downarrow}$.
2. $x \prec_w u,\ y \prec_w v \;\Rightarrow\; x \circ y \prec_w u^{\downarrow} \circ v^{\downarrow}$.

Proof. We first show that $(x_1u_1, \dots, x_nu_n) \prec_w (y_1^\downarrow u_1^\downarrow, \dots, y_n^\downarrow u_n^\downarrow)$. By setting $u_{k+1}^\downarrow = \cdots = u_n^\downarrow = 0$ in Theorem 10.5, we obtain

$$\sum_{i=1}^{k} x_i^\downarrow u_i^\downarrow \le \sum_{i=1}^{k} y_i^\downarrow u_i^\downarrow, \quad k = 1, \dots, n. \tag{10.3}$$

Note that all components of x, y, and u are nonnegative. For any positive integer $k \le n$ and sequence $1 \le i_1 < \cdots < i_k \le n$, we have $x_{i_t} \le x_t^\downarrow$ and $u_{i_t} \le u_t^\downarrow$, $t = 1, \dots, k$. It follows that

$$\sum_{t=1}^{k} x_{i_t} u_{i_t} \le \sum_{t=1}^{k} x_t^\downarrow u_t^\downarrow, \quad k = 1, \dots, n.$$

With (10.3), (1) is proven. For (2), apply (1) twice. ∎

Problems

1. Find two vectors x, $y \in \mathbb{R}^3$ such that neither $x \prec y$ nor $x \succ y$ holds.
2. Let $y = (2, 1) \in \mathbb{R}^2$. Sketch the following sets in the plane $\mathbb{R}^2$:
$$\{x \in \mathbb{R}^2 : x \prec y\} \quad \text{and} \quad \{x \in \mathbb{R}^2 : x \prec_w y\}.$$
3. Let x, $y \in \mathbb{R}^n$. Show that $(-x)^\downarrow = -(x^\uparrow)$ and $x - y \prec x^\downarrow - y^\uparrow$.
4. Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Show that $x_i^\downarrow = x_{n-i+1}^\uparrow$, $i = 1, 2, \dots, n$.
5. Let $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ be in $\mathbb{R}^n$. Show that $x \prec_w y \Leftrightarrow (x_1^\downarrow, x_2^\downarrow, \dots, x_k^\downarrow) \prec_w (y_1^\downarrow, y_2^\downarrow, \dots, y_k^\downarrow)$, $k = 1, 2, \dots, n$.
6. Let $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$. If $x_1 \ge x_2 \ge \cdots \ge x_n$, show that $\frac{1}{m}\sum_{i=1}^m x_i \ge \frac{1}{n}\sum_{i=1}^n x_i$ for any positive integer $m \le n$.
7. Let $a_1, a_2, \dots, a_n$ be nonnegative numbers such that $\sum_{i=1}^n a_i = 1$. Show that $(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}) \prec (a_1, a_2, \dots, a_n) \prec (1, 0, \dots, 0)$ and that
$$v_n \prec (v_{n-1}, 0) \prec \cdots \prec (v_2, 0, \dots, 0) \prec (1, 0, \dots, 0),$$
where $v_k = (\frac{1}{k}, \frac{1}{k}, \dots, \frac{1}{k})$ with k copies of $\frac{1}{k}$ for $k = 1, 2, \dots, n$.
8. Referring to Theorem 10.6 (2), can $u^\downarrow \circ v^\downarrow$ be replaced by $u \circ v$?
9. Let x, $y \in \mathbb{R}^n$. Show that $x \prec y \Leftrightarrow -x \prec -y$. If $x \prec_w y$, does it necessarily follow that $-x \prec_w -y$ (or $-y \prec_w -x$)?

10. Let $x, y, z \in \mathbb{R}^n$. If $x + y \prec_w z$, show that $x \prec_w z - y^\uparrow$.

11. Let $x, y \in \mathbb{R}^n$ and $x \le y$ (componentwise). Show that $xP \le yP$ for any $n \times n$ permutation matrix P and consequently $x^\downarrow \le y^\downarrow$.

12. Let $e = (1, 1, \ldots, 1) \in \mathbb{R}^n$. Find all $x \in \mathbb{R}^n$ such that $x \prec e$.

13. Let $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and $\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$. Show that $\bar{x}e \prec x$, where $e = (1, 1, \ldots, 1) \in \mathbb{R}^n$. State the case of $n = 2$.

14. Let $x, y \in \mathbb{R}^n_+$. Show that $(x, y) \prec_w (x + y, 0)$.

15. Let $x, y \in \mathbb{R}^n$. If $x \prec y$ and if $0 \le \alpha \le \beta \le 1$, show that
$$\beta x^\downarrow + (1 - \beta) y^\downarrow \prec \alpha x^\downarrow + (1 - \alpha) y^\downarrow.$$

16. Let $x, y \in \mathbb{R}^n$. Show that $x \prec y \Leftrightarrow (x, z) \prec (y, z)$ for all $z \in \mathbb{R}^m$.

17. Let $x, y \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$. Show that
$$(x, z) \prec_w (y, z) \;\Rightarrow\; x \prec_w y.$$
Consider the more general case. If $(x, u) \prec (y, v)$ for some $u, v \in \mathbb{R}^m$ satisfying $u \prec v$, does it necessarily follow that $x \prec y$ or $x \prec_w y$?

18. Let $x, y \in \mathbb{R}^n$ such that $x \prec_w y$. Show that (i) there exists $\tilde{y} \in \mathbb{R}^n$ which differs from y by at most one component such that $x \prec \tilde{y}$; and (ii) there exist $a, b \in \mathbb{R}$ such that $(x, a) \prec (y, b)$.

19. Let $x, y, z \in \mathbb{R}^n_+$. If $2x \prec_w y^\downarrow + z^\downarrow$, show that
$$(x, x) \prec_w (y, z) \prec_w (y^\downarrow + z^\downarrow, 0).$$

20. Let $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and α be a real number such that $x_n^\downarrow \le \alpha \le x_1^\downarrow$. Let $\beta = x_1 + x_2 + \cdots + x_n$. Show that $(\alpha, \frac{\beta - \alpha}{n-1}, \ldots, \frac{\beta - \alpha}{n-1}) \prec x$.

21. Let $x, y \in \mathbb{R}^n$. If $x \prec y$, show that $y_m^\downarrow \ge x_m^\downarrow \ge y_{m+1}^\downarrow$ for some m.

22. Let $t \in \mathbb{R}$ and denote $t^+ = \max\{t, 0\}$. For $x, y \in \mathbb{R}^n$, if $x \prec_w y$, show that $(x_1^+, \ldots, x_n^+) \prec_w (y_1^+, \ldots, y_n^+)$. Is the converse true?

23. Give an example that Theorem 10.6 is not valid for some $x, y \in \mathbb{R}^n$.

24. Let $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$. If $x \prec y$, show that
$$\sum_{i=1}^n u_i^\downarrow y_i^\uparrow \le \sum_{i=1}^n u_i^\downarrow x_i \le \sum_{i=1}^n u_i^\downarrow x_i^\downarrow \le \sum_{i=1}^n u_i^\downarrow y_i^\downarrow, \quad u = (u_1, \ldots, u_n) \in \mathbb{R}^n.$$

10.2 Majorization and Stochastic Matrices

Recall that a *doubly stochastic matrix* is a square nonnegative matrix whose row sums and column sums are all equal to 1. In symbols, A is doubly stochastic if A is nonnegative and for $e = (1, \dots, 1) \in \mathbb{R}^n$,

$$eA = e \quad \text{and} \quad Ae^T = e^T.$$

In this section, we show a close relation between majorization and this type of matrices. Our goal is to present two fundamental results: $x \prec y$ if and only if $x = yP$ for some doubly stochastic matrix P; $x \prec_w y$ if and only if $x = yQ$ for some doubly substochastic matrix Q. A *doubly substochastic matrix* is a square nonnegative matrix whose row and column sums are each at most 1, i.e., $eA \le e$, $Ae^T \le e^T$.

Theorem 10.7 *Let A be an $n \times n$ nonnegative matrix. Then A is doubly stochastic if and only if $xA \prec x$ for all (row vectors) $x \in \mathbb{R}^n$.*

Proof. For necessity, by Theorem 5.21, we write A as a convex combination of permutation matrices $P_1, P_2, \dots, P_m$:

$$A = \sum_{i=1}^m \alpha_i P_i, \quad \sum_{i=1}^m \alpha_i = 1, \ \ \alpha_i \ge 0.$$

Because $xP_i \prec x$ for each i, we have

$$xA = \sum_{i=1}^m \alpha_i x P_i \prec \sum_{i=1}^m \alpha_i x = x.$$

For sufficiency, take $x = e = (1, \dots, 1)$. Then $eA \prec e$ says that each column sum of A is at most 1. However, adding up all the column sums of A, one should get n. It follows that every column sum of A is 1. Now set $x = e_i = (0, \dots, 1, \dots, 0)$, where 1 is the ith component of e_i and all other components are 0. Then $e_i A \prec e_i$ means that the ith row sum of A is 1 for $i = 1, 2, \dots, n$. ∎

We next take a closer look at the relation between two vectors when one is majorized by the other. Specifically, we want to see how to get one vector from the other. This can be accomplished by successively applying a finite number of so-called *T-transforms.*

A 2×2 *T-transformation*, or *transform* for short, is a matrix of the form $\begin{pmatrix} t & 1-t \\ 1-t & t \end{pmatrix}$, where $0 \leq t \leq 1$. For higher dimension, we call a matrix a *T-transform* if it is obtained from the identity matrix I by replacing a 2×2 principal submatrix of I with a 2×2 T-transform. It is readily seen that a T-transform is a doubly stochastic matrix and it can be written as $tI+(1-t)P$, where $0 \leq t \leq 1$ and P is a permutation matrix that just interchanges two columns of the identity matrix I. Note that when $t = 0$, the T-transform is a permutation matrix. Since a permutation on the set $\{1, 2, \ldots, n\}$ can be obtained by a sequence of interchanges (Problem 22, Section 5.6), every permutation matrix can be factorized as a product of T-transforms.

Theorem 10.8 *Let x, $y \in \mathbb{R}^n$. Then $x \prec y$ if and only if there exist T-transforms $T_1, \ldots, T_m$ such that $x = yT_1 \cdots T_m$. Consequently,*

$$x \prec y \quad \Leftrightarrow \quad x = yD \ \textit{ for some doubly stochastic matrix } D.$$

Proof. Sufficiency: This is immediate from the previous theorem. To prove the necessity, we use induction on n. If $n = 1$, there is nothing to show. Suppose $n > 1$ and the result is true for $n - 1$.

If $x_1 = y_1$, then $(x_2, \ldots, x_n) \prec (y_2, \ldots, y_n)$. By induction hypothesis, there exist T-transforms $S_1, S_2, \ldots, S_m$, all of order $n - 1$, such that $(x_2, \ldots, x_n) = (y_2, \ldots, y_n)S_1S_2 \cdots S_m$. Let $T_i = \begin{pmatrix} 1 & 0 \\ 0 & S_i \end{pmatrix}$. Then T_i is also a T-transform, $i = 1, 2, \ldots, m$, and $x = yT_1T_2 \cdots T_m$.

If $x_i = y_j$ for some i and j, we may apply permutations on x and y so that x_i and y_j are the first components of the resulting vectors. Since every permutation matrix is a product of T-transforms, the argument reduces to the case $x_1 = y_1$ that we have settled.

Let $x_i \neq y_j$ for all i, j. We may further assume that x and y are in decreasing order. By Problem 21 of Section 10.1, $x \prec y$ implies

$$y_k > x_k > y_{k+1}, \quad \text{for some } k.$$

So

$$x_k = ty_k + (1 - t)y_{k+1}, \quad \text{for some } t \in (0, 1).$$

Let T_0 be the T-transform with $\left(\begin{smallmatrix} t & 1-t \\ 1-t & t \end{smallmatrix}\right)$ lying in the rows k and $k+1$ and columns k and $k+1$. Set $z = yT_0$. Then z and y have the same components except the kth and $(k+1)$th components:

$$z_k = ty_k + (1-t)y_{k+1} = x_k$$

and

$$z_{k+1} = (1-t)y_k + ty_{k+1} = y_k + y_{k+1} - x_k.$$

Note that

$$\sum_{i=1}^{k} x_i \le \sum_{i=1}^{k-1} y_i + x_k = \sum_{i=1}^{k} z_i \le \sum_{i=1}^{k} z_i^{\downarrow}.$$

For $r \neq k$, i.e., $r < k$ or $r > k$, bearing in mind that $z_k + z_{k+1} = y_k + y_{k+1}$, we always have

$$\sum_{i=1}^{r} x_i \le \sum_{i=1}^{r} y_i = \sum_{i=1}^{r} z_i \le \sum_{i=1}^{r} z_i^{\downarrow},$$

and equality holds when $r = n$. Hence, $x \prec z$.

Now x and z have the same component $x_k = z_k$. By the above argument, we have that $x = zT_1T_2 \cdots T_m = yT_0T_1T_2 \cdots T_m$, where T_is are T-transforms. ∎

Every doubly stochastic matrix is a convex combination of (finite) permutation matrices (Theorem 5.21, Section 5.6), therefore we may restate the second part of the theorem as $x \prec y$ if and only if there exist permutation matrices $P_1, \ldots, P_m$ such that $x = t_1yP_1 + \cdots + t_myP_m$, where $t_1 + \cdots + t_m = 1$, and all $t_i \ge 0$. Thus, for given $y \in \mathbb{R}^n$, $\{x : x \prec y\}$ is the convex hull of all points in $\mathbb{R}^n$ obtained from y by permuting its components.

We now study the analogue of Theorem 10.8 for weak majorization. A weak majorization $\prec_w$ becomes the componentwise inequality $\le$ via T-transforms. When the vectors are nonnegative, the weak majorization can be characterized by doubly substochastic matrices.

Theorem 10.9 *Let $x, y \in \mathbb{R}^n$. Then $x \prec_w y$ if and only if there exist T-transforms $T_1, T_2, \ldots, T_m$ such that $x \le yT_1T_2 \cdots T_m$.*

Proof. If $x \prec_w y$, by Theorem 10.2, there exists $z \in \mathbb{R}^n$ such that $x \leq z \prec y$. By Theorem 10.8, $z = yT_1T_2\cdots T_m$, where T_is are T-transforms. It follows that $x \leq yT_1T_2\cdots T_m$.

Conversely, let $u = yT_1T_2\cdots T_m$. Then $x \leq u$, so $x \prec_w u$. However, $u \prec y$ as $T_1T_2\cdots T_m$ is doubly stochastic. Thus, $x \prec_w y$. ∎

Theorem 10.10 *Let $x, y \in \mathbb{R}^n_+$. Then $x \prec_w y$ if and only if $x = yS$ for some doubly substochastic matrix S.*

Proof. Let $x = yS$ for some doubly substochastic matrix S. Then there exists a stochastic matrix P such that $P - S$ is a nonnegative matrix. Thus, $yS \leq yP$ for $y \in \mathbb{R}^n_+$. As $yP \prec y$, $x = yS \leq yP \prec y$.

For the converse, since $x \prec_w y$, by Theorem 10.9, there exist T-transforms $T_1, T_2, \ldots, T_m$ such that $x \leq yT_1T_2\cdots T_m$. Denote $z = yT_1T_2\cdots T_m$. Then $x \leq z$. Note that x is nonnegative. By scaling the components of z to get x; that is, taking r_i so that $x_i = z_ir_i$, $0 \leq r_i \leq 1$, $i = 1, 2, \ldots, n$, we have a diagonal matrix $R = \operatorname{diag}(r_1, r_2, \ldots, r_n)$ such that $x = zR$. Set $S = T_1T_2\cdots T_mR$. Then S is doubly substochastic and $x = yS$. ∎

For given $y \in \mathbb{R}^n_+$, $\{x : x \prec_w y\}$ is the convex hull of all points $(t_1y_{p(1)}, t_2y_{p(2)}, \ldots, t_ny_{p(n)})$, where p runs over all permutations and each t_i is either 0 or 1. Note that both x and y are required to be nonnegative vectors in Theorem 10.10. The conclusion does not necessarily follow otherwise (Problem 14).

Problems

1. Find a doubly stochastic matrix P such that $(1, 2, 3) = (6, 0, 0)P$.

2. Show that $Q = (q_{ij})$ is a doubly substochastic matrix if there exists a doubly stochastic matrix $D = (d_{ij})$ such that $q_{ij} \leq d_{ij}$ for all i, j.

3. Let $A = (a_{ij})$ and $B = (b_{ij})$ be $n \times n$ matrices. Show that

   (a) If A and B are doubly substochastic, then $C = (a_{ij}b_{ij})$ and $D = (\sqrt{a_{ij}b_{ij}}\,)$ are also doubly substochastic.

   (b) If A and B are unitary, then $E = (|a_{ij}b_{ij}|)$ is doubly substochastic, but $F = (\sqrt{|a_{ij}b_{ij}|}\,)$ is not in general.

4. Show that the following matrix A satisfies (i) $eA \geq e$, $Ae^T \geq e^T$; (ii) $A - Q$ is never nonnegative for any doubly stochastic matrix Q:

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad e = (1, 1, 1).$$

5. Show that the following doubly stochastic matrix cannot be expressed as a product of T-transforms:

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

6. Let P be a square matrix. If both P and its inverse P^{-1} are doubly stochastic, show that P is a permutation matrix.

7. Show each of the following statements.

   (a) The (ordinary) product of two doubly stochastic matrices is a doubly stochastic matrix.

   (b) The (ordinary) product of two doubly substochastic matrices is a doubly substochastic matrix.

   (c) The Hadamard product of two doubly substochastic matrices is a doubly substochastic matrix.

   (d) The Kronecker product of two doubly substochastic matrices is a doubly substochastic matrix.

   (e) The convex combination of finite doubly stochastic matrices is a doubly stochastic matrix.

   (f) The convex combination of finite doubly substochastic matrices is a doubly substochastic matrix.

8. Show that any square submatrix of a doubly stochastic matrix is doubly substochastic and that every doubly substochastic matrix can be regarded as a square submatrix of a doubly stochastic matrix.

9. A square (0,1)-matrix is called *sub-permutation* matrix if each row and each column have at most one 1. Show that a matrix is doubly substochastic if and only if it is a convex combination of finite subpermutation matrices.

10. Let $A = (a_{ij})$ be an $n \times n$ doubly stochastic matrix. Show that n entries of A can be chosen from different rows and columns so that their product is positive; that is, there exists a permutation p such that $\prod_{i=1}^{n} a_{ip(i)} > 0$. [Hint: Use the Frobenius–König theorem.]

11. A matrix A of order $n \geq 2$ is said to be *reducible* if $PAP^T = \left(\begin{smallmatrix} B & 0 \\ C & D \end{smallmatrix}\right)$ for some permutation matrix P, where B and D are some square matrices; A is *irreducible* if it is not reducible. A matrix A is said to be *decomposable* if $PAQ = \left(\begin{smallmatrix} B & 0 \\ C & D \end{smallmatrix}\right)$ for some permutation matrices P and Q, where B and D are some square matrices; A is *indecomposable* if it is not decomposable. Prove each of the following statements.

    (a) If A is a nonnegative indecomposable matrix of order n, then the entries of A^{n-1} are all positive.

    (b) The product of two nonnegative indecomposable matrices is indecomposable.

    (c) The product of two nonnegative irreducible matrices need not be irreducible.

12. Let $x, y \in \mathbb{R}^n_+$. Show that $x \prec_w y$ if and only if x is a convex combination of the vectors $yQ_1, yQ_2, \ldots, yQ_m$, where $Q_1, Q_2, \ldots, Q_m$ are subpermutation matrices.

13. Let A be an $n \times n$ nonnegative matrix. Show that A is doubly substochastic if and only if $Ax \prec_w x$ for all column vectors $x \in \mathbb{R}^n_+$.

14. Give an example that $x \in \mathbb{R}^n_+$, $y \in \mathbb{R}^n$, $x = yS$ for some substochastic matrix S, but $x \prec_w y$ does not hold.

15. Let $x, y \in \mathbb{R}^n$. Show that for any real numbers a and b,

$$x \prec y \;\Rightarrow\; (ax_1+b, ax_2+b, \ldots, ax_n+b) \prec (ay_1+b, ay_2+b, \ldots, ay_n+b).$$

16. Let $T = \left(\begin{smallmatrix} 1-a & b \\ a & 1-b \end{smallmatrix}\right)$, $S = \left(\begin{smallmatrix} -a & b \\ a & -b \end{smallmatrix}\right)$, $0 < a < 1$, $0 < b < 1$. Show that

$$T^n = I + \frac{1 - r^n}{1 - r} S, \quad r = 1 - (a + b),$$

for every positive integer n. Find T^n as $n \to \infty$.

. ———————— ⊙ ———————— .

10.3 Majorization and Convex Functions

This section is devoted to majorization and convex functions. Recall that a real-valued function $f(t)$ defined on an interval of $\mathbb{R}$ is said to be *increasing* if $x \leq y$ implies $f(x) \leq f(y)$, and *convex* if for all x, y in the interval and all nonnegative numbers α, β such that $\alpha+\beta = 1$,

$$f(\alpha x + \beta y) \leq \alpha f(x) + \beta f(y). \tag{10.4}$$

A function is *strictly convex* if the above strict inequality holds whenever $x \neq y$, $\alpha, \beta \in (0,1)$, $\alpha + \beta = 1$. f is *concave* if $-f$ is convex.

A general and equivalent form of (10.4) is known as *Jensen's inequality*: let $f : \mathbb{I} \mapsto \mathbb{R}$ be a convex function on an interval $\mathbb{I} \subseteq \mathbb{R}$. Let $t_1, \ldots, t_n$ be nonnegative numbers such that $\sum_{i=1}^n t_i = 1$. Then

$$f\Big(\sum_{i=1}^n t_i x_i\Big) \leq \sum_{i=1}^n t_i f(x_i), \quad \text{whenever all } x_i \in \mathbb{I}.$$

If $f(t)$ is a differentiable function, from calculus we know that $f(t)$ is increasing if the first derivative is nonnegative, i.e., $f'(t) \geq 0$ on the interval, and $f(t)$ is convex if the second derivative is nonnegative, i.e., $f''(t) \geq 0$. Geometrically, the graph of a convex function is concave upwards. For example, $|t|$ is convex on $(-\infty, \infty)$ Also, if $f(x)$ is twice differentiable and $f''(x) > 0$ then $f(x)$ is strictly convex. For instance, t^2 and e^t are strictly convex on $(-\infty, \infty)$, whereas $\ln t$ is strictly concave on $(0, \infty)$.

Below is a reversal inequality of Jensen type.

