SECOND EDITION

MATRIX ANALYSIS

ROGER A. HORN ■ CHARLES R. JOHNSON

CAMBRIDGE

Matrix Analysis

Second Edition

Linear algebra and matrix theory are fundamental tools in mathematical and physical science, as well as fertile fields for research. This new edition of the acclaimed text presents results of both classic and recent matrix analysis using canonical forms as a unifying theme, and demonstrates their importance in a variety of applications.

The authors have thoroughly revised, updated, and expanded on the first edition. The book opens with an extended summary of useful concepts and facts and includes numerous new topics and features, such as:

- New sections on the singular value and CS decompositions
- New applications of the Jordan canonical form
- A new section on the Weyr canonical form
- Expanded treatments of inverse problems and of block matrices
- A central role for the von Neumann trace theorem
- A new appendix with a modern list of canonical forms for a pair of Hermitian matrices and for a symmetric–skew symmetric pair
- Expanded index with more than 3,500 entries for easy reference
- More than 1,100 problems and exercises, many with hints, to reinforce understanding and develop auxiliary themes such as finite-dimensional quantum systems, the compound and adjugate matrices, and the Loewner ellipsoid
- A new appendix provides a collection of problem-solving hints.

Roger A. Horn is a Research Professor in the Department of Mathematics at the University of Utah. He is the author of *Topics in Matrix Analysis* (Cambridge University Press 1994).

Charles R. Johnson is the author of *Topics in Matrix Analysis* (Cambridge University Press 1994).

Matrix Analysis

Second Edition

Roger A. Horn

University of Utah

Charles R. Johnson

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521548236

© Roger A. Horn and Charles R. Johnson 1985, 2013

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 1985
First paperback edition 1990
Second edition first published 2013

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Horn, Roger A.
Matrix analysis / Roger A. Horn, Charles R. Johnson. – 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-521-83940-2 (hardback)
1. Matrices. I. Johnson, Charles R. II. Title.
QA188.H66 2012
512.9′434–dc23 2012012300

ISBN 978-0-521-83940-2 Hardback
ISBN 978-0-521-54823-6 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

To the matrix theory community

Contents

Preface to the Second Edition *page* xi
Preface to the First Edition xv

0 Review and Miscellanea **1**
0.0 Introduction 1
0.1 Vector spaces 1
0.2 Matrices 5
0.3 Determinants 8
0.4 Rank 12
0.5 Nonsingularity 14
0.6 The Euclidean inner product and norm 15
0.7 Partitioned sets and matrices 16
0.8 Determinants again 21
0.9 Special types of matrices 30
0.10 Change of basis 39
0.11 Equivalence relations 40

1 Eigenvalues, Eigenvectors, and Similarity **43**
1.0 Introduction 43
1.1 The eigenvalue–eigenvector equation 44
1.2 The characteristic polynomial and algebraic multiplicity 49
1.3 Similarity 57
1.4 Left and right eigenvectors and geometric multiplicity 75

2 Unitary Similarity and Unitary Equivalence **83**
2.0 Introduction 83
2.1 Unitary matrices and the QR factorization 83
2.2 Unitary similarity 94
2.3 Unitary and real orthogonal triangularizations 101
2.4 Consequences of Schur's triangularization theorem 108
2.5 Normal matrices 131

2.6 Unitary equivalence and the singular value decomposition 149
2.7 The *CS* decomposition 159

3 Canonical Forms for Similarity and Triangular Factorizations 163
3.0 Introduction 163
3.1 The Jordan canonical form theorem 164
3.2 Consequences of the Jordan canonical form 175
3.3 The minimal polynomial and the companion matrix 191
3.4 The real Jordan and Weyr canonical forms 201
3.5 Triangular factorizations and canonical forms 216

4 Hermitian Matrices, Symmetric Matrices, and Congruences 225
4.0 Introduction 225
4.1 Properties and characterizations of Hermitian matrices 227
4.2 Variational characterizations and subspace intersections 234
4.3 Eigenvalue inequalities for Hermitian matrices 239
4.4 Unitary congruence and complex symmetric matrices 260
4.5 Congruences and diagonalizations 279
4.6 Consimilarity and condiagonalization 300

5 Norms for Vectors and Matrices 313
5.0 Introduction 313
5.1 Definitions of norms and inner products 314
5.2 Examples of norms and inner products 320
5.3 Algebraic properties of norms 324
5.4 Analytic properties of norms 324
5.5 Duality and geometric properties of norms 335
5.6 Matrix norms 340
5.7 Vector norms on matrices 371
5.8 Condition numbers: inverses and linear systems 381