Theorem 10.11 *Let $f : \mathbb{R}_+ \mapsto \mathbb{R}$ be a strictly convex function with $f(0) \leq 0$. If $x_1, \ldots, x_n$ are nonnegative numbers and at least two x_i are nonzero, then*

$$\sum_{i=1}^n f(x_i) < f\Big(\sum_{i=1}^n x_i\Big).$$

Proof. We prove the case $n = 2$. The general case is shown by induction. Let $x_1, x_2 > 0$. Write $x_1 = \frac{x_1}{x_1+x_2}(x_1 + x_2) + \frac{x_2}{x_1+x_2}\, 0$. Then

$$f(x_1) < \frac{x_1}{x_1 + x_2} f(x_1 + x_2) + \frac{x_2}{x_1 + x_2} f(0).$$

In a similar way, we have

$$f(x_2) < \frac{x_1}{x_1 + x_2} f(0) + \frac{x_2}{x_1 + x_2} f(x_1 + x_2).$$

Adding both sides of the inequalities, we arrive at

$$f(x_1) + f(x_2) < f(x_1 + x_2) + f(0).$$

The desired inequality follows at once as $f(0) \le 0$. ∎

We may generalize the definitions of convex and concave functions defined above to functions on $\mathbb{R}^n$ or on a convex subset of $\mathbb{R}^n$. For instance, f is convex on $\mathbb{R}^n$ if (10.4) holds for all $x, y \in \mathbb{R}^n$. A real-valued function ϕ defined on $\mathbb{R}^n$ ($\mathbb{R}^n_+$, in most cases, or even more generally, a convex set) is called *Schur-convex* if

$$x \prec y \quad \Rightarrow \quad \phi(x) \le \phi(y).$$

As an example, the function $\phi(x) = |x_1| + \cdots + |x_n|$ is Schur-convex on $\mathbb{R}^n$, where $x = (x_1, \ldots, x_n)$. Since if $x \prec y$, we can write $x = yA$, where $A = (a_{ij})$ is an $n \times n$ doubly stochastic matrix. Then

$$\begin{aligned}
\phi(x) &= \sum_{i=1}^n |x_i| = \sum_{i=1}^n \Big| \sum_{j=1}^n a_{ji} y_j \Big| \\
&\le \sum_{i=1}^n \sum_{j=1}^n a_{ji} |y_j| = \sum_{j=1}^n \Big(\sum_{i=1}^n a_{ji} \Big) |y_j| \\
&= \sum_{j=1}^n |y_j| = \phi(y).
\end{aligned}$$

One may prove that $\phi(x) = |x_1|^2 + \cdots + |x_n|^2$ is Schur-convex on $\mathbb{R}^n$ too. In fact, if $f(t)$ is convex on $\mathbb{R}$, then $\phi(x) = f(x_1) + \cdots + f(x_n)$ is Schur-convex on $\mathbb{R}^n$ (Problem 6).

The Schur-convex functions have been extensively studied. In this book we are more focused on the functions defined on $\mathbb{R}$ that preserve majorizations. When we write $f(x)$, where $x \in \mathbb{R}^n$, we mean, conventionally, that f is a function defined on an interval that contains all components of x, and that f is applied to every component of x; that is, if $x = (x_1, \ldots, x_n)$, then $f(x) = (f(x_1), \ldots, f(x_n))$.

In such a way, $x^2 = (x_1^2, \ldots, x_n^2)$ and $\ln x = (\ln x_1, \ldots, \ln x_n)$. By $x \in \mathbb{R}^n$, we automatically assume that the ith component of x is x_i.

The following theorem is useful in deriving inequalities.

Theorem 10.12 *Let x, $y \in \mathbb{R}^n$. If f is convex, then*

$$x \prec y \quad \Rightarrow \quad f(x) \prec_w f(y);$$

if f is increasing and convex, then

$$x \prec_w y \quad \Rightarrow \quad f(x) \prec_w f(y).$$

Proof. Since $x \prec y$, by Theorem 10.8, there exists a doubly stochastic matrix $A = (a_{ij})$ such that $x = yA$. This reveals

$$x_i = \sum_{j=1}^{n} a_{ji} y_j, \quad i = 1, 2, \ldots, n.$$

Applying f to both sides, and because f is convex, we have

$$f(x_i) \le \sum_{j=1}^{n} a_{ji} f(y_j), \quad i = 1, 2, \ldots, n.$$

Therefore,

$$(f(x_1), \ldots, f(x_n)) \le (f(y_1), \ldots, f(y_n))A.$$

It follows that $f(x) \le f(y)A$. By Theorem 10.8, the first part of the conclusion is immediate. For the second part, it suffices to note that $x \prec_w y$ ensures $x \le z \prec y$ for some $z \in \mathbb{R}^n$ (see Theorem 10.2). ∎

Corollary 10.1 *Let x, $y \in \mathbb{R}^n$. Then*

1. $x \prec y \;\Rightarrow\; |x| \prec_w |y|$, *i.e.,* $(|x_1|, \ldots, |x_n|) \prec_w (|y_1|, \ldots, |y_n|)$.
2. $x \prec y \;\Rightarrow\; x^2 \prec_w y^2$, *i.e.,* $(x_1^2, \ldots, x_n^2) \prec_w (y_1^2, \ldots, y_n^2)$.
3. $\ln x \prec_w \ln y \;\Rightarrow\; x \prec_w y$, *where all x_i and y_i are positive.*

Proof. For (1) and (2), it suffices to notice that $|t|$ and t^2 are convex, whereas (3) is due to the fact that e^t is increasing and convex. ∎

Majorization may be characterized in terms of convex functions.

Theorem 10.13 *Let $x, y \in \mathbb{R}^n$. Then*

1. $x \prec y \Leftrightarrow \sum_{i=1}^n f(x_i) \le \sum_{i=1}^n f(y_i)$ *for all convex functions f.*
2. $x \prec_w y \Leftrightarrow \sum_{i=1}^n f(x_i) \le \sum_{i=1}^n f(y_i)$ *for all increasing and convex functions f.*

Proof. Necessities are immediate from Theorem 10.12. For sufficiencies, we need to show that $\sum_{i=1}^k x_i^\downarrow \le \sum_{i=1}^k y_i^\downarrow$, $k = 1, 2, \ldots, n$. For any fixed k, take $f(r) = (r - y_k^\downarrow)^+$, where $t^+ = \max\{t, 0\}$; that is, $f(r) = r - y_k^\downarrow$ if $r \ge y_k^\downarrow$, 0 otherwise. Then $f(r)$ is a convex function in r and the condition $\sum_{i=1}^n f(y_i) \ge \sum_{i=1}^n f(x_i)$ reveals

$$\begin{aligned}\sum_{i=1}^k (y_i^\downarrow - y_k^\downarrow) &= \sum_{i=1}^k (y_i^\downarrow - y_k^\downarrow)^+ = \sum_{i=1}^n (y_i^\downarrow - y_k^\downarrow)^+ \\ &\ge \sum_{i=1}^n (x_i^\downarrow - y_k^\downarrow)^+ \ge \sum_{i=1}^k (x_i^\downarrow - y_k^\downarrow)^+ \\ &\ge \sum_{i=1}^k (x_i^\downarrow - y_k^\downarrow).\end{aligned}$$

This implies (2):

$$\sum_{i=1}^k x_i^\downarrow \le \sum_{i=1}^k y_i^\downarrow, \quad \text{i.e.,} \quad x \prec_w y.$$

For (1), take $g(x) = -x$. Then $g(x)$ is a convex function and this gives $\sum_{i=1}^n x_i \ge \sum_{i=1}^n y_i$. So equality has to hold and $x \prec y$. ∎

The next theorem is useful when an equality is in consideration.

Theorem 10.14 *Let $x, y \in \mathbb{R}^n$. If y is not a permutation of x, then for any strictly increasing and strictly convex function f that contains all the components of x and y,*

$$x \prec_w y \quad \Rightarrow \quad \sum_{i=1}^n f(x_i) < \sum_{i=1}^n f(y_i).$$

Proof. Since $x \prec_w y$, by Theorem 10.7, there exists a $z \in \mathbb{R}^n$ such that $x \prec z \le y$. Let $z = (z_1, \ldots, z_n)$. By Theorem 10.13, we have

$$\sum_{i=1}^{n} f(x_i) \le \sum_{i=1}^{n} f(z_i) \le \sum_{i=1}^{n} f(y_i).$$

We show that in these inequalities at least one strict inequality holds.

If z is a permutation of x, then z cannot be a permutation of y. Thus $z_k < y_k$ for some k. As f is strictly increasing, $f(z_k) < f(y_k)$. The second inequality is strict because $f(z_i) \le f(y_i)$ for all $i \neq k$.

If z is not a permutation of x, there exists a nonpermutation, doubly stochastic matrix $A = (a_{ij})$ such that $x = zA$. This reveals

$$x_i = \sum_{j=1}^{n} a_{ji} z_j, \quad i = 1, 2, \ldots, n.$$

Applying f to both sides, and since f is strictly convex, we have

$$f(x_i) \le \sum_{j=1}^{n} a_{ji} f(z_j), \quad i = 1, 2, \ldots, n,$$

and at least one strict inequality holds. It follows that

$$\sum_{i=1}^{n} f(x_i) < \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ji} f(z_j) \le \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ji} f(y_j) \le \sum_{j=1}^{n} f(y_j). \quad \blacksquare$$

Our next result is useful in deriving matrix inequalities and is used repeatedly in later sections. For this purpose, we introduce *log-majorization*. Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}_+^n$; that is, all x_i and y_i are nonnegative. We say that x is *weakly log-majorized* by y and write it as $x \prec_{\text{wlog}} y$ if

$$\prod_{i=1}^{k} x_i^{\downarrow} \le \prod_{i=1}^{k} y_i^{\downarrow}, \quad k = 1, \ldots, n.$$

If equality holds when $k = n$, we say that x is *log-majorized* by y and write it as $x \prec_{\log} y$. In the event that all components of x and y are positive, $x \prec_{\text{wlog}} y$ is the same as $\ln x \prec_w \ln y$, and $x \prec_{\log} y$ if and only if $\ln x \prec \ln y$, for $\ln t$ and e^t are strictly increasing functions.

Theorem 10.15 *Let $x,\ y \in \mathbb{R}^n_+$. Then*

$$x \prec_{\text{wlog}} y \quad \Rightarrow \quad x \prec_w y;$$

that is,

$$\prod_{i=1}^{k} x_i^{\downarrow} \le \prod_{i=1}^{k} y_i^{\downarrow}, k = 1, \ldots, n \quad \Rightarrow \quad \sum_{i=1}^{k} x_i^{\downarrow} \le \sum_{i=1}^{k} y_i^{\downarrow}, k = 1, \ldots, n.$$

Proof. If all components of x are positive, then all components of y are positive. In this case, $x \prec_{\text{wlog}} y$ yields $\ln x \prec_w \ln y$ which results in $x \prec_w y$ by Corollary 10.1(3).

If x contains zero components, say, $x_k^{\downarrow} = 0$ and $x_i^{\downarrow} > 0$ for all $i < k$, then, by the above argument, $(x_1^{\downarrow}, \ldots, x_{k-1}^{\downarrow}) \prec_w (y_1^{\downarrow}, \ldots, y_{k-1}^{\downarrow})$. So $x = (x_1^{\downarrow}, \ldots, x_{k-1}^{\downarrow}, 0, \ldots, 0) \prec_w (y_1^{\downarrow}, \ldots, y_{k-1}^{\downarrow}, y_k^{\downarrow} \ldots, y_n^{\downarrow}) = y$. ∎

Note that the converse of Theorem 10.15 is not true. For example, take $x = (3, 2, 1)$ and $y = (4, 1, 1)$. Then $x \prec y$, but $x \not\prec_{\text{wlog}} y$.

Theorem 10.16 *Let $x,\ y \in \mathbb{R}^n_+$. Then*

$$x^{\downarrow} \circ y^{\uparrow} \prec_w x \circ y \prec_w x^{\downarrow} \circ y^{\downarrow} \tag{10.5}$$

and

$$\prod_{i=1}^{n} (x_i^{\downarrow} + y_i^{\uparrow}) \ge \prod_{i=1}^{n} (x_i + y_i) \ge \prod_{i=1}^{n} (x_i^{\downarrow} + y_i^{\downarrow}). \tag{10.6}$$

Proof. We may only consider the case of positive x and y; otherwise we replace the zero components of x or y with arbitrarily small positive numbers and use a continuity argument. Note that $(\ln x)^{\downarrow} = \ln(x^{\downarrow})$. By Theorem 10.4(3), $\ln x^{\downarrow} + \ln y^{\uparrow} \prec \ln x + \ln y \prec \ln x^{\downarrow} + \ln y^{\downarrow}$; that is, $\ln(x^{\downarrow} \circ y^{\uparrow}) \prec \ln(x \circ y) \prec \ln(x^{\downarrow} \circ y^{\downarrow})$. Corollary 10.1(3) reveals (10.5). For (10.6), applying the convex function $f(t) = -\ln t$ to $x^{\downarrow} + y^{\uparrow} \prec x + y \prec x^{\downarrow} + y^{\downarrow}$ (Theorem 10.4(3)), Theorem 10.12 reveals

$$-\ln(x^{\downarrow} + y^{\uparrow}) \prec_w -\ln(x + y) \prec_w -\ln(x^{\downarrow} + y^{\downarrow}).$$

This yields

$$\begin{aligned} &\ln(x_1^{\downarrow} + y_1^{\uparrow}) + \cdots + \ln(x_n^{\downarrow} + y_n^{\uparrow}) \\ &\qquad \ge \ln(x_1 + y_1) + \cdots + \ln(x_n + y_n) \\ &\qquad\qquad \ge \ln(x_1^{\downarrow} + y_1^{\downarrow}) + \cdots + \ln(x_n^{\downarrow} + y_n^{\downarrow}). \end{aligned}$$

The desired inequalities in (10.6) then follow immediately. ■

We point out that if x or y contains negative components, then the conclusion does not necessarily follow. For instance, take $x = (-1, -2)$, $y = (1, 2)$. Then $x \circ y \prec_w x^{\downarrow} \circ y^{\downarrow}$ does not hold.

We end this section with the result (Problem 13) for $x, y \in \mathbb{R}_+^n$:

$$x \prec y \quad \Rightarrow \quad \prod_{i=1}^{n} x_i \geq \prod_{i=1}^{n} y_i.$$

Problems

1. Let $x, y \in \mathbb{R}^n$. If $x \leq y$ (componentwise), show that $x \prec_w y$.
2. Let $\alpha > 1$. Show that $f(t) = t^\alpha$ is strictly increasing and strictly convex on $\mathbb{R}_+$ and that $g(t) = |t|^\alpha$ is strictly convex on $\mathbb{R}$.
3. Show that $f(t) = e^{\alpha t}$, $\alpha > 0$, is strictly increasing and strictly convex on $\mathbb{R}$ and that $g(t) = \sqrt{t}$ is strictly concave and increasing on $\mathbb{R}_+$.
4. Show that $f(t) = \ln(\frac{1}{t} - 1)$ is convex on $(0, \frac{1}{2})$ but not on $(\frac{1}{2}, 1)$.
5. The following inequalities are of fundamental importance. They can be shown in various ways. One way is to use induction; another way is to use Jensen inequality with convex functions ($f(x) = -\ln x$, say).

   (a) Use Jensen inequality to show the general *arithmetic mean–geometric mean inequality*: if all $a_i \geq 0$, $p_i > 0$ and $\sum_{i=1}^n p_i = 1$,
   $$\prod_{i=1}^{n} a_i^{p_i} \leq \sum_{i=1}^{n} p_i a_i.$$

   (b) Use (a) to show the *Hölder inequality*: if $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$,
   $$\left| \sum_{i=1}^{n} a_i b_i \right| \leq \left(\sum_{i=1}^{n} |a_i|^p \right)^{1/p} \left(\sum_{i=1}^{n} |b_i|^q \right)^{1/q}$$
   for complex numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$.

   (c) Use (b) to show the *Minkowski inequality*: if $1 \leq p < \infty$,
   $$\left(\sum_{i=1}^{n} |a_i + b_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^{n} |a_i|^p \right)^{1/p} + \left(\sum_{i=1}^{n} |b_i|^p \right)^{1/p}$$
   for complex numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$.

6. Show that (i) if $f(t)$ is convex on $\mathbb{R}$, then $f_1(x) = f(x_1) + \cdots + f(x_n)$ and $f_2(x) = (f(x_1), \ldots, f(x_n))$ are Schur-convex on $\mathbb{R}^n$; and (ii) if g is convex on $\mathbb{R}^n$, then g is Schur-convex on $\mathbb{R}^n$.

7. Let $f(t)$ be a nonnegative continuous function defined on an interval $\mathbb{I} \subseteq \mathbb{R}$. If f is (strictly) convex on $\mathbb{I}$, show that $F(x) = \prod_{i=1}^n f(x_i)$ is (strictly) Schur-convex on $\mathbb{I}^n = \{(x_1, \ldots, x_n) : x_1, \ldots, x_n \in \mathbb{I}\} \subseteq \mathbb{R}^n$.

8. Give an example of convex, nonincreasing function $f(t)$ for which $(f(x_1), \ldots, f(x_n)) \prec_w (f(y_1), \ldots, f(y_n))$ is not true even if $x \prec_w y$.

9. Let $x, y \in \mathbb{R}^n$. Prove or disprove:

   (a) $x \prec_w y \Rightarrow |x| \prec_w |y|$, i.e., $(|x_1|, \ldots, |x_n|) \prec_w (|y_1|, \ldots, |y_n|)$.

   (b) $|x| \prec_w |y| \Rightarrow x^2 \prec_w y^2$, i.e., $(x_1^2, \ldots, x_n^2) \prec_w (y_1^2, \ldots, y_n^2)$.

   (c) $x \prec_w y \Rightarrow x^2 \prec_w y^2$, i.e., $(x_1^2, \ldots, x_n^2) \prec_w (y_1^2, \ldots, y_n^2)$.

   (d) $x \prec y \Rightarrow x^3 \prec_w y^3$, i.e., $(x_1^3, \ldots, x_n^3) \prec_w (y_1^3, \ldots, y_n^3)$.

   (e) $x \prec y \Rightarrow |x|^3 \prec_w |y|^3$, i.e., $(|x_1|^3, \ldots, |x_n|^3) \prec_w (|y_1|^3, \ldots, |y_n|^3)$.

   (f) $x \prec_w y \Rightarrow e^x \prec_w e^y$, i.e., $(e^{x_1}, \ldots, e^{x_n}) \prec_w (e^{y_1}, \ldots, e^{y_n})$.

10. Let $x, y \in \mathbb{R}^n$. Show that $|x^\downarrow - y^\downarrow| \prec_w |x - y|$.

11. Let $x, y \in \mathbb{R}_+^n$, $x \prec y$. Show that $\sum_{i=k}^n \sqrt{x_i^\downarrow} \geq \sum_{i=k}^n \sqrt{y_i^\downarrow}$ for each k.

12. Show that the following functions are Schur-convex on $\mathbb{R}^n$.

   (a) $f(x) = \max_i |x_i|$.

   (b) $g(x) = \sum_{i=1}^n |x_i|^p$, $p \geq 1$.

   (c) $h(x) = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$, $p \geq 1$.

   (d) $p(x) = \sum_{i=1}^n \frac{1}{x_i}$, where all $x_i > 0$.

13. Let $x, y \in \mathbb{R}_+^n$. If $x \prec y$, show that $\prod_{i=1}^n x_i \geq \prod_{i=1}^n y_i$ and that the strict inequality holds if y is not a permutation of x. Show by example that this is invalid if x or y contains nonnegative components.

14. Let $x, y \in \mathbb{R}_+^n$. Show that the sum inequalities $\sum_{i=1}^k x_i^\uparrow \leq \sum_{i=1}^k y_i^\uparrow$ $(k \leq n)$ imply the product inequalities $\prod_{i=1}^k x_i^\uparrow \leq \prod_{i=1}^k y_i^\uparrow$ $(k \leq n)$.

15. Let x, y, z be the three interior angles of any triangle. Show that

$$0 < \sin x + \sin y + \sin z \leq \frac{3}{2}\sqrt{3}.$$

16. Let $a_1, \dots, a_n$ and $b_1, \dots, b_n$ be positive numbers. Show that

$$\left(\frac{a_1^\downarrow}{b_1^\downarrow}, \dots, \frac{a_n^\downarrow}{b_n^\downarrow}\right) \prec_w \left(\frac{a_1}{b_1}, \dots, \frac{a_n}{b_n}\right) \prec_w \left(\frac{a_1^\downarrow}{b_n^\downarrow}, \dots, \frac{a_n^\downarrow}{b_1^\downarrow}\right).$$

17. Let $x_1, \dots, x_n$ be positive numbers such that $x_1 + \dots + x_n = 1$. Prove
    (a) $\sum_{i=1}^n \frac{1-x_i}{x_i} \geq n(n-1)$; $\sum_{i=1}^n \frac{x_i}{1-x_i} \geq \frac{n}{(n-1)}$.
    (b) $\sum_{i=1}^n \frac{1+x_i}{x_i} \geq n(n+1)$; $\frac{n}{(n+1)} \geq \sum_{i=1}^n \frac{x_i}{1+x_i} \geq \frac{1}{2}$.
    (c) $\sum_{i=1}^n \frac{1+x_i}{1-x_i} \geq \frac{n(n+1)}{n-1}$; $n-1 \geq \sum_{i=1}^n \frac{1-x_i}{1+x_i} \geq \frac{n(n-1)}{n+1}$.
    (d) $\sum_{i=1}^n x_i \ln \frac{1}{x_i} \leq \ln n$.

18. Let A be an $n \times n$ positive semidefinite matrix, $n > 1$. Show that $\operatorname{tr} e^A \leq e^{\operatorname{tr} A} + (n-1)$ with equality if and only if $\operatorname{rank}(A) \leq 1$.

19. Let $x, y \in \mathbb{R}_+^n$. If $x \prec_{\log} y$ and $x \prec y$, show that $x^\downarrow = y^\downarrow$.

20. For real number t, denote $t^+ = \max\{t, 0\}$. Let $x, y \in \mathbb{R}^n$. Show that
    (a) $x \prec y$ if and only if $\sum_{i=1}^n (x_i - t)^+ \leq \sum_{i=1}^n (y_i - t)^+$ for all $t \in \mathbb{R}$ and $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$.
    (b) $x \prec y$ if and only if $\sum_{i=1}^n |x_i - t| \leq \sum_{i=1}^n |y_i - t|$ for all $t \in \mathbb{R}$.

21. Show that both words "strictly" in Theorem 10.14 are necessary.

22. Let $x, y \in \mathbb{R}_+^n$. If $x \prec_w y$, show that $x^m \prec_w y^m$ for all integers $m \geq 1$, where $z^m = (z_1^m, \dots, z_n^m)$ for $z = (z_1, \dots, z_n) \in \mathbb{R}^n$. If $x^m \prec y^m$ for all integers $m \geq 1$, show that $x^\downarrow = y^\downarrow$.