6 Location and Perturbation of Eigenvalues 387
6.0 Introduction 387
6.1 Geršgorin discs 387
6.2 Geršgorin discs – a closer look 396
6.3 Eigenvalue perturbation theorems 405
6.4 Other eigenvalue inclusion sets 413

7 Positive Definite and Semidefinite Matrices 425
7.0 Introduction 425
7.1 Definitions and properties 429
7.2 Characterizations and properties 438
7.3 The polar and singular value decompositions 448
7.4 Consequences of the polar and singular value decompositions 458
7.5 The Schur product theorem 477
7.6 Simultaneous diagonalizations, products, and convexity 485
7.7 The Loewner partial order and block matrices 493
7.8 Inequalities involving positive definite matrices 505

8 Positive and Nonnegative Matrices **517**
8.0 Introduction 517
8.1 Inequalities and generalities 519
8.2 Positive matrices 524
8.3 Nonnegative matrices 529
8.4 Irreducible nonnegative matrices 533
8.5 Primitive matrices 540
8.6 A general limit theorem 545
8.7 Stochastic and doubly stochastic matrices 547

Appendix A Complex Numbers **555**

Appendix B Convex Sets and Functions **557**

Appendix C The Fundamental Theorem of Algebra **561**

Appendix D Continuity of Polynomial Zeroes and Matrix Eigenvalues **563**

Appendix E Continuity, Compactness, and Weierstrass's Theorem **565**

Appendix F Canonical Pairs **567**

References 571
Notation 575
Hints for Problems 579
Index 607

Preface to the Second Edition

The basic structure of the first edition has been preserved in the second because it remains congruent with the goal of writing "a book that would be a useful modern treatment of a broad range of topics. . . [that] may be used as an undergraduate or graduate text and as a self-contained reference for a variety of audiences." The quotation is from the Preface to the First Edition, whose declaration of goals for the work remains unchanged.

What is different in the second edition?

The core role of canonical forms has been expanded as a unifying element in understanding similarity (complex, real, and simultaneous), unitary equivalence, unitary similarity, congruence, *congruence, unitary congruence, triangular equivalence, and other equivalence relations. More attention is paid to cases of equality in the many inequalities considered in the book. Block matrices are a ubiquitous feature of the exposition in the new edition.

Learning mathematics has never been a spectator sport, so the new edition continues to emphasize the value of exercises and problems for the active reader. Numerous 2-by-2 examples illustrate concepts throughout the book. Problem threads (some span several chapters) develop special topics as the foundation for them evolves in the text. For example, there are threads involving the adjugate matrix, the compound matrix, finite-dimensional quantum systems, the Loewner ellipsoid and the Loewner–John matrix, and normalizable matrices; see the index for page references for these threads. The first edition had about 690 problems; the second edition has more than 1,100. Many problems have hints; they may be found in an appendix that appears just before the index.

A comprehensive index is essential for a book that is intended for sustained use as a reference after initial use as a text. The index to the first edition had about 1,200 entries; the current index has more than 3,500 entries. An unfamiliar term encountered in the text should be looked up in the index, where a pointer to a definition (in Chapter 0 or elsewhere) is likely to be found.

New discoveries since 1985 have shaped the presentation of many topics and have stimulated inclusion of some new ones. A few examples of the latter are the Jordan

canonical form of a rank-one perturbation, motivated by enduring student interest in the *Google* matrix; a generalization of real normal matrices (normal matrices A such that $A\bar{A}$ is real); computable block matrix criteria for simultaneous unitary similarity or simultaneous unitary congruence; G. Belitskii's discovery that a matrix commutes with a Weyr canonical form if and only if it is block upper triangular and has a special structure; the discovery by K. C. O'Meara and C. Vinsonhaler that, unlike the corresponding situation for the Jordan canonical form, a commuting family can be simultaneously upper triangularized by similarity in such a way that any one specified matrix in the family is in Weyr canonical form; and canonical forms for congruence and *congruence.

Queries from many readers have motivated changes in the way that some topics are presented. For example, discussion of Lidskii's eigenvalue majorization inequalities was moved from a section primarily devoted to singular value inequalities to the section where majorization is discussed. Fortunately, a splendid new proof of Lidskii's inequalities by C. K. Li and R. Mathias became available and was perfectly aligned with Chapter 4's new approach to eigenvalue inequalities for Hermitian matrices. A second example is a new proof of Birkhoff's theorem, which has a very different flavor from the proof in the first edition.

Instructors accustomed to the order of topics in the first edition may be interested in a chapter-by-chapter summary of what is different in the new edition:

0. Chapter 0 has been expanded by about 75% to include a more comprehensive summary of useful concepts and facts. It is intended to serve as an as-needed reference. Definitions of terms and notations used throughout the book can be found here, but it has no exercises or problems. Formal courses and reading for self-study typically begin with Chapter 1.
1. Chapter 1 contains new examples related to similarity and the characteristic polynomial, as well as an enhanced emphasis on the role of left eigenvectors in matrix analysis.
2. Chapter 2 contains a detailed presentation of real orthogonal similarity, an exposition of McCoy's theorem on simultaneous triangularization, and a rigorous treatment of continuity of eigenvalues that makes essential use of both the *unitary* and *triangular* aspects of Schur's unitary triangularization theorem. Section 2.4 (Consequences of Schur's triangularization theorem) is almost twice the length of the corresponding section in the first edition. There are two new sections, one devoted to the singular value decomposition and one devoted to the CS decomposition. Early introduction of the singular value decomposition permits this essential tool of matrix analysis to be used throughout the rest of the book.
3. Chapter 3 approaches the Jordan canonical form via the Weyr characteristic; it contains an exposition of the Weyr canonical form and its unitary variant that were not in the first edition. Section 3.2 (Consequences of the Jordan canonical form) discusses many new applications; it contains 60% more material than the corresponding section in the first edition.
4. Chapter 4 now has a modern presentation of variational principles and eigenvalue inequalities for Hermitian matrices via subspace intersections. It contains an expanded treatment of inverse problems associated with interlacing and other

classical results. Its detailed treatment of unitary congruence includes Youla's theorem (a normal form for a square complex matrix A under unitary congruence that is associated with the eigenstructure of $A\bar{A}$), as well as canonical forms for conjugate normal, congruence normal, and squared normal matrices. It also has an exposition of recently discovered canonical forms for congruence and *congruence and new algorithms to construct a basis of a coneigenspace.

5. Chapter 5 contains an expanded discussion of norm duality, many new problems, and a treatment of semi-inner products that finds application in a discussion of finite-dimensional quantum systems in Chapter 7.
6. Chapter 6 has a new treatment of the "disjoint discs" aspect of Geršgorin's theorem and a reorganized discussion of eigenvalue perturbations, including differentiability of a simple eigenvalue.
7. Chapter 7 has been reorganized now that the singular value decomposition is introduced in Chapter 2. There is a new treatment of the polar decomposition, new factorizations related to the singular value decomposition, and special emphasis on row and column inclusion. The von Neumann trace theorem (proved via Birkhoff's theorem) is now the foundation on which many applications of the singular value decomposition are built. The Loewner partial order and block matrices are treated in detail with new techniques, as are the classical determinant inequalities for positive definite matrices.
8. Chapter 8 uses facts about left eigenvectors developed in Chapter 1 to streamline its exposition of the Perron–Frobenius theory of positive and nonnegative matrices.

D. Appendix D contains new explicit perturbation bounds for the zeroes of a polynomial and the eigenvalues of a matrix.

F. Appendix F tabulates a modern list of canonical forms for a pair of Hermitian matrices, or a pair of matrices, one of which is symmetric and the other is skew symmetric. These canonical pairs are applications of the canonical forms for congruence and *congruence presented in Chapter 4.

Readers who are curious about the technology of book making may be interested to know that this book began as a set of LaTeX files created manually by a company in India from hard copy of the first edition. Those files were edited and revised using the Scientific WorkPlace® graphical user interface and typesetting system.

The cover art for the second edition was the result of a lucky encounter on a Delta flight from Salt Lake City to Los Angeles in spring 2003. The young man in the middle seat said he was an artist who paints abstract paintings that are sometimes mathematically inspired. In the course of friendly conversation, he revealed that his special area of mathematical enjoyment was linear algebra, and that he had studied *Matrix Analysis*. After mutual expressions of surprise at the chance nature of our meeting, and a pleasant discussion, we agreed that appropriate cover art would enhance the visual appeal of the second edition; he said he would send something to consider. In due course a packet arrived from Seattle. It contained a letter and a stunning 4.5- by 5-inch color photograph, identified on the back as an image of a 72- by 66-inch oil on canvas, painted in 2002. The letter said that "the painting is entitled *Surprised Again on the Diagonal* and is inspired by the recurring prevalence of the diagonal in math whether it be in geometry, analysis, algebra, set theory or logic. I think that it would

be an attractive addition to your wonderful book." Thank you, Lun-Yi Tsai, for your wonderful cover art!

A great many students, instructors, and professional colleagues have contributed to the evolution of this new edition since its predecessor appeared in 1985. Special thanks are hereby acknowledged to T. Ando, Wayne Barrett, Ignat Domanov, Jim Fill, Carlos Martins da Fonseca, Tatiana Gerasimova, Geoffrey Goodson, Robert Guralnick, Thomas Hawkins, Eugene Herman, Khakim Ikramov, Ilse Ipsen, Dennis C. Jespersen, Hideki Kosaki, Zhongshan Li, Teck C. Lim, Ross A. Lippert, Roy Mathias, Dennis Merino, Arnold Neumaier, Kevin O'Meara, Peter Rosenthal, Vladimir Sergeichuk, Wasin So, Hugo Woerdeman, and Fuzhen Zhang.