23. Let g be a differentiable function on an interval $\mathbb{I} \subseteq \mathbb{R}$. Show that
    (a) g is convex if and only if $g(\frac{a+b}{2}) \leq \frac{1}{2}(g(a) + g(b))$ for all $a, b \in \mathbb{I}$.
    (b) g is linear if and only if $g(\frac{a+b}{2}) = \frac{1}{2}(g(a) + g(b))$ for all $a, b \in \mathbb{I}$.
    (c) $g(x) \prec g(y)$ whenever $x \prec y$, $x, y \in \mathbb{R}^n$, if and only if g is linear.

24. If $x, y, u, v \in \mathbb{R}_+^n$, show that $x \prec_{\text{wlog}} u, y \prec_{\text{wlog}} v \Rightarrow x \circ y \prec_{\text{wlog}} u^\downarrow \circ v^\downarrow$.

25. Let $a, b \in \mathbb{R}^n$ and all components of a and b be positive. Show that

$$\sum_{i=1}^n a_i^\downarrow b_i^\downarrow \leq \frac{r+s}{2\sqrt{rs}} \sum_{i=1}^n a_i b_i,$$

where r and s are the numbers such that $r \geq \frac{a_i}{b_i} \geq s > 0$, $i = 1, \dots, n$. [Hint: Use the Kantorovich inequality for $A = \operatorname{diag}(\frac{a_1}{b_1}, \dots, \frac{a_n}{b_n})$.]

. ———— ⊙ ———— .

10.4 Majorization of Diagonal Entries, Eigenvalues, and Singular Values

This section presents some elegant matrix inequalities involving diagonal entries, eigenvalues, and singular values in terms of *majorization*. Especially, we show that the relationship between the diagonal entries and eigenvalues of a Hermitian matrix is precisely characterized by majorization. It has been evident that majorization is a very useful tool in deriving matrix inequalities.

We denote the vectors of the diagonal entries, eigenvalues, and singular values of an n-square complex matrix A, respectively, by

$$\begin{aligned} d(A) &= (d_1(A), d_2(A), \ldots, d_n(A)), \\ \lambda(A) &= (\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A)), \\ \sigma(A) &= (\sigma_1(A), \sigma_2(A), \ldots, \sigma_n(A)). \end{aligned}$$

The singular values are always arranged in decreasing order. In the case where A is Hermitian, all $d_i(A)$ and $\lambda_i(A)$ are real; we assume that the components of $d(A)$ and $\lambda(A)$ are in decreasing order too.

Theorem 10.17 (Schur) *Let H be a Hermitian matrix. Then*

$$d(H) \prec \lambda(H).$$

Proof. Let H_k, $1 \leq k \leq n$, be the $k \times k$ principal submatrix of H with diagonal entries $d_1(H), d_2(H), \ldots, d_k(H)$. Theorem 8.10 yields

$$\sum_{i=1}^{k} d_i(H) = \operatorname{tr} H_k = \sum_{i=1}^{k} \lambda_i(H_k) \leq \sum_{i=1}^{k} \lambda_i(H).$$

Equality holds when $k = n$, for both sides are equal to $\operatorname{tr} H$. ∎

It is immediate that if H is a Hermitian matrix and U is any unitary matrix of the same size, then

$$d(U^*HU) \prec \lambda(H). \tag{10.7}$$

The next result shows the validity of the converse of Theorem 10.17. Thus the relationship between the diagonal entries and eigenvalues of a Hermitian matrix is precisely characterized by majorization.

Theorem 10.18 *Let* $d = (d_1, d_2, \dots, d_n)$, $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n) \in \mathbb{R}^n$. *If* $d \prec \lambda$, *then there exists an* $n \times n$ *real symmetric matrix* H *that has diagonal entries* $d_1, d_2, \dots, d_n$ *and eigenvalues* $\lambda_1, \lambda_2, \dots, \lambda_n$.

Proof. The proof goes as follows. We first deal with the case of $n = 2$, then proceed inductively. For $n > 2$, we construct a new vector (sequence) λ' that differs from λ by at most two components one of which is some d_i, i.e., λ' is a step "closer" to d. An application of induction results in the desired conclusion. We may assume that $d_1, d_2, \dots, d_n$ and $\lambda_1, \lambda_2, \dots, \lambda_n$ are arranged in decreasing order. Denote $\operatorname{diag} d = \operatorname{diag}(d_1, d_2, \dots, d_n)$ and $\operatorname{diag} \lambda = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$.

If $n = 2$ and $\lambda_1 = \lambda_2$, then $d_1 = d_2$. The statement is obviously true. If $n = 2$ and $\lambda_1 > \lambda_2$, because $(d_1, d_2) \prec (\lambda_1, \lambda_2)$, we have $\lambda_2 \le d_2 \le d_1 \le \lambda_1$ and $d_1 + d_2 = \lambda_1 + \lambda_2$. Take

$$U = (\lambda_1 - \lambda_2)^{-1/2} \begin{pmatrix} (d_1 - \lambda_2)^{1/2} & -(\lambda_1 - d_1)^{1/2} \\ (\lambda_1 - d_1)^{1/2} & (d_1 - \lambda_2)^{1/2} \end{pmatrix}.$$

Then it is routine to check that U is real orthogonal and that $H = U^T (\operatorname{diag} \lambda)\, U = \left(\begin{smallmatrix} d_1 & * \\ * & d_2 \end{smallmatrix}\right)$ is real symmetric with diagonal entries d_1, d_2.

Now suppose $n > 2$. If $d_1 = \lambda_1$, then $(d_2, \dots, d_n) \prec (\lambda_2, \dots, \lambda_n)$. By induction hypothesis on $n-1$, there exists an $(n-1)$-square real symmetric matrix G having diagonal entries $d_2, \dots, d_n$ and eigenvalues $\lambda_2, \dots, \lambda_n$. Then take $H = (\lambda_1) \oplus G$, as desired.

If $d_1 < \lambda_1$, since the sum of all d_is equals that of all λ_is, it is impossible that $d_i < \lambda_i$ for all i. Let $d_1 < \lambda_1, \dots, d_k < \lambda_k$, and $d_{k+1} \ge \lambda_{k+1}$ for some $k \ge 1$. Then $\lambda_{k+1} \le d_{k+1} \le d_k < \lambda_k$. Put $\lambda'_{k+1} = \lambda_{k+1} + \lambda_k - d_k$. It follows that $(d_k, \lambda'_{k+1}) \prec (\lambda_k, \lambda_{k+1})$. By the above argument for $n = 2$, there is a 2×2 real orthogonal matrix U_2 such that the diagonal entries of $U_2^T \operatorname{diag}(\lambda_k, \lambda_{k+1}) U_2$ are d_k, λ'_{k+1}.

Now replace the 2×2 principal submatrix in rows k and $k+1$ of the identity matrix I_n by U_2 to get an $n \times n$ matrix V. Then V is real orthogonal. Set $F = V^T (\operatorname{diag} \lambda)\, V$. Then F is real symmetric, having eigenvalues $\lambda_1, \dots, \lambda_k, \lambda_{k+1}, \dots, \lambda_n$ and diagonal entries $\lambda_1, \dots d_k, \lambda'_{k+1}, \dots, \lambda_n$. Let $\lambda' = (\lambda_1, \dots d_k, \lambda'_{k+1}, \dots, \lambda_n)$. Then $d \prec \lambda' \prec \lambda$; λ' and λ differ by only two elements, and d and λ' both contain d_k. Let $\tilde{d}$ be the vector by deleting d_k from d and $\tilde{\lambda}$ by deleting d_k from λ'. Then $\tilde{d} \prec \tilde{\lambda}$. By induction, there exists an $(n-1)$-square real orthogonal matrix U_{n-1} such that $U_{n-1}^T(\operatorname{diag}\tilde{\lambda})\,U_{n-1}$ has diagonal entries $\tilde{d}$. Now construct an n-square matrix $W = (w_{ij})$ by inserting a row and column in U_{n-1} so that $w_{kk} = 1$ and all other ws are 0 in the row and column. Then W is real orthogonal and $H = W^T F W$ has diagonal entries $d_1, d_2, \ldots, d_n$, as desired. ∎

Theorem 10.19 *Let A be an n-square complex matrix. Then*

$$|d(A)| \prec_w \sigma(A) \tag{10.8}$$

and

$$|\lambda(A)| \prec_w \sigma(A). \tag{10.9}$$

Proof. We may assume that the absolute values of the diagonal entries of A are in decreasing order. For each a_{ii}, let t_i be such that

$$t_i a_{ii} = |a_{ii}|, \quad |t_i| = 1, \quad i = 1, \ldots, n.$$

Let $B = A\operatorname{diag}(t_1, \ldots, t_n)$ and let C be the leading $k \times k$ principal submatrix of B. Then B has the same singular values as A, and

$$d(C) = (|a_{11}|, \ldots, |a_{kk}|).$$

By Theorem 8.14, we have $\sigma_i(C) \le \sigma_i(B),\ i = 1, \ldots, k$.

Applying Theorem 9.6 reveals (10.8):

$$\begin{aligned} |\operatorname{tr} C| &= |a_{11}| + \cdots + |a_{kk}| \\ &\le \sigma_1(C) + \cdots + \sigma_k(C) \\ &\le \sigma_1(B) + \cdots + \sigma_k(B) \\ &= \sigma_1(A) + \cdots + \sigma_k(A). \end{aligned}$$

For (10.9), let $A = U^*TU$ be a Schur decomposition of A, where U is unitary and T is upper-triangular. Then

$$|\lambda(A)| = |d(T)| \quad \text{and} \quad \sigma(A) = \sigma(T).$$

An application of (10.8) gives (10.9). ∎

Theorem 10.20 *Let A, B, C be $n \times n$ complex matrices such that*

$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \geq 0.$$

Then

$$\sigma(B) \prec_{\text{wlog}} \lambda^{1/2}(A) \circ \lambda^{1/2}(C) \tag{10.10}$$

and

$$|\lambda(B)| \prec_{\text{wlog}} \lambda^{1/2}(A) \circ \lambda^{1/2}(C). \tag{10.11}$$

Proof. For any $n \times p$ matrix U and any $n \times q$ matrix V, we have

$$\begin{pmatrix} U^* & 0 \\ 0 & V^* \end{pmatrix} \begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} U^*AU & U^*BV \\ V^*B^*U & V^*CV \end{pmatrix} \geq 0.$$

If $B = 0$, there is nothing to prove. Let $\operatorname{rank}(B) = r > 0$ and choose U and V so that $B = UDV^*$ is a singular value decomposition of B, where D is the $r \times r$ diagonal matrix $\operatorname{diag}(\sigma_1(B), \ldots, \sigma_r(B))$, and U and V are $n \times r$ partial unitary matrices; that is, $U^*U = V^*V = I_r$. Then $U^*BV = D$.

Denote by $[X]_k$ the $k \times k$ leading principal submatrix of matrix X. Extracting such submatrix from each block for every $k \leq n$:

$$\begin{pmatrix} [U^*AU]_k & [D]_k \\ [D]_k & [V^*CV]_k \end{pmatrix} \geq 0.$$

Taking the determinant for each block and then for the 2×2 matrix,

$$\det[D]_k^2 \leq \det([U^*AU]_k) \det([V^*CV]_k).$$

Or equivalently, for each $1 \leq k \leq r$,

$$\prod_{i=1}^{k} \sigma_i^2(B) \leq \prod_{i=1}^{k} \lambda_i([U^*AU]_k) \lambda_i([V^*CV]_k).$$

By the eigenvalue interlacing theorem (Section 8.3), we arrive at

$$\prod_{i=1}^{k} \sigma_i^2(B) \leq \prod_{i=1}^{k} \lambda_i(A) \lambda_i(C).$$

(10.10) follows by taking the square roots of both sides. (10.11) is similarly obtained by letting $B = WTW^*$, where W is unitary and T is an upper-triangular matrix with eigenvalues $\lambda_1(B), \lambda_2(B), \ldots, \lambda_n(B)$ on the main diagonal (Schur decomposition). ∎

Corollary 10.2 (Weyl) *Let A be any $n \times n$ complex matrix. Then*

$$|\lambda(A)| \prec_{\log} \sigma(A). \tag{10.12}$$

Proof 1. Note that $|\lambda_1(A) \cdots \lambda_n(A)| = \sigma_1(A) \cdots \sigma_n(A) = |\det A|$. To show the log-majorization, let $A = UDV$ be a singular value decomposition of A, where U, V are unitary, and D is diagonal. Then

$$\begin{pmatrix} |A^*| & A \\ A^* & |A| \end{pmatrix} = \begin{pmatrix} U & 0 \\ 0 & V^* \end{pmatrix} \begin{pmatrix} D & D \\ D & D \end{pmatrix} \begin{pmatrix} U^* & 0 \\ 0 & V \end{pmatrix} \geq 0.$$

Applying (10.11) to the block matrix on the left gives the inequality.

Proof 2. Let $|\lambda_{\max}(X)|$ represent the largest modulus of the eigenvalues of X. Then $|\lambda_{\max}(X)| \leq \sigma_1(X)$. For any given positive integer $k \leq n$, consider the compound matrix $A^{(k)}$. For any $j_1 < \cdots < j_k$,

$$\prod_{i=1}^{k} |\lambda_{j_i}(A)| \leq |\lambda_{\max}(A^{(k)})| \leq \sigma_1(A^{(k)}) = \prod_{i=1}^{k} \sigma_i(A). \quad ∎$$

As log-majorization implies weak majorization, we have

$$|\lambda(A)| \prec_w \sigma(A), \quad \text{i.e.,} \quad |\lambda(A)| \prec_w \lambda(|A|),$$

which is (10.9). Note that this is weaker than the log-majorization.

Corollary 10.3 (Horn) *Let $A, B \in \mathbb{M}_n$. Then*

$$\sigma(AB) \prec_{\log} \sigma(A) \circ \sigma(B).$$

Proof. This is immediate from (10.10) by observing that

$$(A^*, B)^*(A^*, B) = \begin{pmatrix} AA^* & AB \\ B^*A^* & B^*B \end{pmatrix} \geq 0. \tag{10.13}$$

Note that for any n-square matrix X, $\prod_{i=1}^n \sigma_i(X) = |\det X|$. ∎

Corollary 10.4 (Bhatia–Kittaneh) *Let $A, B \in \mathbb{M}_n$. Then*

$$2\sigma_i(AB) \le \sigma_i(A^*A + BB^*), \quad i = 1, 2, \dots, n.$$

Proof. Use the block matrix in (10.13). By Problem 11, we have

$$\begin{aligned} 2\sigma_i(AB) &\le \lambda_i\Big((A^*, B)^*(A^*, B)\Big) \\ &= \lambda_i\Big((A^*, B)(A^*, B)^*\Big) \\ &= \lambda_i(A^*A + BB^*) = \sigma_i(A^*A + BB^*). \quad \blacksquare \end{aligned}$$

Problems

1. Let $A \in \mathbb{M}_n$. Prove or disprove each of the following identities.
   (a) $\sigma(A) = \sigma(A^*)$.
   (b) $\sigma(|A|) = \sigma(|A^*|)$.
   (c) $\lambda(A) = \lambda(A^*)$.
   (d) $\lambda(|A|) = \lambda(|A^*|)$.
   (e) $d(A) = d(A^*)$.
   (f) $d(|A|) = d(|A^*|)$.
   (g) $\sigma(A) = \lambda(|A|)$.
   (h) $\sigma(A) = \sigma(|A|)$.

2. Let A be a Hermitian matrix partitioned as $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$. Show that $A_{11} \oplus A_{22} = \frac{1}{2}(A + UAU^*)$, where $U = I \oplus (-I)$, and that
$$\lambda(A_{11} \oplus A_{22}) = (\lambda(A_{11}), \lambda(A_{22})) \prec \lambda(A).$$

3. For any square complex matrix $A = (a_{ij})$, show that
$$\max_i |a_{ii}| \le |\lambda_{\max}(A)| \le \sigma_{\max}(A).$$

4. Show by example that $|d(A)| \prec_{\text{wlog}} \sigma(A)$ is not true in general.

5. Show that $|d(A)| \prec_w |\lambda(A)|$ for all normal matrices A and that it is not true for nonnormal matrix $B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$.

6. Let $A = (a_{ij}) \in \mathbb{M}_n$ and p be a permutation on $\{1, 2, \dots, n\}$. Denote $d_p(A) = (a_{1p(1)}, a_{2p(2)}, \dots, a_{np(n)})$. Show that $|d_p(A)| \prec_w \sigma(A)$.

7. Verify that $\left(\begin{smallmatrix} 1 & 0 \\ c & 1 \end{smallmatrix}\right)^{-1} \left(\begin{smallmatrix} \lambda_1 & 1 \\ 0 & \lambda_2 \end{smallmatrix}\right) \left(\begin{smallmatrix} 1 & 0 \\ c & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} d_1 & * \\ * & d_2 \end{smallmatrix}\right)$, where $c = d_1 - \lambda_1$.

8. Let A be a square complex matrix. Majorization (10.8) ensures that $|d(A)| \prec_w \sigma(A)$. Show that this $\prec_w$ becomes $\prec$ if and only if $A = PU$ for a positive semidefinite matrix P and a diagonal unitary matrix U.

9. Let A be a square complex matrix. Majorization (10.9) ensures that $|\lambda(A)| \prec_w \sigma(A)$. Show that this $\prec_w$ becomes $\prec$ if and only if A is normal; that is, $|\lambda(A)| \prec \lambda(|A|)$ if and only if A is normal.

10. Let A be an $n \times n$ positive definite matrix having diagonal entries $d_1 \geq d_2 \geq \cdots \geq d_n$ and eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Show that
$$\prod_{i=k}^{n} d_i \geq \prod_{i=k}^{n} \lambda_i, \quad \sum_{i=k}^{n} \frac{1}{d_i} \leq \sum_{i=k}^{n} \frac{1}{\lambda_i}, \qquad k = 1, 2, \dots, n.$$

11. Let A, B, C be $n \times n$ complex matrices such that $M = \left(\begin{smallmatrix} A & B \\ B^* & C \end{smallmatrix}\right) \geq 0$. Show that $M - 2N \geq 0$, where $N = \left(\begin{smallmatrix} 0 & B \\ B^* & 0 \end{smallmatrix}\right) \geq 0$, and that
$$2\sigma_i(B) \leq \lambda_i(M), \quad i = 1, 2, \dots, n.$$

12. Referring to Corollary 10.4, show by example that A^*A cannot be replaced with AA^*; that is, it is not true in general that $2\sigma_i(AB) \leq \sigma_i(AA^* + BB^*)$, even though $2\sigma_i(AB) \leq \sigma_i(A^*A + BB^*)$ for all i.

13. Show that Corollary 10.3 implies Theorem 10.20 via Theorem 6.8.

14. Let $A \in \mathbb{M}_n$. Show that

    (a) $|\lambda(A)|^2 \prec_w \sigma^2(A)$.

    (b) $|d(A)|^2 \prec_{\text{wlog}} d(|A|) \circ \lambda(|A^*|)$.

    (c) $d(|A|) \circ d(|A^*|) \prec_w \lambda(A^*A)$.

    (d) $|\lambda(A + A^*)| \prec_w \lambda(|A| + |A^*|)$ and $|\lambda(A \circ A^*)| \prec_w \lambda(|A| \circ |A^*|)$.

15. Let A, B, $C \in \mathbb{M}_n$ be such that $\left(\begin{smallmatrix} A & B \\ B^* & C \end{smallmatrix}\right) \geq 0$. Show that
$$\begin{pmatrix} \|A\|_{\text{op}} & \|B\|_{\text{op}} \\ \|B^*\|_{\text{op}} & \|C\|_{\text{op}} \end{pmatrix} \geq 0 \text{ and } \begin{pmatrix} \rho(A) & \rho(B) \\ \rho(B^*) & \rho(C) \end{pmatrix} \geq 0,$$
where $\|X\|_{\text{op}}$ and $\rho(X)$ are the spectral norm and spectral radius of square matrix X, respectively. Do these hold for 3×3 block matrices?

10.5 Majorization for Matrix Sum

Now we turn our attention to the eigenvalue and singular value inequalities in majorization for the sum of Hermitian matrices.

Theorem 10.21 (Fan) *Let A, $B \in \mathbb{M}_n$ be Hermitian. Then*

$$\lambda(A+B) \prec \lambda(A) + \lambda(B).$$

Proof. By Theorem 8.17, with S_k denoting a set of any k orthonormal vectors $x_1, x_2, \ldots, x_k \in \mathbb{C}^n$, we see that the weak majorization $\lambda(A+B) \prec_w \lambda(A) + \lambda(B)$ is equivalent to, for each $k \leq n$,

$$\max_{S_k} \sum_{i=1}^{k} x_i^*(A+B)x_i \leq \max_{S_k} \sum_{i=1}^{k} x_i^* A x_i + \max_{S_k} \sum_{i=1}^{k} x_i^* B x_i.$$

Since $\operatorname{tr}(A+B) = \operatorname{tr} A + \operatorname{tr} B$, the desired majorization follows. ∎

Theorem 10.22 *Let A, $B \in \mathbb{M}_n$ be Hermitian. Then*

$$\lambda(A) - \lambda(B) \prec \lambda(A-B).$$

Proof. Write $A = B + (A - B)$. By Theorem 8.18,

$$\sum_{t=1}^{k} \lambda_{i_t}(A) \leq \sum_{t=1}^{k} \lambda_{i_t}(B) + \sum_{j=1}^{k} \lambda_j(A-B),$$

which yields, for $k = 1, 2, \ldots, n$,

$$\max_{1 \leq i_1 < \cdots < i_k \leq n} \sum_{t=1}^{k} \Big(\lambda_{i_t}(A) - \lambda_{i_t}(B)\Big) \leq \sum_{j=1}^{k} \lambda_j(A-B);$$

that is,

$$\lambda(A) - \lambda(B) \prec_w \lambda(A-B).$$

As equality holds when $k = n$, the desired majorization follows. ∎

Note that $\lambda^{\uparrow}(H) = (\lambda_n(H), \ldots, \lambda_1(H)) = -\lambda(-H)$ for any $n \times n$ Hermitian matrix H. We may rewrite the above two theorems as

$$\lambda(A) + \lambda^{\uparrow}(B) \prec \lambda(A+B) \prec \lambda(A) + \lambda(B); \tag{10.14}$$

$$\lambda(A) - \lambda(B) \prec \lambda(A-B) \prec \lambda(A) - \lambda^{\uparrow}(B). \tag{10.15}$$

For singular value majorizations, if A is an $m \times n$ complex matrix, as usual, we denote by $\sigma(A)$ the singular value vector of A; the singular values of A are the eigenvalues of $|A| = (A^*A)^{1/2}$. As we know, for any matrix X, matrix $\hat{X} = \left(\begin{smallmatrix} 0 & X \\ X^* & 0 \end{smallmatrix}\right)$ is Hermitian and has eigenvalues $\sigma_1(X), \ldots, \sigma_r(X), 0, \ldots, 0, -\sigma_r(X), \ldots, -\sigma_1(X)$, where r is the rank of X. Applying Theorem 10.21 to $\hat{A} = \left(\begin{smallmatrix} 0 & A \\ A^* & 0 \end{smallmatrix}\right)$ and $\hat{B} = \left(\begin{smallmatrix} 0 & B \\ B^* & 0 \end{smallmatrix}\right)$ gives the analogous majorization for singular values.

Theorem 10.23 *Let A and B be $m \times n$ complex matrices. Then*

$$\sigma(A+B) \prec_w \sigma(A) + \sigma(B). \tag{10.16}$$

Applying Theorem 10.22 to $\hat{A}$ and $\hat{B}$ reveals the following majorization on the difference (in absolute value) of singular values and the singular values of the difference of matrices.

Theorem 10.24 *Let A and B be $m \times n$ complex matrices. Then*

$$|\sigma(A) - \sigma(B)| \prec_w \sigma(A-B). \tag{10.17}$$

Proof. By Theorem 10.22, $\lambda(\hat{A}) - \lambda(\hat{B}) \prec \lambda(\hat{A} - \hat{B})$; that is,

$$(\sigma_1(A) - \sigma_1(B), \ldots, 0, \ldots, 0, \ldots, \sigma_1(B) - \sigma_1(A)) \prec \lambda(\hat{A} - \hat{B}).$$

It follows that, for $k = 1, 2, \ldots, n$,

$$\max_{1 \le i_1 < \cdots < i_k \le n} \sum_{t=1}^{k} \left| \sigma_{i_t}(A) - \sigma_{i_t}(B) \right| \le \sum_{j=1}^{k} \sigma_j(A-B). \quad \blacksquare$$

Theorem 10.25 *Let A and B be n-square positive semidefinite matrices and let z be any complex number. Then*

$$\sigma(A - |z|B) \prec_{\text{wlog}} \sigma(A + zB) \prec_{\text{wlog}} \sigma(A + |z|B). \tag{10.18}$$

Proof. For the second part, by (10.10), it is sufficient to notice that

$$\begin{pmatrix} A+|z|B & A+zB \\ A+z^*B & A+|z|B \end{pmatrix} \geq 0.$$

Now we show the first majorization in (10.18). If A and B are nonnegative diagonal matrices, invoking the elementary inequality $|a-|z|b| \leq |a+zb|$, where $z \in \mathbb{C}$, $a, b \geq 0$, we arrive at

$$|\det(A-|z|B)| \leq |\det(A+zB)|. \tag{10.19}$$

Inequality (10.19) actually holds for all positive semidefinite A and B due to the fact that there exists an invertible matrix P such that P^*AP and P^*BP are both nonnegative diagonal (Theorem 7.6).

Note that $A-|z|B$ is Hermitian. By (10.19) and (10.12), we have

$$\prod_{i=1}^n \sigma_i(A-|z|B) \leq \prod_{i=1}^n |\lambda_i(A+zB)| \leq \prod_{i=1}^n \sigma_i(A+zB). \tag{10.20}$$

We claim that the upper limit n for the products in (10.20) can be replaced by any positive integer $k \leq n$. To show this, let U be an n-square unitary matrix such that $U^*(A-|z|B)U = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, where $|\lambda_i| = \sigma_i(A-|z|B)$. Write $U = (U_1, U_2)$, where U_1 consists of the first k columns of U. Then $U_1^*AU_1$ and $U_1^*BU_1$ are $k \times k$. Thus,

$$\begin{aligned}
\prod_{i=1}^k \sigma_i(A-|z|B) &= \prod_{i=1}^k \sigma_i(U_1^*(A-|z|B)U_1) \\
&= \prod_{i=1}^k \sigma_i(U_1^*AU_1 - |z|U_1^*BU_1) \\
&\leq \prod_{i=1}^k \sigma_i(U_1^*AU_1 + zU_1^*BU_1) \quad \text{(by (10.20))} \\
&= \prod_{i=1}^k \sigma_i(U_1^*(A+zB)U_1) \\
&\leq \prod_{i=1}^k \sigma_i(A+zB).
\end{aligned}$$

The last inequality is by Problem 11 of Section 8.3. ∎

The following result is immediate since "$\prec_{\text{wlog}}$" implies "$\prec_w$".

Corollary 10.5 *Let A and B be n-square positive semidefinite matrices and let z be any complex number with $|z| = 1$. Then*

$$\sigma(A - B) \prec_w \sigma(A + zB) \prec_w \sigma(A + B).$$

Theorem 10.26 *Let A be an $n \times n$ complex matrix and let $\mathcal{H}(A)$ be the Hermitian part of A; that is, $\mathcal{H}(A) = (A + A^*)/2$. Then for any $n \times n$ Hermitian matrix G,*

$$\sigma(A - \mathcal{H}(A)) \prec_w \sigma(A - G). \tag{10.21}$$

Proof. By Theorem 10.23,

$$\begin{aligned} \sigma(A - \mathcal{H}(A)) &= \sigma\Big((A - G)/2 - (A - G)^*/2\Big) \\ &\prec_w \sigma((A - G)/2) + \sigma((A - G)^*/2) \\ &= \sigma(A - G). \quad \blacksquare \end{aligned}$$

Theorem 10.27 *Let A be an $n \times n$ positive semidefinite matrix. Then for any $n \times n$ unitary matrix U,*

$$\sigma(A - I_n) \prec_w \sigma(A - U) \prec_w \sigma(A + I_n). \tag{10.22}$$

Proof. Because (10.22) holds if and only if it holds when A is replaced with W^*AW, where W is unitary, we may assume that A is a diagonal matrix with nonnegative diagonal entries. An application of Theorem 10.24 results in

$$\begin{aligned} \sigma(A - I_n) &= (|\sigma(A) - \sigma(U)|)^{\downarrow} \prec_w \sigma(A - U) \\ &\prec_w \sigma(A) + \sigma(U) = \sigma(A + I). \quad \blacksquare \end{aligned}$$

Corollary 10.6 *Let A be an $n \times n$ complex matrix and $A = UP$ be a polar decomposition of A, where U is unitary and P is positive semidefinite. Then for any $n \times n$ unitary matrix V*

$$\sigma(A - U) \prec_w \sigma(A - V) \prec_w \sigma(A + U). \tag{10.23}$$

Proof. Notice that

$$\sigma(A-U)=\sigma(UP-U)=\sigma(P-I).$$

Invoking Theorem 10.27 reveals the desired majorizations. ∎

We end this section by presenting another result of Fan.

Theorem 10.28 (Fan) *Let A be an $n\times n$ matrix with eigenvalues $\lambda_1(A),\dots,\lambda_n(A)$. Let $\operatorname{Re}\lambda(A)=(\operatorname{Re}\lambda_1(A),\dots,\operatorname{Re}\lambda_n(A))$. Then*

$$\operatorname{Re}\lambda(A)\prec\lambda(\mathcal{H}(A)),$$

where $\mathcal{H}(A)=(A+A^*)/2$ is the Hermitian part of the matrix A.

Proof. By the Schur triangularization theorem, we can write $A=V^*TV$, where V is unitary and T is upper-triangular. Moreover, we may assume that $\operatorname{Re}\lambda_1(A)\geq\cdots\geq\operatorname{Re}\lambda_n(A)$. Note that

$$\sum_{i=1}^n\lambda_i(\mathcal{H}(A))=\operatorname{tr}\mathcal{H}(A)=\operatorname{Re}\operatorname{tr}A=\operatorname{Re}\sum_{i=1}^n\lambda_i(A)=\sum_{i=1}^n\operatorname{Re}\lambda_i(A).$$

Now for partial sum, by min-max representation, we have

$$\begin{aligned}\sum_{i=1}^k\lambda_i(\mathcal{H}(A)) &= \max_{UU^*=I_k}\operatorname{tr}U(\mathcal{H}(A))U^*\\ &= \max_{UU^*=I_k}\operatorname{tr}UV^*(\mathcal{H}(T))VU^*\\ &= \max_{WW^*=I_k}\operatorname{tr}W(\mathcal{H}(T))W^*.\end{aligned}$$

Take $W=(I_k,0)$, $k\times n$. Then

$$\sum_{i=1}^k\lambda_i(\mathcal{H}(A))\geq\sum_{i=1}^k\Big(\frac{\lambda_i(A)+\overline{\lambda}_i(A)}{2}\Big)=\sum_{i=1}^k\operatorname{Re}\lambda_i(A).$$ ∎

Problems

1. Let z be a complex number. Show that $|1-|z||\leq|1-z|\leq1+|z|$.

2. Show that the following matrix inequalities do not hold in general

$$|A - |z|B| \leq |A - zB| \leq A + |z|B$$

by taking $z = i$, $A = \left(\begin{smallmatrix} 4 & 2 \\ 2 & 1 \end{smallmatrix}\right)$, and $B = \left(\begin{smallmatrix} 1 & 2 \\ 2 & 4 \end{smallmatrix}\right)$.

3. Let $A \in \mathbb{M}_n$ and x be any n-row complex vector. Show that

$$\lambda(A + x^*x) \prec \big(\lambda_1(A) + xx^*,\ \lambda_2(A), \ldots,\ \lambda_n(A)\big).$$

4. Show by example that neither of the following holds in general.

$$\sigma_i(A + B) \geq \sigma_i(A) + \sigma_n(B), \quad i = 1, 2, \ldots, n;$$

$$\sum_{i=1}^{k} \sigma_i(A + B) \geq \sum_{i=1}^{k} \sigma_i(A) + \sum_{i=1}^{k} \sigma_{n-i+1}(B), \quad k = 1, 2, \ldots, n,$$

where A, $B \in \mathbb{M}_n$. However, the second inequality is true if the plus sign "+" on the right-hand side is replace by the minus sign "−".

5. Let A be an $n \times n$ complex matrix. Denote $\mathcal{S}(A) = \frac{A - A^*}{2}$. Show that for any $n \times n$ skew-Hermitian matrix G

$$\sigma(A - \mathcal{S}(A)) \prec_w \sigma(A - G).$$

6. Let $A \in \mathbb{M}_n$ have a singular value decomposition $A = UDV$, where U and V are unitary. Show that for any $n \times n$ unitary matrix W,

$$\sigma(A - UV) \prec_w \sigma(A - W) \prec_w \sigma(A + UV).$$

7. If $A = (a_{ij}) \in \mathbb{M}_n$ is normal and $d(A) = (a_{11}, \ldots, a_{nn})$, show that

$$\operatorname{Re} d(A) \prec \operatorname{Re} \lambda(A).$$

8. **(Fan–Hoffman)** Let A be a square complex matrix. Show that

$$\lambda\left(\frac{A^* + A}{2}\right) \prec_w \left|\lambda\left(\frac{A^* + A}{2}\right)\right| \prec_w \sigma(A).$$

9. Let $A \in \mathbb{M}_n$. Denote the (necessarily real) eigenvalues of the Hermitian matrix $\mathcal{H}(A) = \frac{A + A^*}{2}$ by $h_1 \geq h_2 \geq \cdots \geq h_n$. Show that

$$h_i \leq \sigma_i(A), \quad i = 1, 2, \ldots, n.$$

However, if we rearrange the absolute values of these eigenvalues of $\mathcal{H}(A)$ in the decreasing order $|h_{i_1}| \geq |h_{i_2}| \geq \cdots \geq |h_{i_n}|$, show by example that $|h_{i_t}| \leq \sigma_t(A)$ does not hold in general.

10. Let A and B be $n \times n$ complex matrices. Show that
$$\sigma(A+B) \prec_w \sigma(A \oplus B) \prec_w \sigma(|A|+|B|).$$

11. Let $A_1, \ldots, A_m$ be n-square positive semidefinite matrices and let $\lambda_1, \ldots, \lambda_m$ be complex numbers. Show that
$$|\det(\lambda_1 A_1 + \cdots + \lambda_m A_m)| \le \det(|\lambda_1| A_1 + \cdots + |\lambda_m| A_m).$$

12. Let A and B be $n \times n$ positive semidefinite matrices. Show that
$$|\lambda(A-B)| \prec_{\text{wlog}} \lambda(A+B).$$

13. Let A and B be $n \times n$ positive semidefinite matrices. Show that
$$\sigma_i(A-B) \le \sigma_i(A \oplus B), \quad i = 1, 2, \ldots, n.$$

14. Let A and B be $n \times n$ positive semidefinite matrices. Show that
$$\big(\lambda(A+B), 0\big) \prec \big(\lambda(A), \lambda(B)\big).$$

15. Let A and B be $n \times n$ positive definite matrices. Show that
$$\sum_{i=1}^{n} \frac{1}{\lambda_i(A) + \lambda_{n-i+1}(B)} \le \operatorname{tr}\big((A+B)^{-1}\big) \le \sum_{i=1}^{n} \frac{1}{\lambda_i(A) + \lambda_i(B)}$$
and
$$\prod_{i=1}^{n} \big(\lambda_i(A) + \lambda_{n-i+1}(B)\big) \le \det(A+B) \le \prod_{i=1}^{n} \big(\lambda_i(A) + \lambda_i(B)\big).$$

16. Let A, B be $n \times n$ strict contractions, i.e., $\sigma_1(A) < 1, \sigma_1(B) < 1$. Let
$$H = \begin{pmatrix} (I - A^*A)^{-1} & (I - A^*B)^{-1} \\ (I - B^*A)^{-1} & (I - B^*B)^{-1} \end{pmatrix}.$$
Show that
$$(1, 1, \ldots, 1) \le d(H) \prec \big(\lambda((I - A^*A)^{-1}), \lambda((I - B^*B)^{-1})\big) \prec \lambda(H).$$
Show also that $H \ge \left(\begin{smallmatrix} I & I \\ I & I \end{smallmatrix}\right)$ to conclude that H has at least n eigenvalues greater than or equal to 2. [Hint: Expand $(I - XY)^{-1}$.]

. ——————— $\odot$ ——————— .

10.6 Majorization for Matrix Product

We have shown in the previous section (see (10.16)) that for complex matrices A and B of the same size

$$\sigma(A+B) \prec_w \sigma(A) + \sigma(B).$$

The analogue of this for product is: if A and B are $n \times n$ matrices,

$$\sigma(AB) \prec_w \sigma(A) \circ \sigma(B). \tag{10.24}$$

In particular, for $n \times n$ positive semidefinite matrices A and B,

$$\lambda(AB) \prec_w \lambda(A) \circ \lambda(B).$$

(10.24) was proved in Section 10.4 (see Corollary 10.3). In this section, we first present a different classic proof of it, study the majorization inequalities of the matrix product AB, then move on to show the majorization inequalities concerning the power of product (i.e., $(AB)^m$) and the product of the powers (i.e., $A^m B^m$).

Theorem 10.29 (Lidskiǐ) *Let A and B be $n \times n$ positive semidefinite matrices and let $1 \le i_1 < \cdots < i_k \le n$. Then*

$$\prod_{t=1}^{k} \lambda_{i_t}(AB) \le \prod_{t=1}^{k} \lambda_{i_t}(A)\lambda_t(B). \tag{10.25}$$

Equality holds when $k = n$.

Proof. Note that $\lambda_i(AB) \le \lambda_i(A)\lambda_1(B)$ for any positive semidefinite matrices A and B and any index i (Theorem 8.12). Apply this to the compound matrix $(AB)^{(k)}$. Because $\prod_{t=1}^{k} \lambda_{i_t}(AB)$ is an eigenvalue of $(AB)^{(k)}$ indexed by $\alpha = (i_1, \ldots, i_k)$, we have

$$\begin{aligned}
\prod_{t=1}^{k} \lambda_{i_t}(AB) &= \lambda_\alpha((AB)^{(k)}) = \lambda_\alpha((A^{(k)}B^{(k)}) \\
&\le \lambda_\alpha(A^{(k)})\lambda_1(B^{(k)}) = \prod_{t=1}^{k} \lambda_{i_t}(A)\lambda_t(B).
\end{aligned}$$

Equality holds when $k = n$ as $\det(AB) = \det A \det B$. ∎

Since $\sigma_j^2(X) = \lambda_j(X^*X)$ for any matrix X and index j, we have

Corollary 10.7 *Let $A, B \in \mathbb{M}_n$ and let $1 \le i_1 < \cdots < i_k \le n$. Then*

$$\prod_{t=1}^{k} \sigma_{i_t}(AB) \le \prod_{t=1}^{k} \sigma_{i_t}(A)\sigma_t(B). \tag{10.26}$$

Equality holds when $k = n$.

Taking $i_t = t$ in the corollary reveals the log-majorization

$$\sigma(AB) \prec_{\log} \sigma(A) \circ \sigma(B). \tag{10.27}$$

In particular, if A and B are positive semidefinite matrices, then

$$\lambda(AB) \prec_{\log} \lambda(A) \circ \lambda(B).$$

By Theorem 10.15, (10.27) implies the weak majorization (10.24).

The next theorem gives lower bounds for $\sigma(AB)$ and $\lambda(AB)$.

Theorem 10.30 *Let A and B be $n \times n$ complex matrices. Then*

$$\sigma(A) \circ \sigma^{\uparrow}(B) \prec_{\log} \sigma(AB); \quad \sigma(A) \circ \sigma^{\uparrow}(B) \prec_w \sigma(AB).$$

In particular, if A and B are positive semidefinite, then

$$\lambda(A) \circ \lambda^{\uparrow}(B) \prec_{\log} \lambda(AB); \quad \lambda(A) \circ \lambda^{\uparrow}(B) \prec_w \lambda(AB).$$

Proof. It is sufficient to show the first log-majorization. We may assume that A and B are nonsingular by continuity. Note that (10.26) also holds when AB on the left-hand side is replaced by BA. Taking logarithm for both sides of (10.26) with BA for AB, we get

$$\max_{1 \le i_1 < \cdots < i_k \le n} \sum_{i=1}^{k} (\ln \sigma_{i_t}(BA) - \ln \sigma_{i_t}(A)) \le \sum_{t=1}^{k} \ln \sigma_t(B).$$

It follows that $\ln \sigma(BA) - \ln \sigma(A) \prec \ln \sigma(B)$. With A replaced by $B^{-1}A$ and then B by AB, we obtain $\ln \sigma(A) - \ln \sigma(B^{-1}) \prec \ln \sigma(AB)$, i.e., $\ln(\sigma(A) \circ \sigma^{\uparrow}(B)) \prec \ln \sigma(AB)$, or $\sigma(A) \circ \sigma^{\uparrow}(B) \prec_{\log} \sigma(AB)$. ∎

In summary, we have for $n \times n$ complex matrices A and B

$$\sigma(A) \circ \sigma^{\uparrow}(B) \prec_{\log} \sigma(AB) \prec_{\log} \sigma(A) \circ \sigma(B); \tag{10.28}$$

$$\sigma(A) \circ \sigma^{\uparrow}(B) \prec_{w} \sigma(AB) \prec_{w} \sigma(A) \circ \sigma(B).$$

In particular, if A and B are positive semidefinite, then

$$\lambda(A) \circ \lambda^{\uparrow}(B) \prec_{\log} \lambda(AB) \prec_{\log} \lambda(A) \circ \lambda(B); \tag{10.29}$$

$$\lambda(A) \circ \lambda^{\uparrow}(B) \prec_{w} \lambda(AB) \prec_{w} \lambda(A) \circ \lambda(B).$$

We now study the majorization inequalities concerning the power of product, $(AB)^m$, and the product of the powers, $A^m B^m$.

By (10.27), it is immediate that for any positive integer m and $n \times n$ positive semidefinite matrices A and B,

$$\lambda(A^m B^m) \prec_{\log} \lambda^m(A) \circ \lambda^m(B)$$

(here $\lambda^{\alpha}(X) = (\lambda_1^{\alpha}(X), \ldots, \lambda_n^{\alpha}(X)) = ((\lambda_1(X))^{\alpha}, \ldots, (\lambda_n(X))^{\alpha})$). Thus

$$\lambda^{1/m}(A^m B^m) \prec_{\log} \lambda(A) \circ \lambda(B).$$

This says that $\lambda^{1/m}(A^m B^m)$ is bounded above in majorization by $\lambda(A) \circ \lambda(B)$. In what follows we show it is bounded below by $\lambda(AB)$, or equivalently, $\lambda((AB)^m) \prec_{\log} \lambda(A^m B^m)$. Putting these together,

$$\lambda(AB) \prec_{\log} \lambda^{1/m}(A^m B^m) \prec_{\log} \lambda(A) \circ \lambda(B).$$

In fact, more can be said about $\lambda^{1/m}(A^m B^m)$, as we show.