R.A.H.

Preface to the First Edition

Linear algebra and matrix theory have long been fundamental tools in mathematical disciplines as well as fertile fields for research in their own right. In this book, and in the companion volume, *Topics in Matrix Analysis*, we present classical and recent results of matrix analysis that have proved to be important to applied mathematics. The book may be used as an undergraduate or graduate text and as a self-contained reference for a variety of audiences. We assume background equivalent to a one-semester elementary linear algebra course and knowledge of rudimentary analytical concepts. We begin with the notions of eigenvalues and eigenvectors; no prior knowledge of these concepts is assumed.

Facts about matrices, beyond those found in an elementary linear algebra course, are necessary to understand virtually any area of mathematical science, whether it be differential equations; probability and statistics; optimization; or applications in theoretical and applied economics, the engineering disciplines, or operations research, to name only a few. But until recently, much of the necessary material has occurred sporadically (or not at all) in the undergraduate and graduate curricula. As interest in applied mathematics has grown and more courses have been devoted to advanced matrix theory, the need for a text offering a broad selection of topics has become more apparent, as has the need for a modern reference on the subject.

There are several well-loved classics in matrix theory, but they are not well suited for general classroom use, nor for systematic individual study. A lack of problems, applications, and motivation; an inadequate index; and a dated approach are among the difficulties confronting readers of some traditional references. More recent books tend to be either elementary texts or treatises devoted to special topics. Our goal was to write a book that would be a useful modern treatment of a broad range of topics.

One view of "matrix analysis" is that it consists of those *topics* in linear algebra that have arisen out of the needs of mathematical analysis, such as multivariable calculus, complex variables, differential equations, optimization, and approximation theory. Another view is that matrix analysis is an *approach* to real and complex linear

algebraic problems that does not hesitate to use notions from analysis – such as limits, continuity, and power series – when these seem more efficient or natural than a purely algebraic approach. Both views of matrix analysis are reflected in the choice and treatment of topics in this book. We prefer the term *matrix analysis* to *linear algebra* as an accurate reflection of the broad scope and methodology of the field.

For review and convenience in reference, Chapter 0 contains a summary of necessary facts from elementary linear algebra, as well as other useful, though not necessarily elementary, facts. Chapters 1, 2, and 3 contain mainly core material likely to be included in any second course in linear algebra or matrix theory: a basic treatment of eigenvalues, eigenvectors, and similarity; unitary similarity, Schur triangularization and its implications, and normal matrices; and canonical forms and factorizations, including the Jordan form, LU factorization, QR factorization, and companion matrices. Beyond this, each chapter is developed substantially independently and treats in some depth a major topic:

1. *Hermitian and complex symmetric matrices* (Chapter 4). We give special emphasis to variational methods for studying eigenvalues of Hermitian matrices and include an introduction to the notion of majorization.
2. *Norms on vectors and matrices* (Chapter 5) are essential for error analyses of numerical linear algebraic algorithms and for the study of matrix power series and iterative processes. We discuss the algebraic, geometric, and analytic properties of norms in some detail and make a careful distinction between those norm results for matrices that depend on the submultiplicativity axiom for matrix norms and those that do not.
3. *Eigenvalue location and perturbation results* (Chapter 6) for general (not necessarily Hermitian) matrices are important for many applications. We give a detailed treatment of the theory of Geršgorin regions, and some of its modern refinements, and of relevant graph theoretic concepts.
4. *Positive definite matrices* (Chapter 7) and their applications, including inequalities, are considered at some length. A discussion of the polar and singular value decompositions is included, along with applications to matrix approximation problems.
5. *Entry-wise nonnegative and positive matrices* (Chapter 8) arise in many applications in which nonnegative quantities necessarily occur (probability, economics, engineering, etc.), and their remarkable theory reflects the applications. Our development of the theory of nonnegative, positive, primitive, and irreducible matrices proceeds in elementary steps based on the use of norms.

In the companion volume, further topics of similar interest are treated: the field of values and generalizations; inertia, stable matrices, M-matrices and related special classes; matrix equations, Kronecker and Hadamard products; and various ways in which functions and matrices may be linked.

This book provides the basis for a variety of one- or two-semester courses through selection of chapters and sections appropriate to a particular audience. We recommend that an instructor make a careful preselection of sections and portions of sections of the book for the needs of a particular course. This would probably include Chapter 1, much of Chapters 2 and 3, and facts about Hermitian matrices and norms from Chapters 4 and 5.

Most chapters contain some relatively specialized or nontraditional material. For example, Chapter 2 includes not only Schur's basic theorem on unitary triangularization of a single matrix but also a discussion of simultaneous triangularization of families of matrices. In the section on unitary equivalence, our presentation of the usual facts is followed by a discussion of trace conditions for two matrices to be unitarily equivalent. A discussion of complex symmetric matrices in Chapter 4 provides a counterpoint to the development of the classical theory of Hermitian matrices. Basic aspects of a topic appear in the initial sections of each chapter, while more elaborate discussions occur at the ends of sections or in later sections. This strategy has the advantage of presenting topics in a sequence that enhances the book's utility as a reference. It also provides a rich variety of options to the instructor.

Many of the results discussed are valid or can be generalized to be valid for matrices over other fields or in some broader algebraic setting. However, we deliberately confine our domain to the real and complex fields where familiar methods of classical analysis as well as formal algebraic techniques may be employed.

Though we generally consider matrices to have complex entries, most examples are confined to real matrices, and no deep knowledge of complex analysis is required. Acquaintance with the arithmetic of complex numbers is necessary for an understanding of matrix analysis and is covered to the extent necessary in an appendix. Other brief appendices cover several peripheral, but essential, topics such as Weierstrass's theorem and convexity.

We have included many exercises and problems because we feel these are essential to the development of an understanding of the subject and its implications. The exercises occur throughout as part of the development of each section; they are generally elementary and of immediate use in understanding the concepts. We recommend that the reader work at least a broad selection of these. Problems are listed (in no particular order) at the end of each section; they cover a range of difficulties and types (from theoretical to computational) and they may extend the topic, develop special aspects, or suggest alternate proofs of major ideas. Significant hints are given for the more difficult problems. The results of some problems are referred to in other problems or in the text itself. We cannot overemphasize the importance of the reader's active involvement in carrying out the exercises and solving problems.

While the book itself is not about applications, we have, for motivational purposes, begun each chapter with a section outlining a few applications to introduce the topic of the chapter.

Readers who wish to consult alternative treatments of a topic for additional information are referred to the books listed in the References section following the appendices.

The list of book references is not exhaustive. As a practical concession to the limits of space in a general multitopic book, we have minimized the number of citations in the text. A small selection of references to papers – such as those we have explicitly used – does occur at the end of most sections accompanied by a brief discussion, but we have made no attempt to collect historical references to classical results. Extensive bibliographies are provided in the more specialized books we have referenced.

We appreciate the helpful suggestions of our colleagues and students who have taken the time to convey their reactions to the class notes and preliminary manuscripts

that were the precursors of the book. They include Wayne Barrett, Leroy Beasley, Bryan Cain, David Carlson, Dipa Choudhury, Risana Chowdhury, Yoo Pyo Hong, Dmitry Krass, Dale Olesky, Stephen Pierce, Leiba Rodman, and Pauline van den Driessche.

R.A.H.
C.R.J.

CHAPTER 0

Review and Miscellanea

0.0 Introduction

In this initial chapter we summarize many useful concepts and facts, some of which provide a foundation for the material in the rest of the book. Some of this material is included in a typical elementary course in linear algebra, but we also include additional useful items, even though they do not arise in our subsequent exposition. The reader may use this chapter as a review before beginning the main part of the book in Chapter 1; subsequently, it can serve as a convenient reference for notation and definitions that are encountered in later chapters. We assume that the reader is familiar with the basic concepts of linear algebra and with mechanical aspects of matrix manipulations, such as matrix multiplication and addition.

0.1 Vector spaces

A finite dimensional vector space is the fundamental setting for matrix analysis.

0.1.1 Scalar field. Underlying a vector space is its *field*, or set of scalars. For our purposes, that underlying field is typically the real numbers $\mathbf{R}$ or the complex numbers $\mathbf{C}$ (see Appendix A), but it could be the rational numbers, the integers modulo a specified prime number, or some other field. When the field is unspecified, we denote it by the symbol $\mathbf{F}$. To qualify as a field, a set must be closed under two binary operations: "addition" and "multiplication." Both operations must be associative and commutative, and each must have an identity element in the set; inverses must exist in the set for all elements under addition and for all elements except the additive identity under multiplication; multiplication must be distributive over addition.

0.1.2 Vector spaces. A *vector space* V over a field $\mathbf{F}$ is a set V of objects (called *vectors*) that is closed under a binary operation ("addition") that is associative and commutative and has an identity (the *zero vector*, denoted by 0) and additive inverses

in the set. The set is also closed under an operation of "scalar multiplication" of the vectors by elements of the scalar field $\mathbf{F}$, with the following properties for all $a, b \in \mathbf{F}$ and all $x, y \in V$: $a(x + y) = ax + ay$, $(a + b)x = ax + bx$, $a(bx) = (ab)x$, and $ex = x$ for the multiplicative identity $e \in \mathbf{F}$.