We have seen that for $n \times n$ positive semidefinite matrices A, B,

$$\lambda_1(AB) \leq \lambda_1(A)\lambda_1(B), \quad \lambda_n(AB) \geq \lambda_n(A)\lambda_n(B),$$

and that for arbitrary $n \times n$ matrices X, Y,

$$|\lambda_1(XY)| \leq \sigma_1(XY) \leq \sigma_1(X)\sigma_1(Y)$$

and

$$|\lambda_n(XY)| \geq \sigma_n(XY) \geq \sigma_n(X)\sigma_n(Y).$$

Theorem 10.31 *Let A and B be $n \times n$ positive semidefinite matrices. Then for any positive integers p and q, where $p \leq q$,*

$$\lambda_1(AB) \leq \lambda_1^{1/p}(A^p B^p) \leq \lambda_1^{1/q}(A^q B^q) \leq \lambda_1(A)\lambda_1(B)$$

and

$$\lambda_n(AB) \geq \lambda_n^{1/p}(A^p B^p) \geq \lambda_n^{1/q}(A^q B^q) \geq \lambda_n(A)\lambda_n(B).$$

Proof. We use induction on positive integer m to show that

$$\lambda_1^{1/m}(A^m B^m) \leq \lambda_1^{1/m+1}(A^{m+1} B^{m+1}). \tag{10.30}$$

If $m = 1$, then

$$\lambda_1(AB) \leq \sigma_1(AB) = \lambda_1^{1/2}(ABBA) = \lambda_1^{1/2}(A^2 B^2).$$

Suppose that the inequalities hold for integers no more than m. Put

$$X = A^{(m+1)/2} B^{(m+1)/2}, \quad Y = B^{(m-1)/2} A^{(m-1)/2}.$$

Then $\lambda_1(XY) = \lambda_1(A^m B^m)$. However,

$$\lambda_1(XY) \leq \sigma_1(X)\sigma_1(Y) = \lambda_1^{1/2}(A^{m+1} B^{m+1})\lambda_1^{1/2}(A^{m-1} B^{m-1}).$$

By induction hypothesis on $m - 1$, we have

$$\begin{aligned} \lambda_1(A^m B^m) &\leq \lambda_1^{1/2}(A^{m+1} B^{m+1})\lambda_1^{1/2}(A^{m-1} B^{m-1}) \\ &\leq \lambda_1^{1/2}(A^{m+1} B^{m+1})\lambda_1^{(m-1)/2m}(A^m B^m). \end{aligned}$$

Thus

$$\lambda_1^{(m+1)/2m}(A^m B^m) \leq \lambda_1^{1/2}(A^{m+1} B^{m+1});$$

that is,

$$\lambda_1^{1/m}(A^m B^m) \leq \lambda_1^{1/(m+1)}(A^{m+1} B^{m+1}).$$

This proves (10.30). On the other hand, for any positive integer m,

$$\lambda_1^{1/m}(A^m B^m) \leq \lambda_1^{1/m}(A^m)\lambda_1^{1/m}(B^m) = \lambda_1(A)\lambda_1(B).$$

The second part is similarly proven. ∎

Theorem 10.32 (Wang–Gong) *Let A and B be $n \times n$ positive semidefinite matrices. Then for any positive integers p and q, $p \leq q$,*

$$\lambda(AB) \prec_{\log} \lambda^{1/p}(A^p B^p) \prec_{\log} \lambda^{1/q}(A^q B^q) \prec_{\log} \lambda(A) \circ \lambda(B).$$

Proof. We only show that for any positive integer m,

$$\lambda^{1/m}(A^m B^m) \prec_{\log} \lambda^{1/m+1}(A^{m+1} B^{m+1}).$$

Considering the compound matrix $(A^m B^m)^{(k)}$, where $k \leq n$, we have

$$(A^m B^m)^{(k)} = (A^m)^{(k)} (B^m)^{(k)} = (A^{(k)})^m (B^{(k)})^m.$$

Moreover,

$$\lambda_1\Big((A^{(k)})^m (B^{(k)})^m\Big) = \lambda_1\Big((A^m B^m)^{(k)}\Big) = \prod_{i=1}^{k} \lambda_i(A^m B^m).$$

Note that $A^{(k)}$ and $B^{(k)}$ are positive semidefinite. By (10.30),

$$\begin{aligned}
\prod_{i=1}^{k} \lambda_i^{1/m}(A^m B^m) &= \lambda_1^{1/m}\Big((A^{(k)})^m (B^{(k)})^m\Big) \\
&\leq \lambda_1^{1/m+1}\Big((A^{(k)})^{m+1} (B^{(k)})^{m+1}\Big) \\
&= \prod_{i=1}^{k} \lambda_i^{1/m+1}(A^{m+1} B^{m+1}).
\end{aligned}$$

When $k = n$, equality holds as for any positive integer m,

$$\prod_{i=1}^{n} \lambda_i^{1/m}(A^m B^m) = \det A \det B. \quad \blacksquare$$

The previous theorem refines $\lambda(AB) \prec_{\log} \lambda(A) \circ \lambda(B)$. Our next theorem shows an analogous result for $\lambda(A) \circ \lambda^{\uparrow}(B) \prec_{\log} \lambda(AB)$.

Theorem 10.33 (Wang–Gong) *Let A and B be $n \times n$ positive semidefinite matrices. Then for any positive integers p and q, $p \leq q$,*

$$\lambda(A) \circ \lambda^{\uparrow}(B) \prec_{\log} \lambda^{q}(A^{1/q} B^{1/q}) \prec_{\log} \lambda^{p}(A^{1/p} B^{1/p}) \prec_{\log} \lambda(AB).$$

Proof. By the first log-majorization of (10.29), we have

$$\lambda(A^{1/m}) \circ \lambda^{\uparrow}(B^{1/m}) \prec_{\log} \lambda(A^{1/m}B^{1/m}).$$

Raising both sides to the mth power, we arrive at

$$\lambda(A) \circ \lambda^{\uparrow}(B) \prec_{\log} \lambda^m(A^{1/m}B^{1/m}).$$

We now show that

$$\lambda^{m+1}(A^{1/m+1}B^{1/m+1}) \prec_{\log} \lambda^m(A^{1/m}B^{1/m}).$$

By Theorem 10.31, we have (Problem 5)

$$\lambda_1^{m+1}(A^{1/m+1}B^{1/m+1}) \leq \lambda_1^m(A^{1/m}B^{1/m}) \leq \lambda_1(AB).$$

The desired log-majorizations follow by making use of compound matrices as we did in the proof of Theorem 10.32. ∎

The following corollaries are immediate from the theorems. For instance, with A replaced by A^m and B by B^m in the log-majorization $\lambda^m(A^{1/m}B^{1/m}) \prec_{\log} \lambda(AB)$, we obtain $\lambda^m(AB) \prec_{\log} \lambda(A^mB^m)$, which results in the weak majorization $\lambda^m(AB) \prec_w \lambda(A^mB^m)$.

Corollary 10.8 *Let A and B be $n \times n$ positive semidefinite matrices. Then for any positive integer m,*

$$\lambda^m(A) \circ (\lambda^m(B))^{\uparrow} \prec_{\log} \lambda^m(AB) \prec_{\log} \lambda(A^mB^m) \prec_{\log} \lambda^m(A) \circ \lambda^m(B)$$

and

$$\lambda^m(A) \circ (\lambda^m(B))^{\uparrow} \prec_w \lambda^m(AB) \prec_w \lambda(A^mB^m) \prec_w \lambda^m(A) \circ \lambda^m(B).$$

Corollary 10.9 (Lieb–Thirring) *Let A and B be $n \times n$ positive semidefinite matrices. Then for any positive integer m,*

$$\operatorname{tr}(AB)^m \leq \operatorname{tr}(A^mB^m). \tag{10.31}$$

Equality holds if and only if $m = 1$ or $AB = BA$.

Proof. The inequality is readily seen from the previous corollary. We show that equality occurs if and only if $m = 1$ or $AB = BA$.

Sufficiency is obvious. For necessity, assume that equality holds for some $m \geq 2$. We first consider the case $m = 2$; that is, $\operatorname{tr}(AB)^2 = \operatorname{tr}(A^2B^2)$. Without loss of generality, we may assume that A is a diagonal matrix with diagonal entries $a_1, \dots, a_n$. Then

$$\begin{aligned}\operatorname{tr}(A^2B^2) - \operatorname{tr}(AB)^2 &= \sum_{i,j} a_i^2 |b_{ij}|^2 - \sum_{i,j} a_i a_j |b_{ij}|^2 \\ &= \sum_{i<j} (a_i - a_j)^2 |b_{ij}|^2 = 0.\end{aligned}$$

Thus $a_i b_{ij} = a_j b_{ij}$ for every pair of i and j; that is, $AB = BA$.

Let $m > 2$. We show that $\operatorname{tr}(AB)^m = \operatorname{tr}(A^mB^m)$ implies $\operatorname{tr}(AB)^2 = \operatorname{tr}(A^2B^2)$, which leads to $AB = BA$, as we have just proven.

If $\operatorname{tr}(AB)^2 \neq \operatorname{tr}(A^2B^2)$, we apply the strictly increasing and strictly convex function $f(t) = t^{m/2}$, $t > 0$, to the weak majorization $\lambda^2(AB) \prec_w \lambda(A^2B^2)$. By Theorem 10.14, we arrive at

$$\operatorname{tr}(AB)^m = \sum_{i=1}^n \lambda_i^{m/2}((AB)^2) < \sum_{i=1}^n \lambda_i^{m/2}(A^2B^2).$$

On the other hand, since $x \prec_w y \Rightarrow x^m \prec_w y^m$ for $x, y \in \mathbb{R}_+^n$ (Problem 22, Section 10.3) and $\lambda^{1/2}(A^2B^2) \prec_w \lambda^{1/m}(A^mB^m)$ (Theorem 10.32), which results in $\lambda^{m/2}(A^2B^2) \prec_w \lambda(A^mB^m)$, we have

$$\sum_{i=1}^n \lambda_i^{m/2}(A^2B^2) \leq \sum_{i=1}^n \lambda_i(A^mB^m) = \operatorname{tr}(A^mB^m).$$

This contradicts the assumption that $\operatorname{tr}(AB)^m = \operatorname{tr}(A^mB^m)$. ∎

Problems

1. Show that the inequalities in Theorem 10.31 hold for singular values σ_1 (resp., σ_n) in place of λ_1 (resp., λ_n) when A and B are normal matrices. If A or B is not normal, then they are not true in general: Take $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $B = I_2$ as an example.

2. Let A and B be positive semidefinite matrices of the same size, and p and q be positive numbers with $1 < p < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$. Show that
$$\operatorname{tr}(A^{1/p}B^{1/q}) \le (\operatorname{tr} A)^{1/p}(\operatorname{tr} B)^{1/q}$$
and
$$\big(\operatorname{tr}(A+B)\big)^{1/p} \le (\operatorname{tr} A^p)^{1/p} + (\operatorname{tr} B^q)^{1/q}.$$

3. Let A be an $n \times n$ positive semidefinite matrix, and p and q be positive numbers such that $1 < p < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$. Show that for all $n \times n$ positive semidefinite matrices B with $\operatorname{tr} B^q = 1$, the inequality
$$\operatorname{tr}(AB) \le (\operatorname{tr} A^p)^{1/p}$$
holds. Show that equality occurs if and only if $B^q = \frac{1}{\operatorname{tr} A^p} A^p$.

4. Let A and B be positive semidefinite matrices of the same size. Show that $\lambda_1^m(AB) \le \lambda_1(A^m)\lambda_1(B^m)$ for any positive integer m.

5. Let A and B be positive semidefinite matrices of the same size. Show that $\lambda_1^m(A^{1/m}B^{1/m})$ decreases as m increases; that is,
$$\lambda_1^{m+1}(A^{1/m+1}B^{1/m+1}) \le \lambda_1^m(A^{1/m}B^{1/m}).$$

6. Show Theorem 10.32 for singular values with A and B both normal.

7. Let $A \circ B = (a_{ij}b_{ij})$ be the Hadamard product of $A = (a_{ij})$ and $B = (b_{ij})$. Show, with $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, that the inequality
$$\sum_{i=1}^{n} \lambda_i^m(A)\lambda_{n-i+1}^m(B) \le \operatorname{tr}\left((A \circ B)^m\right)$$
does not hold in general for positive semidefinite matrices A and B.

8. Let A be a square matrix. Show that
$$\sigma^2(A) \prec \sigma(A^2) \quad \Leftrightarrow \quad A \text{ is normal}.$$

9. Referring to (10.28), show that for $n \times n$ complex matrices A and B, $\sigma(A) \circ \sigma^{\uparrow}(B) \prec_{\log} \sigma(BA)$. Note that $\sigma(AB) \ne \sigma(BA)$ in general.

10. Let A and B be positive semidefinite and $0 \le r \le 1$. Show that
$$\lambda_1(A^r B^r) \le \lambda_1^r(AB).$$
The inequality is reversed for $r \ge 1$.

11. Let A and B be $n \times n$ positive semidefinite matrices and k be a positive integer, $k \le n$. If $1 \le i_1 < i_2 < \cdots < i_k \le n$, show that

    (a) $\prod_{t=1}^{k} \lambda_t(AB) \ge \prod_{t=1}^{k} \lambda_{i_t}(A)\lambda_{n-i_t+1}(B)$.

    (b) $\sum_{t=1}^{k} \lambda_t(AB) \ge \sum_{t=1}^{k} \lambda_{i_t}(A)\lambda_{n-i_t+1}(B)$.

    (c) $\operatorname{tr}(AB) \ge \sum_{t=1}^{n} \lambda_t(A)\lambda_{n-t+1}(B)$.

12. Let A and B be $n \times n$ complex matrices and k be a positive integer, $k \le n$. If $1 \le i_1 < i_2 < \cdots < i_k \le n$, show that
$$\prod_{t=1}^{k} \sigma_{i_t}(AB) \ge \prod_{t=1}^{k} \sigma_{i_t}(A)\sigma_{n-t+1}(B),$$
but it is not true in general that
$$\sum_{t=1}^{k} \sigma_{i_t}(AB) \ge \sum_{t=1}^{k} \sigma_{i_t}(A)\sigma_{n-t+1}(B).$$

13. Let A, B, and C be $n \times n$ positive definite matrices. Show that
$$\ln \lambda(A^{-1}C) \prec \ln \lambda(A^{-1}B) + \ln \lambda(B^{-1}C).$$

14. Let A be an $n \times n$ matrix. Show that for any positive integer k,
$$\sigma(A^k) \prec_{\log} \sigma^k(A).$$
Deduce that
$$\operatorname{tr}\left((A^*)^k A^k\right) \le \operatorname{tr}(A^*A)^k.$$
Show that equality holds if and only if A is a normal matrix.

15. Let A and B be $n \times n$ positive semidefinite matrices and k and m be positive integers, $k \le m$. Show that
$$\operatorname{tr}(A^k B^k)^m \le \operatorname{tr}(A^m B^m)^k.$$

16. Let A and B be $n \times n$ Hermitian matrices and m be a positive integer. Show that
$$|\operatorname{tr}(AB)^{2m}| \le \operatorname{tr}(A^{2m}B^{2m});$$
$$|\operatorname{tr}(A^m B^m)^2| \le \operatorname{tr}(A^{2m}B^{2m}).$$
Show by the example $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and $B = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}$ that in general
$$|\operatorname{tr}(AB)^3| \not\le \operatorname{tr}(A^3B^3).$$

10.7 Majorization and Unitarily Invariant Norms

This section shows a close relation between the weak majorization and unitarily invariant matrix (vector) norms. To be precise, for $A, B \in \mathbb{M}_{m\times n}$, we show that $\sigma(A) \prec_w \sigma(B)$ if and only if $\|A\| \leq \|B\|$ for all unitarily invariant matrix norms $\|\cdot\|$ on $\mathbb{M}_{m\times n}$. Note that $\sigma(A) = \lambda^{1/2}(A^*A)$ is an n-vector for $m \times n$ matrix A. The symmetric gauge functions that we introduce below serve as a "bridge" between majorization and the matrix norm.

A *symmetric gauge function* is a real-valued function defined on $\mathbb{R}^n$ that is invariant under any permutation of the components of the vectors and any change of signs of the components. To be exact, $\phi : \mathbb{R}^n \mapsto \mathbb{R}$ is a *symmetric gauge function* if the following are satisfied.

a. $\phi(x) \geq 0$. And $\phi(x) = 0$ if and only if $x = 0$.

b. $\phi(cx) = |c|\phi(x)$, $c \in \mathbb{R}$.

c. $\phi(x + y) \leq \phi(x) + \phi(y)$.

d. $\phi(x_1, x_2, \dots, x_n) = \phi(x_{p(1)}, x_{p(2)}, \dots, x_{p(n)})$, where $p \in S_n$.

e. $\phi(x_1, x_2, \dots, x_n) = \phi(\epsilon_1 x_1, \epsilon_2 x_2, \dots, \epsilon_n x_n)$, where all $\epsilon_i = \pm 1$.

Apparently, a symmetric gauge function is a norm on $\mathbb{R}^n$. Recall that a vector norm is a function satisfying (a), (b), and (c).

One may check that $\varphi(x) = \max_i |x_i|$ and $\psi(x) = \sum_i |x_i|$ are symmetric gauge functions. Moreover, if ϕ is a symmetric gauge function, then (a) $\phi(x) = \phi(|x|)$ and (b) $0 \leq x \leq y \Rightarrow \phi(x) \leq \phi(y)$. We demonstrate the proof of (b) for the case $n = 2$. Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$. Let t_1 be such that $t_1 y_1 = x_1$. Then $0 \leq t_1 \leq 1$ and

$$\begin{aligned}
\phi(x) &= \phi(x_1, x_2) = \phi(t_1 y_1, x_2) \\
&= \phi\left(\frac{1+t_1}{2}(y_1, x_2) + \frac{1-t_1}{2}(-y_1, x_2)\right) \\
&\leq \frac{1+t_1}{2}\phi(y_1, x_2) + \frac{1-t_1}{2}\phi(-y_1, x_2) \\
&= \phi(y_1, x_2).
\end{aligned}$$

Likewise, $\phi(y_1, x_2) \leq \phi(y_1, y_2) = \phi(y)$. Thus $\phi(x) \leq \phi(y)$.

One can prove (Problem 5) that the l_p-*norm* on $\mathbb{R}^n$

$$\|x\|_p = \Big(\sum_{i=1}^n |x_i|^p\Big)^{1/p}, \quad p \geq 1$$

is a symmetric gauge function, and so is the *Ky Fan* k-*norm* on $\mathbb{R}^n$

$$\|x\|_{(k)} = \max_{1\leq i_1<\cdots<i_k\leq n} \sum_{t=1}^k |x_{i_t}|.$$

Theorem 10.34 *Symmetric gauge functions are Schur-convex on* $\mathbb{R}^n_+$.

Proof. Let x, $y \in \mathbb{R}^n_+$ and $x \prec y$. We need to show that $\phi(x) \leq \phi(y)$ for any symmetric gauge function ϕ on $\mathbb{R}^n$. By Theorem 10.8, there exists a doubly stochastic matrix A such that $x = yA$. By Theorem 5.21, every doubly stochastic matrix is a convex combination of permutation matrices. Write $x = \sum_{i=1}^m t_i y P_i$, where t_i are positive numbers adding up to 1 and P_i are permutation matrices. Then

$$\phi(x) = \phi\Big(\sum_{i=1}^m t_i y P_i\Big) \leq \sum_{i=1}^m t_i \phi(yP_i) = \sum_{i=1}^m t_i \phi(y) = \phi(y). \quad \blacksquare$$

Theorem 10.35 *Let* x, $y \in \mathbb{R}^n$. *Then* $|x| \prec_w |y|$ *if and only if* $\phi(x) \leq \phi(y)$ *for all symmetric gauge functions* ϕ *on* $\mathbb{R}^n$.

Proof. For sufficiency, take ϕ to be the Ky Fan k-norm, $k = 1, 2, \ldots, n$. Then $\phi(x) \leq \phi(y)$ yields $|x| \prec_w |y|$. For necessity, since $|x| \prec_w |y|$, there exists a nonnegative vector u such that $|x| \prec u \leq |y|$. Theorem 10.34 implies that $\phi(x) = \phi(|x|) \leq \phi(u) \leq \phi(|y|) = \phi(y)$. $\blacksquare$

Recall that a *matrix-vector norm* (see Section 4.2, Chapter 4) on the vector space $\mathbb{M}_{m\times n}$ is a function $\|\cdot\|$ from $\mathbb{M}_{m\times n}$ to $\mathbb{R}_+$ satisfying

i. $\|A\| \geq 0$, and $\|A\| \geq 0$ if and only if $A = 0$,

ii. $\|cA\| = |c|\|A\|$, $c \in \mathbb{C}$, and

iii. $\|A + B\| \leq \|A\| + \|B\|$.

It is further said to be *unitarily invariant* if

v. $\|UAV\| = \|A\|$ for all $A \in \mathbb{M}_{m\times n}$ and unitary U and V.

Note that the multiplicative condition $\|AB\| \leq \|A\|\|B\|$ is not required for a matrix-vector norm.

If ϕ is a symmetric gauge function on $\mathbb{R}^n$ and A is an $m \times n$ matrix, then $\sigma(A) \in \mathbb{R}^n$ and $\phi(\sigma(A))$ is well defined. Moreover, $\phi(\sigma(UAV)) = \phi(\sigma(A))$ for all $m \times m$ unitary U and $n \times n$ unitary V. The following two theorems, due to von Neumann, best characterize the relation between a unitarily invariant matrix-vector norm and symmetric gauge functions through singular values.

Theorem 10.36 *If $\phi : \mathbb{R}^n \mapsto \mathbb{R}$ is a symmetric gauge function, then $\|A\|_\phi = \phi(\sigma(A))$ is a unitarily invariant norm on $\mathbb{M}_{m\times n}$.*

Proof. Conditions (i) and (ii) are obviously satisfied. To show (iii), let A and B be $m \times n$ matrices. Because $\sigma(A+B) \prec_w \sigma(A) + \sigma(B)$,

$$\begin{aligned} \|A+B\|_\phi &= \phi(\sigma(A+B)) \leq \phi(\sigma(A) + \sigma(B)) \\ &\leq \phi(\sigma(A)) + \phi(\sigma(B)) = \|A\|_\phi + \|B\|_\phi. \end{aligned}$$

Thus $\|\cdot\|_\phi$ is a matrix-vector norm. It is unitarily invariant because $\sigma(UAV) = \sigma(A)$ for any $m \times m$ unitary U and $n \times n$ unitary V. ∎

Theorem 10.37 *If $\|\cdot\|$ is a unitarily invariant (matrix-vector) norm on $\mathbb{M}_{m\times n}$, then there exists a symmetric gauge function ϕ on $\mathbb{R}^n$ such that $\|A\| = \phi(\sigma(A))$ for all $m \times n$ complex matrices A.*

Proof. If $m \geq n$, for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, define an $m \times n$ matrix M_x whose (i,i)-entry is x_i, $i = 1, 2, \ldots, n$, and 0 elsewhere. Let $\phi(x) = \|M_x\|$. Then ϕ is a function from $\mathbb{R}^n$ to $\mathbb{R}_+$ and it is readily seen that it satisfies (a) and (b). For (d) and (e), it suffices to note that $\|PM_xQ\| = \|M_x\|$ for permutation matrices P and Q and for diagonal matrices Q with ± 1 on the main diagonal. For (c), we have

$$\begin{aligned} \phi(x+y) &= \|M_{x+y}\| = \|M_x + M_y\| \\ &\leq \|M_x\| + \|M_y\| = \phi(x) + \phi(y). \end{aligned}$$

Thus, $\phi : \mathbb{R}^n \mapsto \mathbb{R}_+$ is a symmetric gauge function. For any $m \times n$ matrix A, $\sigma(A) = \lambda^{1/2}(A^*A) \in \mathbb{R}^n_+$. Let $A = UA_\sigma V$ be a singular value decomposition of A, where $A_\sigma = M_{\sigma(A)}$. Then

$$\|A\| = \|A_\sigma\| = \|M_{\sigma(A)}\| = \phi(\sigma(A)).$$

If $m < n$, we define $\|X\|^{\tilde{}} = \|X^T\|$, where X is $n \times m$. Then $\|\cdot\|^{\tilde{}}$ is a unitarily invariant norm on $\mathbb{M}_{n\times m}$. The above argument ensures a symmetric gauge function $\varphi : \mathbb{R}^m \mapsto \mathbb{R}_+$ such that $\varphi(\sigma(X)) = \|X\|^{\tilde{}} = \|X^T\|$. Define $\phi : \mathbb{R}^n \mapsto \mathbb{R}_+$ by $\phi(x) = \varphi(\tilde{x})$, where $\tilde{x}$ is the m-vector consisting of the first m largest components of x in absolute value. Then ϕ is a symmetric gauge function on $\mathbb{R}^n$ (Problem 4).