For a given field $\mathbf{F}$ and a given positive integer n, the set $\mathbf{F}^n$ of n-tuples with entries from $\mathbf{F}$ forms a vector space over $\mathbf{F}$ under entrywise addition in $\mathbf{F}^n$. Our convention is that *elements of* $\mathbf{F}^n$ *are always presented as column vectors*; we often call them *n-vectors*. The special cases $\mathbf{R}^n$ and $\mathbf{C}^n$ are the basic vector spaces of this book; $\mathbf{R}^n$ is a real vector space (that is, a vector space over the real field), while $\mathbf{C}^n$ is both a real vector space and a complex vector space (a vector space over the complex field). The set of polynomials with real or with complex coefficients (of no more than a specified degree *or* of arbitrary degree) and the set of real-valued or complex-valued functions on subsets of $\mathbf{R}$ or $\mathbf{C}$ (all with the usual notions of addition of functions and multiplication of a function by a scalar) are also examples of real or complex vector spaces.

0.1.3 Subspaces, span, and linear combinations. A *subspace* of a vector space V over a field $\mathbf{F}$ is a subset of V that is, by itself, a vector space over $\mathbf{F}$ using the same operations of vector addition and scalar multiplication as in V. A subset of V is a subspace precisely when it is closed under these two operations. For example, $\{[a, b, 0]^T : a, b \in \mathbf{R}\}$ is a subspace of $\mathbf{R}^3$; see (0.2.5) for the transpose notation. An intersection of subspaces is always a subspace; a union of subspaces need not be a subspace. The subsets $\{0\}$ and V are always subspaces of V, so they are often called *trivial subspaces*; a subspace of V is said to be *nontrivial* if it is different from both $\{0\}$ and V. A subspace of V is said to be a *proper subspace* if it is not equal to V. We call $\{0\}$ the *zero vector space*. Since a vector space always contains the zero vector, a subspace cannot be empty.

If S is a subset of a vector space V over a field $\mathbf{F}$, span S is the intersection of all subspaces of V that contain S. If S is nonempty, then span $S = \{a_1 v_1 + \cdots + a_k v_k : v_1, \ldots, v_k \in S, a_1, \ldots, a_k \in \mathbf{F}, \text{and } k = 1, 2, \ldots\}$; if S is empty, it is contained in every subspace of V. The intersection of every subspace of V is the subspace $\{0\}$, so the definition ensures that span $S = \{0\}$. Notice that span S is always a subspace even if S is not a subspace; S is said to *span* V if span $S = V$.

A *linear combination* of vectors in a vector space V over a field $\mathbf{F}$ is any expression of the form $a_1 v_1 + \cdots + a_k v_k$ in which k is a positive integer, $a_1, \ldots, a_k \in \mathbf{F}$, and $v_1, \ldots, v_k \in V$. Thus, the span of a nonempty subset S of V consists of all linear combinations of finitely many vectors in S. A linear combination $a_1 v_1 + \cdots + a_k v_k$ is *trivial* if $a_1 = \cdots = a_k = 0$; otherwise, it is *nontrivial*. A linear combination is by definition a sum of *finitely many* elements of a vector space.

Let S_1 and S_2 be subspaces of a vector space over a field $\mathbf{F}$. The *sum* of S_1 and S_2 is the subspace

$$S_1 + S_2 = \text{span}\,\{S_1 \cup S_2\} = \{x + y : x \in S_1, y \in S_2\}$$

If $S_1 \cap S_2 = \{0\}$, we say that the sum of S_1 and S_2 is a *direct sum* and write it as $S_1 \oplus S_2$; every $z \in S_1 \oplus S_2$ can be written as $z = x + y$ with $x \in S_1$ and $y \in S_2$ in one and only one way.

0.1.4 Linear dependence and linear independence. We say that a finite list of vectors $v_1, \ldots, v_k$ in a vector space V over a field $\mathbf{F}$ is *linearly dependent* if and only if there are scalars $a_1, \ldots, a_k \in \mathbf{F}$, not all zero, such that $a_1x_1 + \cdots + a_kx_k = 0$. Thus, a list of vectors $v_1, \ldots, v_k$ is linearly dependent if and only if some nontrivial linear combination of $v_1, \ldots, v_k$ is the zero vector. It is often convenient to say that "$v_1, \ldots, v_k$ are linearly dependent" instead of the more formal statement "the list of vectors $v_1, \ldots, v_k$ is linearly dependent." A list of vectors $v_1, \ldots, v_k$ is said to have *length* k. A list of two or more vectors is linearly dependent if one of the vectors is a linear combination of some of the others; in particular, a list of two or more vectors in which two of the vectors in the list are identical is linearly dependent. Two vectors are linearly dependent if and only if one of the vectors is a scalar multiple of the other. A list consisting only of the zero vector is linearly dependent since $a_1 0 = 0$ for $a_1 = 1$.

A finite list of vectors $v_1, \ldots, v_k$ in a vector space V over a field $\mathbf{F}$ is *linearly independent* if it is not linearly dependent. Again, it can be convenient to say that "$v_1, \ldots, v_k$ are linearly independent" instead of "the list of vectors $v_1, \ldots, v_k$ is linearly independent."

Sometimes one encounters natural lists of vectors that have infinitely many elements, for example, the monomials $1, t, t^2, t^3, \ldots$ in the vector space of all polynomials with real coefficients or the complex exponentials $1, e^{it}, e^{2it}, e^{3it}, \ldots$ in the vector space of complex-valued continuous functions that are periodic on $[0, 2\pi]$.

If certain vectors in a list (finite or infinite) are deleted, the resulting list is a *sublist* of the original list. An infinite list of vectors is said to be linearly dependent if some finite sublist is linearly dependent; it is said to be linearly independent if every finite sublist is linearly independent. Any sublist of a linearly independent list of vectors is linearly independent; any list of vectors that has a linearly dependent sublist is linearly dependent. Since a list consisting only of the zero vector is linearly dependent, any list of vectors that contains the zero vector is linearly dependent. A list of vectors can be linearly dependent, while any proper sublist is linearly independent; see (1.4.P12). An empty list of vectors is not linearly dependent, so it is linearly independent.

The *cardinality* of a finite set is the number of its (necessarily distinct) elements. For a given list of vectors $v_1, \ldots, v_k$ in a vector space V, the cardinality of the *set* $\{v_1, \ldots, v_k\}$ is less than k if and only if two or more vectors in the list are identical; if $v_1, \ldots, v_k$ are linearly independent, then the cardinality of the set $\{v_1, \ldots, v_k\}$ is k. The *span* of a list of vectors (finite or not) is the span of the set of elements of the list; a list of vectors *spans* V if V is the span of the list.

A set S of vectors is said to be linearly independent if every finite list of distinct vectors in S is linearly independent; S is said to be linearly dependent if some finite list of distinct vectors in S is linearly dependent.

0.1.5 Basis. A linearly independent list of vectors in a vector space V whose span is V is a *basis* for V. Each element of V can be represented as a linear combination of vectors in a basis in one and only one way; this is no longer true if any element whatsoever is appended to or deleted from the basis. A linearly independent list of vectors in V is a basis of V if and only if no list of vectors that properly contains it is linearly independent. A list of vectors that spans V is a basis for V if and only if no proper sublist of it spans V. The empty list is a basis for the zero vector space.

0.1.6 Extension to a basis. Any linearly independent list of vectors in a vector space V may be extended, perhaps in more than one way, to a basis of V. A vector space can have a basis that is not finite; for example, the infinite list of monomials $1, t, t^2, t^3, \ldots$ is a basis for the real vector space of all polynomials with real coefficients; each polynomial is a unique linear combination of (finitely many) elements in the basis.

0.1.7 Dimension. If there is a positive integer n such that a basis of the vector space V contains exactly n vectors, then every basis of V consists of exactly n vectors; this common cardinality of bases is the *dimension* of the vector space V and is denoted by $\dim V$. In this event, V is *finite-dimensional*; otherwise V is *infinite-dimensional*. In the infinite-dimensional case, there is a one-to-one correspondence between the elements of any two bases. The real vector space $\mathbf{R}^n$ has dimension n. The vector space $\mathbf{C}^n$ has dimension n over the field $\mathbf{C}$ but dimension $2n$ over the field $\mathbf{R}$. The basis $e_1, \ldots, e_n$ of $\mathbf{F}^n$ in which each n-vector e_i has a 1 as its ith entry and 0s elsewhere is called the *standard basis*.

It is convenient to say "V is an n-dimensional vector space" as a shorthand for "V is a finite-dimensional vector space whose dimension is n." Any subspace of an n-dimensional vector space is finite-dimensional; its dimension is strictly less than n if it is a proper subspace.