Now for $A \in \mathbb{M}_{m\times n}$, A^T is $n \times m$, $\sigma(A)$ is an n-vector, and $\tilde{\sigma}(A)$ is an m-vector that differs from $\sigma(A)$ by $n - m$ zeros. Notice that A^*A, AA^*, and $\overline{AA^*} = (A^T)^*A^T$ have the same nonzero eigenvalues. Thus $\tilde{\sigma}(A) = \sigma(A^T)$. It follows that

$$\phi(\sigma(A)) = \varphi(\tilde{\sigma}(A)) = \varphi(\sigma(A^T)) = \|A^T\|^{\tilde{}} = \|A\|. \quad \blacksquare$$

Theorem 10.38 (von Neumann) *Let A, $B \in \mathbb{M}_{m\times n}$. Then*

$$\sigma(A) \prec_w \sigma(B) \quad \Leftrightarrow \quad \|A\| \le \|B\|$$

for all unitarily invariant matrix-vector norms $\|\cdot\|$ on $\mathbb{M}_{m\times n}$.

Proof. Theorem 10.37 says that unitarily invariant norms are essentially the same as symmetric gauge functions. On the other hand, Theorem 10.35 ensures that $\sigma(A) \prec_w \sigma(B)$ if and only if $\phi(\sigma(A)) \le \phi(\sigma(B))$ for all symmetric gauge functions ϕ. $\blacksquare$

Theorem 10.39 (Fan Dominance Theorem) *Let A, $B \in \mathbb{M}_{m\times n}$. Then $\|A\| \le \|B\|$ for all unitarily invariant matrix-vector norm $\|\cdot\|$ on $\mathbb{M}_{m\times n}$ if and only if $\|A\|_{(k)} \le \|B\|_{(k)}$, $k = 1, 2, \ldots, q = \min\{m, n\}$, where $\|A\|_{(k)}$ denote the Ky Fan k-norms on $\mathbb{M}_{m\times n}$:*

$$\|A\|_{(k)} = \sum_{i=1}^{k} \sigma_i(A), \quad k = 1, \ldots, q = \min\{m, n\}.$$

Proof. Ky Fan k-norms on $\mathbb{M}_{m\times n}$ are unitarily invariant; so necessity is obvious. Conversely, that $\|A\|_{(k)} \le \|B\|_{(k)}$ for each k is the same as $\sigma(A) \prec_w \sigma(B)$. By Theorem 10.38, the sufficiency follows. $\blacksquare$

A variety of matrix inequalities on unitarily invariant norms follows at once from Theorem 10.38. For instance, by Theorem 10.25, for any unitarily invariant norm $\|\cdot\|$ on $\mathbb{M}_n$, $A, B \in \mathbb{M}_n$, and $z \in \mathbb{C}$,

$$\|A - |z|B\| \le \|A + zB\| \le \|A + |z|B\|,$$

and Corollary 10.8 ensures that for a unitarily invariant norm $\|\cdot\|$ on $\mathbb{M}_n$, positive semidefinite $A, B \in \mathbb{M}_n$, and positive integers m,

$$\|(AB)^m\| \leq \|A^m B^m\|.$$

Problems

1. If $\|\cdot\|$ is a vector norm on $\mathbb{R}^n$ satisfying $\|xP\| = \|x\|$ for any $n \times n$ permutation matrix P, show that such a vector norm is Schur-convex.

2. Show that a symmetric gauge function on $\mathbb{R}^n$ is a convex function.

3. Let $\|\cdot\|$ be a vector norm on $\mathbb{R}^n$. Show that $\|\,|x|\,\| = \|x\|$ for any $x \in \mathbb{R}^n$ if and only if $0 \leq x \leq y \Rightarrow \|x\| \leq \|y\|$, where $|x| = (|x_1|, \ldots, |x_n|)$.

4. Let φ be a symmetric gauge function on $\mathbb{R}^m$. Let $n > m$. For $x \in \mathbb{R}^n$, let $\tilde{x}$ be the vector of the m largest components of x in absolute value. Show that $\phi(x) = \varphi(\tilde{x})$ is a symmetric gauge function on $\mathbb{R}^n$.

5. Show that the l_p-norm $\|\cdot\|_p$ (or the *Schatten p-norm*) and the Ky Fan k-norm $\|\cdot\|_{(k)}$ defined on $\mathbb{R}^n$ are symmetric gauge functions:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \quad p \geq 1; \quad \|x\|_{(k)} = \max_{1 \leq i_1 < \cdots < i_k \leq n} \sum_{t=1}^{k} |x_{i_t}|.$$

6. For $m \times n$ matrix A, let $\|A\|_{(k)}$ be given as in Theorem 10.39; for $x \in \mathbb{R}^n$, let $\|x\|_{(k)}$ be defined as in the previous problem. Show that

$$\|A\|_{(k)} = \|\sigma(A)\|_{(k)}.$$

7. Extend the definition of symmetric gauge function for $\mathbb{C}^n$ as follows. Replace the conditions (b) and (e) by, respectively, (b'). $\phi(cx) = |c|\phi(x)$, $c \in \mathbb{C}$ and (e'). $\phi(x) = \phi(|x|)$, where $|x| = (|x_1|, \ldots, |x_n|)$. For $x, y \in \mathbb{C}^n$, show that $|x| \prec_w |y|$ if and only if $\phi(x) \leq \phi(y)$ for all symmetric gauge functions ϕ on $\mathbb{C}^n$.

8. We have used $\|\cdot\|_{(k)}$ to denote the Ky Fan k-norms for both matrices (in Theorem 10.39) and vectors (in Problem 5). If M is a matrix with singular value vector $v = \sigma(M)$, show that $\|M\|_{(k)} = \|v\|_{(k)}$. Let $A = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right)$. Compute $\|A\|_{(1)}$. If A is considered as a vector in $\mathbb{R}^4$, say u, find $\|u\|_{(1)}$. Are $\|A\|_{(1)}$ and $\|u\|_{(1)}$ the same?

9. Show that $\|A\| = \|\,|A|\,\|$ for any unitarily invariant norm $\|\cdot\|$ on $\mathbb{M}_n$ and $A \in \mathbb{M}_n$, where $|A| = (A^*A)^{1/2}$.

10. Show that $\|\cdot\|$ defined below is a unitarily invariant norm on $\mathbb{M}_n$:

$$\|A\| = \max\{|\operatorname{tr}(AX)| : X \in \mathbb{M}_n,\ \operatorname{tr}(X^*X) = 1\}.$$

11. Let $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq 0$ be given. Define

$$\|A\|_\alpha = \sum_{i=1}^{n} \alpha_i \sigma_i(A), \quad A \in \mathbb{M}_n.$$

Show that $\|\cdot\|_\alpha$ is a unitarily invariant norm on $\mathbb{M}_n$.

12. Let $A \in \mathbb{M}_n$ and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of A. Denote by Λ the diagonal matrix $\operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Show that for every unitarily invariant norm on $\mathbb{M}_n$, $\|\Lambda\| \leq \|A\|$.

13. Let A and B be $m \times n$ complex matrices with singular values $\sigma_1(A) \geq \cdots \geq \sigma_n(A)$ and $\sigma_1(B) \geq \cdots \geq \sigma_n(B)$, respectively. Show that

$$\|\operatorname{diag}(\sigma_1(A) - \sigma_1(B), \ldots, \sigma_n(A) - \sigma_n(B))\| \leq \|A - B\|$$

for unitarily invariant norms on $\mathbb{M}_{m\times n}$. [Hint: Use Theorem 10.24.]

14. Let $A, B \in \mathbb{M}_n$. Show that for every unitarily invariant norm on $\mathbb{M}_n$,

$$2\|AB\| \leq \|A^*A\| + \|B^*B\|.$$

15. Let $A, B \in \mathbb{M}_n$. As is known, $|A + B| \leq |A| + |B|$ is not true in general, where $|X| = (X^*X)^{1/2}$ (see Section 8.6). However, show by Theorem 8.22 that for any unitarily invariant norm $\|\cdot\|$ on $\mathbb{M}_n$,

$$\|\,|A + B|\,\| \leq \|\,|A|\,\| + \|\,|B|\,\|.$$

16. Let $A, B \in \mathbb{M}_n$. Show that for unitarily invariant norms $\|\cdot\|$ on $\mathbb{M}_{2n}$,

$$\left\|\frac{A+B}{2} \oplus \frac{A+B}{2}\right\| \leq \|A \oplus B\| = \|\,|A| \oplus |B|\,\| \leq \|(|A| + \|B\|) \oplus 0\|.$$

17. Let $A = U\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}V$ be a singular value decomposition of A, where U and V are $m\times m$ and $n\times n$ unitary matrices, respectively, D is an $r\times r$ positive diagonal matrix, and $r = \operatorname{rank}(A)$. Let $A^\dagger$ be the *Moore–Penrose inverse* of A, i.e., $A^\dagger = V^*\begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix}U^*$. Let $A, B \in \mathbb{M}_{m\times n}$. Show that for any $n \times n$ matrix X and unitarily invariant norm $\|\cdot\|$,

$$\|A(A^\dagger B) - B\| \leq \|AX - B\|.$$

18. Let A be an $m \times n$ complex matrix. Show that the Moore–Penrose inverse $A^{\dagger}$ of A is the only matrix satisfying all the equations

$$AXA = A, \quad XAX = X, \quad (AX)^* = AX, \quad (XA)^* = XA.$$

19. Give a unitarily invariant matrix-vector norm $\|\cdot\|$ on $\mathbb{M}_n$, for which $\|AB\| \leq \|A\|\,\|B\|$ does not hold for some matrices $A, B \in \mathbb{M}_n$.

20. Let $A \in \mathbb{M}_n$ and $\|\cdot\|$ denote a unitarily invariant norm on $\mathbb{M}_n$. Prove
    (a) For any Hermitian $X \in \mathbb{M}_n$, $\|A - \frac{1}{2}(A + A^*)\| \leq \|A - X\|$.
    (b) For any skew-Hermitian $X \in \mathbb{M}_n$, $\|A - \frac{1}{2}(A - A^*)\| \leq \|A - X\|$.
    (c) If $A = UP$ is a polar decomposition of A, where U is unitary and P is positive semidefinite, then for any unitary $X \in \mathbb{M}_n$, $\|A - U\| \leq \|A - X\| \leq \|A + U\|$.
    (d) If $A = UDV$ is a singular value decomposition of A, where U and V are unitary, and D is nonnegative diagonal, then for any unitary $X \in \mathbb{M}_n$, $\|A - UV\| \leq \|A - X\| \leq \|A + UV\|$.
    (e) If A is normal, then $\|A\| \leq \|X^{-1}AX\|$ for all invertible $X \in \mathbb{M}_n$.
    (f) If A is positive semidefinite, then $\|A - I\| \leq \|A - X\| \leq \|A + I\|$ for any unitary $X \in \mathbb{M}_n$.

21. Let A and B be $n \times n$ normal matrices. Let $A \circ B$ be the Hadamard (entrywise) product of A and B and $|A| = (A^*A)^{1/2}$. Show that

$$\|A \circ B\| \leq \|\,|A| \circ |B|\,\|$$

for unitarily invariant norms on $\mathbb{M}_n$. [Hint: Note that $\begin{pmatrix} |A| & A^* \\ A & |A^*| \end{pmatrix} \geq 0$.]

22. Let $A, B \in \mathbb{M}_n$. If the product AB is normal, show that $\|AB\| \leq \|BA\|$ for any unitarily invariant matrix-vector norm $\|\cdot\|$.

23. Let p and q be positive real numbers such that $\frac{1}{p} + \frac{1}{q} = 1$. Show that

$$\|AB\|^{pq} = \|AB\|^{p+q} \leq \|\,|A|^p\|\,\|\,|B|^q\|$$

for all $A, B \in \mathbb{M}_n$ and for all unitarily invariant norms $\|\cdot\|$ on $\mathbb{M}_n$.

24. **(Horn–Johnson)** Let $\|\cdot\|$ be a unitarily invariant matrix-vector norm on $\mathbb{M}_n$. Show that the following statements are equivalent.
    (a) $\|\operatorname{diag}(1, 0, \ldots, 0)\| \geq 1$.
    (b) $\|A\| \geq \sigma_{\max}(A)$ for all matrices $A \in \mathbb{M}_n$.
    (c) $\|A\| \geq \rho(A)$, the spectral radius of A, for all matrices $A \in \mathbb{M}_n$.
    (d) $\|AB\| \leq \|A\|\,\|B\|$ for all matrices $A, B \in \mathbb{M}_n$.

References

Books:

T. Ando, *Norms and Cones in Tensor Products of Matrices,* Preprint, 2001.

T. Ando, *Operator-Theoretic Methods for Matrix Inequalities,* Preprint, 1989.

R. B. Bapat and T. E. S. Raghavan, *Nonnegative Matrices and Applications,* Cambridge University Press, New York, 1997.

E. F. Beckenbach and R. Bellman, *Inequalities,* Springer-Verlag, New York, Fourth printing, 1983.

R. Bellman, *Introduction to Matrix Analysis,* SIAM, Philadelphia, Reprint of the Second edition, 1995.

A. Berman and R. Plemmons, *Nonnegative Matrices in the Mathematical Sciences,* SIAM, Philadelphia, 1994.

D. S. Bernstein, *Matrix Mathematics,* Princeton University Press, Princeton, NJ, 2005.

R. Bhatia, *Matrix Analysis,* Springer-Verlag, New York, 1997.

R. Bhatia, *Positive Definite Matrices,* Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2007.

T. S. Blyth and E. F. Robertson, *Further Linear Algebra,* Springer, New York, 2002.

B. Bollobás, *Linear Analysis: An Introductory Course,* 2nd edition, Cambridge University Press, New York, 1999.

R. A. Brualdi, *Combinatorial Matrix Classes* (Encyclopedia of Mathematics and its Applications), Cambridge University Press, New York, 2006.

R. A. Brualdi and D. Cvetkovic, *A Combinatorial Approach to Matrix Theory and Its Applications* (Discrete Mathematics and Its Applications), CRC Press, Boca Raton, FL, 2009.

R. Brualdi and H. J. Ryser, *Combinatorial Matrix Theory,* Cambridge University Press, New York, 1991.

D. Carlson, C. R. Johnson, D. C. Lay, and A. D. Porter, *Linear Algebra Gems: Assets for Undergraduate Mathematics,* Mathematical Association of America, Washington, DC, 2002.

F. Zhang, *Matrix Theory: Basic Results and Techniques*, Universitext,
DOI 10.1007/978-1-4614-1099-7, © Springer Science+Business Media, LLC 2011

D. Carlson, C. R. Johnson, D. C. Lay, A. D. Porter, A. Watkins, and W. Watkins, *Resources for Teaching Linear Algebra,* Mathematical Association of America, Washington, DC, 1997.

P. J. Davis, *Circulant Matrices,* John Wiley & Sons, Inc., New York, 1979.

W. F. Donoghue, Jr., *Monotone Matrix Functions and Analytic Continuation,* Springer-Verlag, New York, 1974.

T. Furuta, *Invitation to Linear Operators*, Taylor & Francis, New York, 2001.

F. R. Gantmacher, *The Theory of Matrices, Volume One,* Chelsea, New York, 1964.

S. K. Godunov, *Modern Aspects of Linear Algebra,* Translations of Mathematical Monographs, Vol. 175, Amer. Math. Soc., Providence, RI, 1998.

I. Gohberg and H. Langer, *Linear Operators and Matrices:* The Peter Lancaster Anniversary Volume (Operator Theory: Advances and Applications), Birkhäuser-Verlag, Berlin, 2002.

I. Gohberg, P. Lancaster, and L. Rodman, *Matrix Polynomials,* Academic Press, New York, 1982.

J. S. Golan, *The Linear Algebra a Beginning Graduate Student Ought to Know,* Kluwer Academic Publishers, Dordrecht, 2004.

G. H. Golub and C. F. Van Loan, *Matrix Computations,* 3rd edition, Johns Hopkins University Press, Baltimore, MD, 1996.

F. Graybill, *Matrices with Applications in Statistics,* Wadsworth, Belmont, CA, 1983.

P. R. Halmos, *Linear Algebra Problem Book,* Mathematical Association of America, Washington, DC , 1995.

D. A. Harville, *Matrix Algebra from a Statistician's Perspective,* Springer-Verlag, New York, 1997.

N. J. Higham, *Handbook of Writing for the Mathematical Sciences,* SIAM, Philadelphia, 1993.

N. J. Higham, *Functions of Matrices: Theory and Computation,* SIAM, Philadelphia, 2008.

K. J. Horadam, *Hadamard Matrices and Their Applications*, Princeton University Press, Princeton, NJ, 2007.

R. A. Horn and C. R. Johnson, *Matrix Analysis,* Cambridge University Press, New York, 1985.

R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis,* Cambridge University Press, New York, 1991.

N. Jacobson, *Basic Algebra I*, W. H. Freeman, San Francisco, 1974.

X.-Q. Jin and Y.-M. Wei, *Numerical Linear Algebra and Its Applications,* Science Press, Beijing, 2004.

E. Jorswieck and H. Boche, *Majorization and Matrix Monotone Functions in Wireless Communications* (Foundations and Trends in Communcations and Information Theory), NOW, Boston - Delft, 2007.

M. S. Klamkin, *Problems in Applied Mathematics,* SIAM, Philadelphia, 1990.

A. I. Kostrikin, *Exercises in Algebra,* Gordon and Breach, Amsterdam, 1996.

P. Lancaster and M. Tismenetsky, *The Theory of Matrices,* Second edition, Academic Press, Orlando, FL, 1985.

A. J. Laub, *Matrix Analysis for Scientists and Enginners*, SIAM, Philadelphia, 2005.

P. D. Lax, *Linear Algebra,* John Wiley & Sons, New York, 1997.

J.-S. Li and J.-G. Cha, *Linear Algebra (in Chinese),* University of Science and Technology of China Press, Hefei, 1989.

Z.-L. Li, *Collection of Math Graduate Entrance Examinations of Beijing Normal University 1978–2007,* Beijing Normal University Press, Beijing, 2007.

S.-Z. Liu, *Contributions to Matrix Calculus and Applications in Econometrics,* Tinbergen Institute Research Series, no. 106, Thesis Publishers, Amsterdam, 1995.

J. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics,* John Wiley & Sons, New York, 1988.

M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities,* Reprint edition, Dover, New York, 1992.

A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications,* Second edition, Springer, New York, 2011.

M. L. Mehta, *Matrix Theory,* Hindustan, New Delhi, 1989.

R. Merris, *Multilinear Algebra,* Gordon & Breach, Amsterdam, 1997.

H. Minc, *Permanents,* Addison-Wesley, New York, 1978.

L. Mirsky, *An Introduction to Linear Algebra,* Reprint edition, Dover, New York, 1990.

S. Montgomery and E. W. Ralston, *Selected Papers on Algebra, Volume Three,* Mathematical Association of America, Washington, DC, 1977.

G.-X. Ni, *Common Methods in Matrix Theory (in Chinese),* Shanghai Science and Technology Press, Shanghai, 1984.

B. N. Parlett, *The Symmetric Eigenvalue Problem* (Classics in Applied Mathematics), SIAM, Philadelphia, 1998.

R. Piziak and P. L. Odell, *Matrix Theory: From Generalized Inverses to Jordan Form*, Chapman & Hall/CRC, Boca Raton, FL, 2007.

V. V. Prasolov, *Problems and Theorems in Linear Algebra,* American Mathematical Society, Providence, RI, 1994.

C. R. Rao and M. B. Rao, *Matrix Algebra and Its Applications to Statistics and Econometrics,* World Scientific, Singapore, 1998.

W. Rudin, *Principles of Mathematical Analysis,* McGraw-Hill, New York, 1976.

J. R. Schott, *Matrix Analysis for Statistics,* John Wiley & Sons, New York, 1997.

D. Serre, *Matrices: Theory and Applications*, Springer, New York, 2002.

M.-R. Shi, *600 Linear Algebra Problems with Solutions (in Chinese),* Beijing Press of Science and Technology, Beijing, 1985.

B. Simon, *Trace Ideals and Their Applications,* American Mathematical Scociety, Providence, RI, 2005.

G. W. Stewart and J.-G. Sun, *Matrix Permutation Theory,* Academic Press, New York, 1990.

G. Strang, *Linear Algebra and Applications,* Academic Press, New York, 1976.

F. Uhlig and R. Grone, *Current Trends in Matrix Theory,* North-Holland, New York, 1986.

R. S. Varga, *Geršgorin and His Circles*, Springer Series in Computational Mathematics, Springer, Berlin, 2004.

B.-Y. Wang, *Foundations of Multilinear Algebra (in Chinese),* Beijing Normal University Press, Beijing, 1985.

B.-Y. Wang, *Introduction to Majorization Inequalities (in Chinese),* Beijing Normal University Press, Beijing, 1991.

S.-G. Wang and Z.-Z. Jia, *Matrix Inequalities (in Chinese),* Anhui Education Press, Hefei, 1994.

S.-G. Wang, M.-X. Wu and Z.-Z. Jia, *Matrix Inequalities (in Chinese),* 2nd edition, Science Press, Beijing, 2006.

S.-F. Xu, *Theory and Methods in Matrix Computation (in Chinese),* Beijing University Press, Beijing, 1995.

Y.-C. Xu, *Introduction to Algebra (in Chinese),* Shanghai Science and Technology Press, Shanghai, 1982.

H. Yanai, *Projection Matrices, Generalized Inverses, and Singular Value Decompositions,* University of Tokyo Press, Tokyo, Second printing, 1993.

X. Zhan, *Matrix Inequalities*, Lecture Notes in Mathematics 1790, Springer, New York, 2002.

X. Zhan, *Matrix Theory (in Chinese)*, Higher Education Press, Beijing, 2008.

F. Zhang (ed.), *The Schur Complement and Its Applications,* Springer, New York, 2005.

F. Zhang, *Linear Algebra: Challenging Problems for Students,* Second edition, Johns Hopkins University Press, Baltimore, MD, 2009.

Papers:

C. Akemann, J. Anderson, and G. Pedersen, *Triangle inequalities in operator algebras,* Linear Multilinear Algebra, Vol. 11, pp. 167–178, 1982.

G. Alpargu and G. P. H. Styan, *Some comments and a future bibliography on the Frucht-Kantorovich and Wielandt inequalities, and on some related inequalities,* Manuscript, 1998.

T. W. Anderson and G. P. H. Styan, *Cochran's theorem, rank additivity and tripotent matrices,* Statistics and Probability: Essays in Honor of C. R. Rao, Edited by G. Kallianpur, P. R. Krishnaiah, and J. K. Ghosh, North-Holland, Amsterdam, pp. 1–23, 1982.

T. Ando, *Concavity of certain maps on positive definite matrices and applications to Hadamard products,* Linear Algebra Appl., Vol. 26, pp. 203–241, 1979.

T. Ando, *Hölder type inequalities for matrices,* Math. Inequal. Appl., Vol. 1, no. 1, pp. 1–30, 1998.

T. Ando and F. Hiai, *Log majorization and complementary Golden-Thompson type inequalities,* Linear Algebra Appl., Vol. 197-198, pp. 113-131, 1994.

T. Ando, R. A. Horn, and C. R. Johnson, *The singular values of a Hadamard product: A basic inequality,* Linear Multilinear Algebra, Vol. 21, pp. 345–365, 1987.

T. Ando, C.-K. Li, and R. Mathias, *Geometric means,* Linear Algebra Appl., Vol. 385, pp. 305–334, 2004.

K. M. R. Audenaert, *A Lieb–Thirring inequality for singular values,* Linear Algebra Appl., Vol. 430, pp. 3053–3057, 2009.

K. M. R. Audenaert, *Spectral radius of Hadamard product versus conventional product for nonnegative matrices,* Linear Algebra Appl., Vol. 432, pp. 366–368, 2010.

J.-S. Aujla and F. C. Silva, *Weak majorization inequalities and convex functions*, Linear Algebra Appl., Vol. 369, pp. 217–233, 2003.

Y.-H. Au-Yeung and Y.-T. Poon, 3×3 *orthostochastic matrices and the convexity of generalized numerical ranges,* Linear Algebra Appl., Vol. 27, pp. 69–79, 1979.

C. S. Ballantine, *A note on the matrix equation* $H = AP + PA^*$, Linear Algebra Appl., Vol. 2, pp. 37–47, 1969.

R. B. Bapat and M. K. Kwong, *A generalization of* $A \circ A^{\mathsf{T}} \geq I$, Linear Algebra Appl., Vol. 93, pp. 107–112, 1987.

R. B. Bapat and V. S. Sunder, *On majorization and Schur products,* Linear Algebra Appl., Vol. 72, pp. 107–117, 1985.

S. J. Bernau and G. G. Gregory, *A Cauchy–Schwarz inequality for determinants,* Amer. Math. Monthly, Vol. 84, pp. 495–496, June–July 1977.

R. Bhatia and C. Davis, *A better bound on the variance,* Amer. Math. Monthly, Vol. 107, pp. 353–356, 2000.

R. Bhatia, C. Davis, and A. McIntosh, *Perturbation of spectral subspaces and solution of linear operator equations,* Linear Algebra Appl., Vol. 52-53, pp. 45-67, 1983.

R. Bhatia and F. Kittaneh, *On the Singular Values of a Product of Operators*, SIAM J. Matrix Anal. Appl., Vol. 11, No. 2, pp. 272–277, 1990.

R. Bhatia, R. Horn, and F. Kittaneh, *Normal approximations to binormal operators,* Linear Algebra Appl., Vol. 147, pp. 169–179, 1991.

J.-V. Bondar, *Schur majorization inequalities for symmetrized sums with applications to tensor products,* Linear Algebra Appl., Vol. 360, pp. 1–13, 2003.

J.-C. Bourin, *Matrix versions of some classical inequalities*, Linear Algebra Appl., Vol. 416, pp. 890–907, 2006.

R. Brualdi and H. Schneider, *Determinantal identities: Gauss, Schur, Cauchy, Sylvester, Kronecker, Jacobi, Binet, Laplace, Muir, and Cayley,* Linear Algebra Appl., Vol. 52/53, pp. 769–791, 1983.

M. Buliga, *Four applications of majorization to convexity in the calculus of variations,* Linear Algebra Appl., Vol 429, pp. 1528–1545, 2008.

B.-E. Cain, R.-A. Horn and L.-L. Li, *Inequalities for monotonic arrangements of eigenvalues,* Linear Algebra Appl., Vol. 222, pp. 1–8, 1995.

E. Carlen and E. Lieb, *Short proofs of theorems of Mirsky and Horn on diagonal and eigenvalues of matrices,* Electron. J. Linear Algebra, Vol. 18, pp. 438–441, 2009.

D. Carlson, C. R. Johnson, D. C. Lay, and A. D. Porter, *Gems of exposition in elementary linear algebra,* College Math. J., Vol. 23, no. 4, pp. 299–303, 1992.

N. N. Chan and M. K. Kwong, *Hermitian matrix inequalities and a conjecture,* Amer. Math. Monthly, Vol. 92, pp. 533–541, October 1985.

R. Chapman, *A polynomial taking integer values,* Math. Mag., Vol. 69, no. 2, p. 121, April 1996.

C.-M. Cheng, R. A. Horn, C.-K. Li, *Inequalities and equalities for the Cartesian decomposition of complex matrices,* Linear Algebra Appl., Vol. 341, pp. 219–237, 2002.

C.-M. Cheng, I.-C. Law, and S.-I. Leong, *Eigenvalue equalities for ordinary and Hadamard products of powers of positive semidefinite matrices,* Linear Algebra Appl., Vol. 422, pp. 771–787, 2007.

M.-D. Choi, *A Schwarz inequality for positive linear maps on C^*-algebras,* Illinois J. Math., Vol. 18, pp. 565–574, 1974.

J. Chollet, *On principal submatrices,* Linear Multilinear Algebra, Vol. 11, pp. 283–285, 1982.

J. Chollet, *Some Inequalities for principal submatrices,* Amer. Math. Monthly, Vol. 104, pp. 609–617, August–September 1997.

G. Dahl, *Matrix majorization,* Linear Algebra Appl., Vol. 288, pp. 53–73, 1999.

C. Davis, *Notions generalizing convexity for functions defined on spaces of matrices,* Proc. Sympos. Pure Math., Amer. Math. Soc., Vol. 7, pp. 187–201, 1963.

D. Z. Djokovic, *On some representations of matrices,* Linear Multilinear Algebra, Vol. 4, pp. 33–40, 1979.

S. W. Drury, *A bound for the determinant of certain Hadamard products and for the determinant of the sum of two normal matrices,* Linear Algebra Appl., Vol. 199, pp. 329–338, 1994.

S. W. Drury, *Operator norms of words formed from positive-definite matrices,* Electron. J. Linear Algebra, Vol. 18, pp. 13–20, 2009.

C. F. Dunkl and K. S. Williams, *A simple norm inequality,* Amer. Math. Monthly, Vol. 71, pp. 53–54, January 1964.

L. Elsner, *On the variation of the spectra of matrices,* Linear Algebra Appl., Vol. 47, pp. 127–138, 1982.

L. Elsner, *The generalized spectral-radius theorem: An analytic-geometric proof,* Linear Algebra Appl., Vol. 220, pp. 151–159, 1995.

L. Elsner and Kh. D. Ikramov, *Normal matrices: An update*, Linear Algebra Appl., Vol. 285, pp. 291–303, 1998.

K. Fan, *On a theorem of Weyl concerning eigenvalues of linear transformations I,* Proc. Natl. Acad. Sci. USA, Vol. 35, pp. 652–655, 1949.

M. Fiedler, *Bounds for the determinant of sum of Hermitian matrices,* Proc. Amer. Math. Soc., Vol. 30, No. 1, pp. 27–31, 1971.

M. Fiedler, *A note on the Hadamard product of matrices,* Linear Algebra Appl., Vol. 49, pp. 233–235, 1983.

S. Fisk, *A note on Weyl's inequality,* Amer. Math. Monthly, Vol. 104, pp. 257–258, March 1997.

T. Furuta, *$A \geq B \geq 0$ assures $(B^r A^p B^r)^{1/q} \geq B^{(p+2r)/q}$ for $r \geq 0, p \geq 0, q \geq 1$ with $(1+2r)q \geq p+2r$,* Proc. Amer. Math. Soc., Vol. 101, pp. 85–88, September 1987.

T. Furuta and M. Yanagida, *Generalized means and convexity of inversion for positive operators,* Amer. Math. Monthly, Vol. 105, No. 3, pp. 258–259, 1998.

A. Galántai and Cs. J. Hegedűs, *Jordan's principal angles in complex vector spaces,* Numer. Linear Algebra Appl., Vol. 13, pp. 589–598, 2006.

G. R. Goodson, *The inverse-similarity problem for real orthogonal matrices,* Amer. Math. Monthly, Vol. 104, pp. 223–230, March 1997.

W. Govaerts and J. D. Pryce, *A singular value inequality for block matrices,* Linear Algebra Appl., Vol. 125, pp. 141–148, 1989.

R. Grone, C. R. Johnson, E. M. Sa, and H. Wolkowicz, *Normal matrices,* Linear Algebra Appl., Vol. 87, pp. 213–225, 1987.

R. Grone, S. Pierce, and W. Watkins., *Extremal correlation matrices,* Linear Algebra Appl., Vol. 134, pp. 63–70, 1990.

J. Groß, G. Trenkler, and S. Troschke, *On a characterization associated with the matrix arithmetic and geometric means,* IMAGE: Bull. Int. Linear Algebra Soc., no. 17, p. 32, Summer 1996.

P. R. Halmos, *Bad products of good matrices,* Linear Multilinear Algebra, Vol. 29, pp. 1–20, 1991.

F. Hansen, *An operator inequality*, Math. Ann., Vol. 246, pp. 249–250, 1980.

F. Hansen and G. K. Pedersen, *Jensen's inequalities for operators and Löwner's theorem*, Math. Ann., Vol. 258, pp. 229–241, 1982.

E. V. Haynsworth, *Applications of an inequality for the Schur complement,* Proc. Amer. Math. Soc., Vol. 24, pp. 512–516, 1970.

G. Herman, *Solution to IMAGE Problem 45-4*, IMAGE: Bull. Int. Linear Algebra Soc., Issue 46, Spring, 2011.

G. Herman, *Solution to Monthly Problem 11488*, Preprint, Feb., 2011.

Y. P. Hong and R. A. Horn, *A canonical form for matrices under consimilarity,* Linear Algebra Appl., Vol. 102, pp. 143–168, 1988.

R. A. Horn, *The Hadamard product,* Proc. Sympos. Appl. Math., Vol. 40, Edited by C. R. Johnson, Amer. Math. Soc., Providence, RI, pp. 87–169, 1990.

R. A. Horn and C. R. Johnson, *Hadamard and conventional submultiplicativity for unitarily invariant norms on matrices,* Linear Multilinear Algebra, Vol. 20, pp. 91–106, 1987.

R. A. Horn and I. Olkin, *When does $A^*A = B^*B$ and why does one want to know?* Amer. Math. Monthly, Vol. 103, pp. 270–482, June–July 1996.

R. A. Horn and F. Zhang, *Bounds on the spectral radius of a Hadamard product of nonnegative or positive semidefinite matrices,* Electron. J. Linear Algebra, Vol. 20, pp. 90–94, 2010.

L.-K. Hua, *Inequalities involving determinants (in Chinese),* Acta Math Sinica, Vol. 5, no. 4, pp. 463–470, 1955.

Z. Huang, *On the spectral radius and the spectral norm of Hadamard products of nonnegative matrices*, Linear Algebra Appl., Vol. 434, pp. 457–462, 2011.

Y. Ikebe, T. Inagaki, and S. Miyamoto, *The monotonicity theorem, Cauchy's interlace theorem, and the Courant–Fischer theorem,* Amer. Math. Monthly, Vol. 94, pp. 352–354, April 1987.

E. I. Im, *Narrower eigenbounds for Hadamard products,* Linear Algebra Appl., Vol. 264, pp. 141–144, 1997.

E.-X. Jiang, *Bounds for the smallest singular value of a Jordan block with an application to eigenvalue permutation,* Linear Algebra Appl., Vol. 197/198, pp. 691–707, 1994.

C. R. Johnson, *An inequality for matrices whose symmetric part is positive definite,* Linear Algebra Appl., Vol. 6, pp. 13–18, 1973.

C. R. Johnson, *Inverse M-matrices,* Linear Algebra Appl., Vol. 47, pp. 195–216, 1982.

C. R. Johnson, *The relationship between AB and BA,* Amer. Math. Monthly, Vol. 103, pp. 578–582, August–September 1996.

C. R. Johnson and F. Zhang, *An operator inequality and matrix normality,* Linear Algebra Appl., Vol. 240, pp. 105–110, 1996.

K. R. Laberteaux, M. Marcus, G. P. Shannon, E. A. Herman, R. B. Israel, and G. Letac, *Hermitian matrices,* Problem Solutions, Amer. Math. Monthly, Vol. 104, p. 277, March 1997.

T.-G. Lei, C.-W. Woo, and F. Zhang, *Matrix inequalities by means of embedding,* Electron. J. Linear Algebra, Vol 11, pp. 66–77, 2004.

C.-K. Li, *Matrices with some extremal properties,* Linear Algebra Appl., Vol. 101, pp. 255–267, 1988.

R.-C. Li, *Norms of certain matrices with applications to variations of the spectra of matrices and matrix pencils,* Linear Algebra Appl., Vol. 182, pp. 199–234, 1993.

Z.-S. Li, F. Hall, and F. Zhang, *Sign patterns of nonnegative normal matrices,* Linear Algebra Appl., Vol. 254, pp. 335–354, 1997.

R. A. Lippert and G. Strang, *The Jordan forms of AB and BA,* Electron. J. Linear Algebra, Vol. 18, pp. 281–288, 2009.

S.-Z. Liu and H. Neudecker, *Several matrix Kantorovich-type inequalities,* J. Math. Anal. Appl., Vol. 197, pp. 23–26, 1996.

D. London, *A determinantal inequality and its permanental counterpart,* Linear Multilinear Algebra, Vol. 42, pp. 281–290, 1997.

M. Lundquist and W. Barrett, *Rank inequalities for positive semidefinite matrices,* Linear Algebra Appl., Vol. 248, pp. 91–100, 1996.

M. Marcus, K. Kidman, and M. Sandy, *Products of elementary doubly stochastic matrices,* Linear Multilinear Algebra, Vol. 15, pp. 331–340, 1984.

T. Markham, *An Application of Theorems of Schur and Albert,* Proc. Amer. Math. Soc., Vol. 59, No. 2, pp. 205–210, September 1976.

T. Markham, *Oppenheim's inequality for positive definite matrices,* Amer. Math. Monthly, Vol. 93, pp. 642–644, October 1986.

G. Marsaglia and G. P. H. Styan, *Equalities and inequalities for ranks of matrices,* Linear Multilinear Algebra, Vol. 2, pp. 269–292, 1974.

A. W. Marshall and I. Olkin, *Reversal of the Lyapunov, Hölder, and Minkowski inequalities and other extensions of the Kantorovich inequality,* J. Math. Anal. Appl., Vol. 8, pp. 503–514, 1964.

A. W. Marshall and I. Olkin, *Matrix versions of the Cauchy and Kantorovich inequalities,* Aequationes Math., Vol. 40, pp. 89–93, 1990.

R. Mathias, *Concavity of monotone matrix functions of finite order,* Linear Multilinear Algebra, Vol. 27, pp. 129–138, 1990.

R. Mathias, *An arithmetic-geometric-harmonic mean inequality involving Hadamard products,* Linear Algebra Appl., Vol. 184, pp. 71–78, 1993.

C. McCarthy and A. T. Benjamin, *Determinants of tournaments,* Math. Mag., Vol. 69, no. 2, pp. 133–135, April 1996.

J. K. Merikoski, *On the trace and the sum of elements of a matrix,* Linear Algebra Appl., Vol. 60, pp. 177–185, 1984.

D. Merino, *Matrix similarity,* IMAGE: Bull. Int. Linear Algebra Soc., no. 21, p. 26, October 1998.

R. Merris, *The permanental dominance conjecture,* in Current Trends in Matrix Theory, Edited by F. Uhlig and R. Grone, Elsevier, New York, pp. 213–223, 1987.

L. Mirsky, *A note on normal matrices,* Amer. Math. Monthly, Vol. 63, p. 479, 1956.

B. Mond and J. E. Pečarić, *Matrix versions of some means inequalities,* Austral. Math. Soc. Gaz., Vol. 20, pp. 117–120, 1993.

B. Mond and J. E. Pečarić, *Inequalities for the Hadamard product of matrices,* SIAM J. Matrix Anal. Appl., Vol. 19, No. 1, pp. 66–70, 1998.

M. Newman, *On a problem of H.J. Ryser,* Linear Multilinear Algebra, Vol. 12, pp. 291–293, 1983.

P. Nylen, T.-Y. Tam, and F. Uhlig, *On the eigenvalues of principal submatrices of normal, Hermitian and symmetric matrices,* Linear Multilinear Algebra, Vol. 36, pp. 69–78, 1993.

K. Okubo, *Hölder-type norm inequalities for Schur products of matrices,* Linear Algebra Appl., Vol. 91, pp. 13–28, 1987.

I. Olkin, *Symmetrized product definiteness?,* IMAGE: Bull. Int. Linear Algebra Soc., no. 19, p. 32, Summer 1997.

D. V. Ouellette, *Schur complements and statistics,* Linear Algebra Appl., Vol. 36, pp. 187–295, 1981.

C. C. Paige, G. P. H. Styan, B.-Y. Wang, and F. Zhang, *Hua's matrix equality and Schur complements,* Internat. J. Inform. Syst. Sci., Vol. 4, no.1, pp. 124–135, 2008.

W. V. Parker, *Sets of complex numbers associated with a matrix,* Duke Math. J., Vol. 15, pp. 711–715, 1948.

R. Patel and M. Toda, *Trace inequalities involving Hermitian matrices,* Linear Algebra Appl., Vol. 23, pp. 13–20, 1979.

J. E. Pečarić, S. Puntanen, and G. P. H. Styan, *Some further matrix extensions of the Cauchy–Schwarz and Kantorovich inequalities, with some statistical applications,* Linear Algebra Appl., Vol. 237/238, pp. 455–476, 1996.

V. Pták, *The Kantorovich inequality,* Amer. Math. Monthly, Vol. 102, pp. 820–821, November 1995.

S. M. Robinson, *A short proof of Cramer's rule,* Math. Mag., Vol. 43, pp. 94–95, 1970.

L. Rodman, *Products of symmetric and skew-symmetric matrices,* Linear Multilinear Algebra, Vol. 43, pp. 19–34, 1997.

B. L. Shader and C. L. Shader, *Scheduling conflict-free parties for a dating service,* Amer. Math. Monthly, Vol. 104, pp. 99–106, February 1997.

C. Shafroth, *A generalization of the formula for computing the inverse of a matrix,* Amer. Math. Monthly, Vol. 88, pp. 614–616, October 1981.

W. So, *Equality cases in matrix expoential inequalities,* SIAM J. Matrix Anal. Appl., Vol. 13, No. 4, pp. 1154–1158, 1992.

W. So and R. C. Thompson, *Products of exponentials of Hermitian and complex symmetric matrices,* Linear Multilinear Algebra, Vol. 29, pp. 225–233, 1991.

G. W. Soules, *An approach to the permanental-dominance conjecture,* Linear Algebra Appl., Vol. 201, pp. 211–229, 1994.

J. Stewart, *Positive definite functions and generalizations: A historical survey,* Rocky Mountain J. Math., Vol. 6, pp. 409–434, 1976.

R. Stong, *Solution to Monthly Problem 11488*, Amer. Math. Monthly, to appear.

G. P. H. Styan, *Hadamard products and multivariate statistical analysis,* Linear Algebra Appl., Vol. 6, pp. 217–240, 1973.

G. P. H. Styan, *Schur complements and linear statistical models,* Proc. 1st Int. Tampere Seminar on Linear Stat. Models and Their Appl., Edited by T. Pukkila and S. Puntanen, Department of Mathematical Sciences, University of Tampere, Tampere, Finland, pp. 37–75, 1985.

J.-G. Sun, *On two functions of a matrix with positive definite Hermitian part,* Linear Algebra Appl., Vol. 244, pp. 55–68, 1996.

R. C. Thompson, *Convex and concave functions of singular values of matrix sums,* Pacific J. Math., Vol. 66, pp. 285–290, 1976.

R. C. Thompson, *Matrix type metric inequalities,* Linear Multilinear Algebra, Vol. 5, pp. 303–319, 1978.

R. C. Thompson, *High, low, and quantitative roads in linear algebra,* Linear Algebra Appl., Vol. 162–164, pp. 23–64, 1992.

L. Tie, K.-Y. Cai, and Y. Lin, *Rearrangement inequalities for Hermitian matrices,* Linear Algebra Appl., Vol. 434, pp. 443–456, 2011.

G. Visick, *Majorizations of Hadamard products of matrix powers,* Linear Algebra Appl., Vol. 269, pp. 233–240, 1998.

B.-Y. Wang and M.-P. Gong, *Some eigenvalue inequalities for positive semidefinite matrix power products,* Linear Algebra Appl., Vol. 184, pp. 249–260, 1993.

B.-Y. Wang and F. Zhang, *A trace inequality for unitary matrices,* Amer. Math. Monthly, Vol. 101, no. 5, pp. 453–455, May 1994.

B.-Y. Wang and F. Zhang, *Words and matrix normality,* Linear Multilinear Algebra, Vol. 34, pp. 93–89, 1995.

B.-Y. Wang and F. Zhang, *Trace and eigenvalue inequalities for ordinary and Hadamard products of positive semidefinite Hermitian matrices,* SIAM J. Matrix Anal. Appl., Vol. 16, pp. 1173–1183, October 1995.

B.-Y. Wang and F. Zhang, *Schur complements and matrix inequalities of Hadamard products,* Linear Multilinear Algebra, Vol. 43, pp. 315–326, 1997.

B.-Y. Wang and F. Zhang, *On zero-one symmetric and normal matrices,* J. Math. Res. Exposition, Vol. 18, No.2, pp. 159–164, May 1998.

W. Watkins, *A determinantal inequality for correlation matrices,* Linear Algebra Appl., Vol. 79, pp. 209–213, 1988.

G. S. Watson, G. Alpargu, and G. P. H. Styan, *Some comments on six inequalities associated with the inefficiency of ordinary least squares with one regressor,* Linear Algebra Appl., Vol. 264, pp. 13–53, 1997.

C.-S. Wong, *Characterizations of products of symmetric matrices,* Linear Algebra Appl., Vol. 42, pp. 243–251, 1982.

P.-Y. Wu, *Products of positive semidefinite matrices,* Linear Algebra Appl., Vol. 111, pp. 53–61, 1988.

C.-Q. Xu, Z.-D. Xu, and F. Zhang, *Revisiting Hua-Marcus-Bellman-Ando Inequalities on Contractive Matrices,* Linear Algebra Appl., Vol. 430, pp. 1499–1508, 2009.

G.-H. Xu, C.-Q. Xu, and F. Zhang, *Contractive matrices of Hua type,* Linear Multilinear Algebra, Vol. 59, No.2, pp. 159–172, 2011.

T. Yamazaki, *Parallelisms between Aluthge transformation and powers of operators,* Acta Sci. Math., Vol. 67, pp. 809–820, 2001.

R. Yarlagadda and J. Hershey, *A note on the eigenvectors of Hadamard matrices of order* 2^n, Linear Algebra Appl., Vol. 45, pp. 43–53, 1982.

X.-Z. Zhan, *Inequalities for the singular values of Hadamard products,* SIAM J. Matrix Anal. Appl., Vol. 18, pp. 1093–1095, October 1997.

X.-Z. Zhan, *Inequalities involving Hadamard products and unitarily invariant norms,* Adv. in Math. (China), 27(1998), no. 5, 416-422.

X.-Z. Zhan, *OThe sharp Rado theorem for majorizations,* Amer. Math. Monthly, Vol. 110, no. 2, pp. 152-153, 2003.

X.-Z. Zhan, *On some matrix inequalities,* Linear Algebra Appl., Vol. 376, pp. 299-303, 2004.

F. Zhang, *Quaternions and matrices of quaternions,* Linear Algebra Appl., Vol. 251, pp. 21–57, 1997.

F. Zhang, *On the operator Bohr inequality*, J. Math. Anal. Appl., Vol. 333, pp. 1264–1271, 2007.

F. Zhang, *Positivity of matrices with generalized matrix functions,* Preprint, 2011.

F. Zhang and Q.-L. Zhang, *Eigenvalue inequalities for matrix product,* IEEE Trans. Automat. Contr., Vol. 51, No. 9, pp. 1506–1509, August 2006.

Notation

$\mathbb{M}_n$, 8	$n \times n$ (i.e., n-square) complex matrices
$\mathbb{M}_{m\times n}$, 8	$m \times n$ complex matrices
$\mathbb{C}$, 1	complex numbers
$\mathbb{R}$, 1	real numbers
$\mathbb{F}$, 1	a field of numbers, i.e., $\mathbb{C}$ or $\mathbb{R}$ in this book
$\mathbb{Q}$, 6	rational numbers
$\mathbb{C}^n$, 3	(column) vectors with n complex components
$\mathbb{R}^n$, 2	(column) vectors with n real components
$\mathbb{R}^n_+$, 331	(column) vectors with n nonnegative components
$\mathbb{F}[x]$, 5	polynomials over field $\mathbb{F}$
$\mathbb{F}_n[x]$, 5	polynomials over field $\mathbb{F}$ with degree at most n
$C[a,b]$, 6	real-valued continuous functions on interval $[a,b]$
$C'(\mathbb{R})$, 6	real-valued functions with continuous derivatives on $\mathbb{R}$
$\operatorname{Re} c$, 129, 195	real part of complex number c
$\operatorname{Im} c$, 294	imaginary part of complex number c
ω, 139	nth primitive root of unity
t^+, 333	$t^+ = t$ if $t \geq 0$; $t^+ = 0$ if $t < 0$
δ_{ij}, 266	Kronecker delta, i.e., $\delta_{ij} = 1$ if $i = j$, and 0 otherwise
$V \cap W$, 4	intersettion of sets V and W
$V \cup W$, 26, 68	union of sets V and W
$P \Rightarrow Q$, 4	statement P implies statement Q
$P \Leftrightarrow Q$, 32	statements P and Q are equivalent
$\operatorname{Span} S$, 3	vector space spanned by the vectors in S
$\dim V$, 3	dimension of the vector space V
$V + W$, 4	sum of subspaces V and W
$V \oplus W$, 4	direct sum of subspaces V and W
$\mathcal{D}_x$, 17	differential operator
S_n, 12	nth symmetric group, i.e., all permutations on $\{1, 2, \ldots, n\}$
e_i, 3	vector with ith component 1 and 0 elsewhere
(u, v), 27	inner product of vectors u and v, i.e, v^*u
$\angle_{x,y}$, 30	angle between real vectors x, y, i.e., $\angle_{x,y} = \cos^{-1} \frac{(x,y)}{\lVert x\rVert\,\lVert y\rVert}$
$<_{x,y}$, 33, 198	angle between complex vectors x, y, i.e., $<_{x,y} = \cos^{-1} \frac{\lvert(x,y)\rvert}{\lVert x\rVert\,\lVert y\rVert}$
$d(x, y)$, 182	distance between x and y in a metric space
$\lvert x\rvert$, 327	absolute value vector $\lvert x\rvert = (\lvert x_1\rvert, \ldots, \lvert x_n\rvert)$
$\lVert x\rVert$, 28	length or norm of vector x
$\lVert x\rVert_p$, 373	l_p-norm of vector x, i.e., $\lVert x\rVert_p = \Big(\sum_{i=1}^n \lvert x_i\rvert^p\Big)^{1/p}$
$\lVert x\rVert_{(k)}$, 373	Ky Fan k-norm of vector x, i.e., $\lVert x\rVert_{(k)} = \max_{i_1<\cdots<i_k} \sum_{t=1}^k \lvert x_{i_t}\rvert$
x^T, 2	transpose of x; it is a column vector if x is a row vector
$x^\perp$, 28	vectors orthogonal to vector x

$S^{\perp}$, 28	vector space orthogonal to set S
$S_1 \perp S_2$, 28	$(x, y) = 0$ for all $x \in S_1$ and $y \in S_2$
V_λ, 23	eigenspace of the eigenvalue λ
I_n, I, 9	identity matrix of order n
$A = (a_{ij})$, 8	matrix with entries a_{ij}
A^T, 9	transpose of matrix A
$\overline{A}$, 9	conjugate of matrix A
A^*, 9	conjugate transpose of matrix A
A^{-1}, 13	inverse of matrix A
$A^{\dagger}$, 377	Moore–Penrose inverse of matrix A
A_{11}, 41, 217	principal submatrix of matrix A in the upper-left corner
$A(i\|j)$, 13	matrix by deleting the ith row and jth column of matrix A
$\operatorname{adj}(A)$, 13	adjoint matrix of matrix A
$\det A$, 12	determinant of matrix A
$\operatorname{rank}(A)$, 11	rank of matrix A
$\operatorname{tr} A$, 21	trace of matrix A
$\operatorname{diag} S$, 70	diagonal matrix with the elements of S on the diagonal
$\begin{vmatrix} A & B \\ C & D \end{vmatrix}$, 11	determinant of the 2×2 block matrix
$(A, B)_{\mathbb{M}}$, 30	matrix inner product, i.e., $(A, B)_{\mathbb{M}} = \operatorname{tr}(B^* A)$
$\operatorname{Im} A$, 17, 51	image of matrix or linear transformation A, i.e., $\operatorname{Im} A = \{Ax\}$
$\operatorname{Ker} A$, 17, 51	kernel or null space of A, i.e., $\operatorname{Ker} A = \{x : Ax = 0\}$
$\mathcal{R}(A)$, 55	row space spanned by the row vectors of matrix A
$\mathcal{C}(A)$, 55	column space spanned by the column vectors of matrix A
$R(A$, 306	row sum vector of matrix A
$C(A)$, 306	column sum vector of matrix A
$\mathcal{H}(A)$, 233	Hermitian part of matrix A, i.e., $\frac{1}{2}(A + A^*)$
$\mathcal{S}(A)$, 361	skew-Hermitian part of matrix A, i.e., $\frac{1}{2}(A - A^*)$
$W(A)$, 107	numerical range of matrix A
J_n, 152	n-square matrix with all entries equal to 1
T_n, 133	n-square tridiagonal matrix
H_n, 150	Hadamard matrix
$V_n(a_i)$, 143	n-square Vandermonde matrix of $a_1, \ldots, a_n$
$G(x_i)$, 225	Gram matrix of $x_1, \ldots, x_n$
$s_k(a_i)$, 124	kth elementary symmetric function of $a_1, \ldots, a_n$
$w(A)$, 109	numerical radius of matrix A
$\rho(A)$, 109	spectral radius of matrix A
$i_+(A)$, 255	number of positive eigenvalues of Hermitian matrix A
$i_-(A)$, 255	number of negative eigenvalues of Hermitian matrix A
$i_0(A)$, 255	number of zero eigenvalues of Hermitian matrix A
$\operatorname{In}(A)$, 256	inertia of Hermitian matrix A, i.e., $\operatorname{In}(A) = (i_+(A), i_-(A), i_0(A))$
$\lambda_{\max}(A)$, 124	largest eigenvalue of matrix A
$\sigma_{\max}(A)$, 109	largest singular value of matrix A, i.e., the spectral norm of A
$\sigma_1(A)$, 266	largest singular value of matrix A; the same as $\sigma_{\max}(A)$
$\lambda_{\min}(A)$, 266	smallest eigenvalue of matrix A
$\sigma_{\min}(A)$, 266	smallest singular value of matrix A

$\lambda_i(A)$, 21, 82	eigenvalue of matrix A
$\sigma_i(A)$, 61, 82	singular value of matrix A
$\lambda(A)$, 349	eigenvalue vector of $A \in \mathbb{M}_n$, i.e., $\lambda(A) = (\lambda_1(A), \dots, \lambda_n(A))$
$\sigma(A)$, 349	singular value vector of $A \in \mathbb{M}_{m\times n}$, i.e., $\sigma(A) = (\sigma_1(A), \dots, \sigma_n(A))$
$\lambda^\alpha(A)$, 365	$\lambda^\alpha(A) = (\lambda_1^\alpha(A), \dots, \lambda_n^\alpha(A)) = ((\lambda_1(A))^\alpha, \dots, (\lambda_n(A))^\alpha)$
$\sigma^\alpha(A)$, 365	$\sigma^\alpha(A) = (\sigma_1^\alpha(A), \dots, \sigma_n^\alpha(A)) = ((\sigma_1(A))^\alpha, \dots, (\sigma_n(A))^\alpha)$
$p(\lambda)\|q(\lambda)$, 94	$p(\lambda)$ divides $q(\lambda)$
$d(\lambda)$, 94	invariant factors of λ-matrix $\lambda I - A$
$d(A)$, 349	vector of diagonal entries of a square matrix A
$m_A(\lambda)$, 88	minimal polynomial of matrix A
$p_A(\lambda)$, 21, 87	characteristic polynomial of matrix A, i.e., $p_A(\lambda) = \det(\lambda I - A)$
$A \geq 0$, 81	A is positive semidefinite (or a nonnegative matrix in Section 5.7)
$A > 0$, 81	A is positive definite (or a positive matrix in Section 5.7)
$A \geq B$, 81	$A - B$ is positive semidefinite (or $a_{ij} \geq b_{ij}$ in Section 5.7)
$A^{1/2}$, 81, 203	square root of positive semidefinite matrix A
A^α, 81	$A^\alpha = U^* \operatorname{diag}(\lambda_1^\alpha, \dots, \lambda_n^\alpha)U$ if $A = U^* \operatorname{diag}(\lambda_1, \dots, \lambda_n)U$
e^A, 66	$\sum_{k=0}^\infty \frac{1}{k!}A^k$
$[A]$, 219	principal submatrix of A
$A[\alpha\|\beta]$, 122	submatrix of A indexed by α and β
$\|A\|$, 83, 287	$\|A\| = (A^*A)^{1/2}$ (or $(\|a_{ij}\|)$ in Section 5.7)
$\widetilde{A_{11}}$, 217	Schur complement of A_{11}
$A^{(k)}$, 122	kth compound matrix of matrix A
$\\|A\\|$, 113	norm of matrix A
$\\|A\\|_{\mathrm{op}}$, 113	operator norm of matrix A, i.e., $\\|A\\|_{\mathrm{op}} = \sup_{\\|x\\|=1} \\|Ax\\|$
$\\|A\\|_F$, 115	Frobenious norm of matrix A, i.e., $\\|A\\|_F = \left(\sum_{i=1}^n \sigma_i^2(A)\right)^{1/2}$
$\\|A\\|_{(k)}$, 115	Ky Fan k-norm of matrix A, i.e., $\\|A\\|_{(k)} = \sum_{i=1}^k \sigma_i(A)$
$\\|A\\|_p$, 115	Schatten p-norm of matrix A, i.e., $\\|A\\|_p = \left(\sum_{i=1}^n \sigma_i^p(A)\right)^{1/p}$
$\\|A\\|_2$, 115	$\\|A\\|_2 = \\|A\\|_F = \left(\sum_{i,j} \|a_{ij}\|^2\right)^{1/2} = \left(\sum_{i=1}^n \sigma_i^2(A)\right)^{1/2}$
$[A, B]$, 305	commutator of A and B, i.e., $[A, B] = AB - BA$
$A \oplus B$, 11	direct sum of matrices A and B, i.e., $A \oplus B = \left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right)$
$A \otimes B$, 117	Kronecker product of matrices A and B
$A \circ B$, 117	Hadamard product of matrices A and B
$x \circ y$, 117, 327	$x \circ y = (x_1y_1, \dots, x_ny_n)$
x^m, 348	$x^m = (x_1^m, \dots, x_n^m)$ if $x = (x_1, \dots, x_n)$
$x^\downarrow$, 325	$x^\downarrow = (x_1^\downarrow, x_2^\downarrow, \dots, x_n^\downarrow)$, where $x_1^\downarrow \geq x_2^\downarrow \geq \cdots \geq x_n^\downarrow$.
$x^\uparrow$, 325	$x^\uparrow = (x_1^\uparrow, x_2^\uparrow, \dots, x_n^\downarrow)$, where $x_1^\uparrow \leq x_2^\uparrow \leq \cdots \leq x_n^\uparrow$.
$x \prec_w y$, 326	x is weakly majorized by y, i.e., $\sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k y_i^\downarrow$, $k \leq n$
$x \prec y$, 326	x is majorized by y, i.e., $x \prec_w y$ and $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$
$x \prec_{\mathrm{wlog}} y$, 344	x is weakly log-majorized by y, i.e., $\prod_{i=1}^k x_i^\downarrow \leq \prod_{i=1}^k y_i^\downarrow$, $k \leq n$
$x \prec_{\log} y$, 344	x is log-majorized by y, i.e., $x \prec_{\mathrm{wlog}} y$ and $\prod_{i=1}^n x_i = \prod_{i=1}^n y_i$

Index

T-transform, 335
T-transformation, 335
$*$-congruency, 256
λ-matrix, 93
λ-matrix standard form, 94
l_p-norm, 373, 376

addition, 1
adjoint, 13
algebraic multiplicity, 24
angle, 30
arithmetic mean–geometric mean
 inequality, 86, 346

backward identity, 78, 100
backward identity matrix, 91
basis, 3
Bhatia–Kittaneh theorem on singular
 values, 354
Binet–Cauchy formula, 123, 221
Birkhoff theorem, 159

Cauchy eigenvalue interlacing
 theorem, 269
Cauchy matrix, 137, 239
Cauchy sequence, 182
Cauchy–Schwarz inequality, xvii, 27,
 245
Cayley–Hamilton theorem, 87
characteristic polynomial, 21
Cholesky factorization, 201, 204
cofactor, 13
column rank, 51
column space, 55
column sum vector, 306
commutator, 305
companion matrix, 89, 91
complete space, 182
conjugate, 9
continuity argument, 62
contraction, 182, 206, 230, 362
 strict, 182
convergency, 182
convex combination, 159
convex hull, 112
coordinate, 4
Courant–Fischer theorem, 268

decomposable matrix, 339
decomposition
 Cholesky, 201, 204
 Jordan, 97
 LU, 85
 polar, 83
 QR, 85
 rank, 11, 201
 Schur, 79
 singular value, 82
 spectral, 81
 triangularization, 80
determinant, 11
differential operator, 17
dimension, 3
 finite, 3
 infinite, 3
dimension identity, 5
direct product, 117
direct sum, 4, 11
dual norm, 116

eigenspace, 23
eigenvalue, 19, 21

eigenvalue interlacing theorem, 269
eigenvector, 19
elementary λ-matrix, 94
elementary column operations, 10
elementary divisors, 94
elementary operation on λ-matrix, 94
elementary operations, 10
elementary row operations, 10
elementary symmetric function, 124, 145
Euclidean norm, 115
Euclidean space, 27, 30
even permutation, 11

Fan dominance theorem, 375
Fan eigenvalue majorization theorem, 356
Fan max-representation, 281
Fan–Hoffman theorem, 361
field of values, 107
Fischer inequality, 217, 225
fixed point, 183
Frobenius norm, 115
Frobenius–König theorem, 158
function
  concave, 340
  convex, 340
  increasing, 340
  Schur-convex, 341
  strictly convex, 340

generalized elementary matrix, 36, 40
geometric multiplicity, 24
Geršgorin disc theorem, 68
Gram matrix, 225
Gram–Schmidt orthonormalization, 33
Gram–Schmidt process, 33

Hölder inequality, 346
Hadamard inequality, 218, 225
Hadamard product, 117
Hilbert–Schmidt norm, 115
Hoffman–Wielandt theorem, 320
Horn theorem on singular values, 353
Horn–Johnson theorem, 378
Hua determinant inequality, 230, 231

image, 17
indecomposable matrix, 339
induced norm, 33, 113
inertia, 255
inner product, 27
inner product of vectors, 27
inner product space, 27
interpolation, 75, 144
invariant factors, 94
invariant subspace, 22, 23
inverse, 13
invertible λ-matrix, 94
invertible matrix, 13
involution, 179
irreducible matrix, 155, 339
isomorphism, 26

Jensen inequality, 340
Jordan block, 93
Jordan canonical form, 93
Jordan decomposition, 97
Jordan decomposition theorem, xvii
Jordan form, 93

Kantorovich inequality, 248, 249, 348
kernel, 17, 51
Kittaneh theorem, 319
Kronecker delta, 266
Kronecker product, 117
Ky Fan k-norm for matrix, 115, 375, 376
Ky Fan k-norm for vector, 373, 376

Löwner (partial) ordering, 207
Löwner ordering, 274
Löwner–Heinz theorem, 211
Laplace formula, 12
Laplace expansion formula, 12
length, 28
Levy–Desplanques theorem, 72
Lidskiĭ theorem, 363
Lieb–Thirring theorem, 368
linear transformation, 17
  addition, 22
  identity, 22, 25
  invertible, 25

product, 22
scalar multiplication, 22
linearly dependent, 3
linearly independent, 3
log-majorization, 344
LU factorization, 85

majorization, 325, 326, 349
matrix
M-, 170
$*$-congruent, 256
λ-, 93
$(0,1)$-, 306
addition, 8
adjoint, 13
backward identity, 78, 91
block, 10
Cauchy, 137, 239
circulant, 138
companion, 89, 91
compound, 122
conjugate, 9
constractive, 230
correlation, 243
decomposable, 339
definition, 8
diagonal, 9
direct sum, 11
doubly stochastic, 158, 334
doubly substochastic, 334
elementary, 10
elementary operations, 10
Fourier, 140, 142
function, 63
generalized elementary, 36, 40
Gram, 225
Hadamard, 150
Hankel, 142
Hermitian, 9, 253
idempotent, 125
identity, 9
indecomposable, 339
inverse, 13
invertible, 12
involutary, 125
involution, 179
irreducible, 155, 339
nilpotent, 125
nonnegative, 164
nonnegative definite, 199
nonsingular, 12
normal, 9, 293
order, 8
orthogonal, 9, 171
partitioned, 10
permuation, 155
positive, 164
positive definite, 199
positive semidefinite, 80, 199
primary permutation, 138, 156
product, 8
projection, 125
rank, 11
reducible, 155, 339
reflection, 177
rotation, 177
scalar, 9
scalar multiplication, 8
sequence, 64
series, 64
similar, 18
size, 8
skew-Hermitian, 257
skew-symmetric, 64
square root, 64, 81, 202
strictly diagonally dominant, 72
subpermutation, 338
symmetric, 9
symmetric unitary, 192
Toeplitz, 142
transpose, 9
tridiagonal, 133
under a basis, 19
unitary, 9, 171
upper-triangular, 9
Vandermonde, 74, 142, 143
zero, 9
matrix addition, 8
matrix function, 63
matrix norm, 113
matrix product, 8

matrix sequence, 64
matrix series, 64
matrix square root, 203
matrix-vector norm, 113, 373
metric, 182
metric space, 182
min-max representation, 266
minimal polynomial, 88
Minkowski inequality, 346
minor, 25
modulus of a matrix, 83, 287, 314
Moore–Penrose inverse, 377
multiplicative matrix norm, 113

norm, 28
  l_p, 373, 376
  Euclidean, 115
  Frobenius, 115
  Hilbert–Schmidt, 115
  induced, 33, 113
  Ky Fan k-matrix, 115, 375, 376
  Ky Fan k-vector, 373, 376
  matrix, 113
  matrix-vector, 113, 373
  multiplicative, 374
  multiplicative matrix, 113
  operator, 113
  Schatten p-matrix, 115
  Schatten p-vector, 376
  spectral, 114
  unitarily invariant, 115, 373
  vector, 34, 113, 372
normed space, 34
null space, 11, 17, 51
numerical radius, 109
numerical range, 107

odd permutation, 11
operator, 22
operator norm, 113
Oppenheim inequality, 242
orthogonal, 28
orthogonal projection, 127, 132
orthogonal set, 28
orthonormal, 28

permutation
  even, 11
  interchange, 163
  odd, 11
  of $\{1, 2, \ldots, n\}$, 163
permutation similarity, 155
Perron root, 169
Perron theorem, 167
Perron vector, 169
Poincaré eigenvalue interlacing theorem, 269, 271
polar decomposition, 83
positive linear transformation, 248
positive semidefinite matrix, 80
primary permutation matrix, 138
primitive root of unity, 139
principal submatrix, 25
product
  linear transformation, 22
  operator, 22
projection, 125, 127, 132

QR factorization, 85

range, 51
rank, 11
rank decomposition, 11
rank factorization, 201
Rayleigh–Ritz theorem, 267
reducible matrix, 155, 339
reflection, 177
rotation, 177
row rank, 51
row space, 55
row sum vector, 306

scalar, 2
scalar multiplication, 1
Schatten p-norm for matrix, 115
Schatten p-norm for vector, 376
Schur complement, 217, 227
Schur decomposition, 79
Schur inequality, 312
Schur product, 117
Schur theorem on Hadamard product, 234

Schur theorem on majorization, 349
Schur triangularization, 80
Schur triangularization theorem, xvii
Schur-convex function, 341
similarity, 18
singular value, 61, 82
singular value decomposition, 82
singular value decomposition theorem, xvii
solution space, 11
span, 3
spectral decomposition, 81
spectral decomposition theorem, xvii
spectral norm, 109, 114
spectral radius, 109, 165
spread, 324
square root, 64, 81
square root of a matrix, 81, 203
square root of a positive semidefinite matrix, 81, 202
standard basis, 3
Sturm eigenvalue interlacing theorem, 269
submatrix, 9
subpermutation, 338
subspace, 4
sum
    direct, 4
    vector spaces, 4
Sun theorem, 322
SVD, 82
Sylvester rank identity, 52
Sylvester's law of inertia, 256
symmetric gauge function, 372

tensor product, 117
Thompson theorem on matrix modulus, 289
Thompson theorem on sum of Hermitian matrices, 281
Toeplitz–Hausdorff theorem, 108
trace, 21
transpose, 9
triangle inequality, 28
triangularization, 80

uniqueness of matrix square root, 81, 202
unit, 28
unital linear transformation, 248
unitarily invariant norm, 115

Vandermonde determinant, 14
Vandermonde matrix, 74, 143
vector, 2
vector norm, 34, 113, 372
vector norm of matrix, 373
vector space, 1
von Neumann theorem, 375

Wang–Gong theorem, 367
weak majorization, 326
Weyl theorem on log-majorization, 353
Wielandt inequality, 247