Let V be a finite-dimensional vector space and let S_1 and S_2 be two given subspaces of V. The *subspace intersection lemma* is

$$\dim(S_1 \cap S_2) + \dim(S_1 + S_2) = \dim S_1 + \dim S_2 \tag{0.1.7.1}$$

Rewriting this identity as

$$\begin{aligned} \dim(S_1 \cap S_2) &= \dim S_1 + \dim S_2 - \dim(S_1 + S_2) \\ &\geq \dim S_1 + \dim S_2 - \dim V \end{aligned} \tag{0.1.7.2}$$

reveals the useful fact that if $\delta = \dim S_1 + \dim S_2 - \dim V \geq 1$, then the subspace $S_1 \cap S_2$ has dimension at least δ, and hence it contains δ linearly independent vectors, namely, any δ elements of a basis of $S_1 \cap S_2$. In particular, $S_1 \cap S_2$ contains a nonzero vector. An induction argument shows that if $S_1, \ldots, S_k$ are subspaces of V, and if $\delta = \dim S_1 + \cdots + \dim S_k - (k-1)\dim V \geq 1$, then

$$\dim(S_1 \cap \cdots \cap S_k) \geq \delta \tag{0.1.7.3}$$

and hence $S_1 \cap \cdots \cap S_k$ contains δ linearly independent vectors; in particular, it contains a nonzero vector.

0.1.8 Isomorphism. If U and V are vector spaces over the same scalar field $\mathbf{F}$, and if $f : U \to V$ is an invertible function such that $f(ax + by) = af(x) + bf(y)$ for all $x, y \in U$ and all $a, b \in \mathbf{F}$, then f is said to be an *isomorphism* and U and V are said to be isomorphic ("same structure"). Two finite-dimensional vector spaces over the same field are isomorphic if and only if they have the same dimension; thus, any n-dimensional vector space over $\mathbf{F}$ is isomorphic to $\mathbf{F}^n$. Any n-dimensional real vector space is, therefore, isomorphic to $\mathbf{R}^n$, and any n-dimensional complex vector space is isomorphic to $\mathbf{C}^n$. Specifically, if V is an n-dimensional vector space over a field $\mathbf{F}$

with specified basis $\mathcal{B} = \{x_1, \ldots, x_n\}$, then, since any element $x \in V$ may be written uniquely as $x = a_1x_1 + \cdots + a_nx_n$ in which each $a_i \in \mathbf{F}$, we may identify x with the n-vector $[x]_{\mathcal{B}} = [a_1 \ \ldots \ a_n]^T$. For any basis $\mathcal{B}$, the mapping $x \to [x]_{\mathcal{B}}$ is an isomorphism between V and $\mathbf{F}^n$.

0.2 Matrices

The fundamental object of study here may be thought of in two important ways: as a rectangular array of scalars and as a linear transformation between two vector spaces, given specified bases for each space.

0.2.1 Rectangular arrays. A *matrix* is an m-by-n array of scalars from a field $\mathbf{F}$. If $m = n$, the matrix is said to be *square*. The set of all m-by-n matrices over $\mathbf{F}$ is denoted by $M_{m,n}(\mathbf{F})$, and $M_{n,n}(\mathbf{F})$ is often denoted by $M_n(\mathbf{F})$. The vector spaces $M_{n,1}(\mathbf{F})$ and $\mathbf{F}^n$ are identical. If $\mathbf{F} = \mathbf{C}$, then $M_n(\mathbf{C})$ is further abbreviated to M_n, and $M_{m,n}(\mathbf{C})$ to $M_{m,n}$. Matrices are typically denoted by capital letters, and their scalar entries are typically denoted by doubly subscripted lowercase letters. For example, if

$$A = \begin{bmatrix} 2 & -\frac{3}{2} & 0 \\ -1 & \pi & 4 \end{bmatrix} = [a_{ij}]$$

then $A \in M_{2,3}(\mathbf{R})$ has entries $a_{11} = 2, a_{12} = -3/2, a_{13} = 0, a_{21} = -1, a_{22} = \pi$, $a_{23} = 4$. A *submatrix* of a given matrix is a rectangular array lying in specified subsets of the rows and columns of a given matrix. For example, $[\pi \ 4]$ is a submatrix (lying in row 2 and columns 2 and 3) of A.

Suppose that $A = [a_{ij}] \in M_{n,m}(\mathbf{F})$. The *main diagonal* of A is the list of entries $a_{11}, a_{22}, \ldots, a_{qq}$, in which $q = \min\{n, m\}$. It is sometimes convenient to express the main diagonal of A as a vector $\operatorname{diag} A = [a_{ii}]_{i=1}^{q} \in \mathbf{F}^q$. The pth *superdiagonal* of A is the list $a_{1,p+1}, a_{2,p+2}, \ldots, a_{k,p+k}$, in which $k = \min\{n, m - p\}$, $p = 0, 1, 2, \ldots, m - 1$; the pth *subdiagonal* of A is the list $a_{p+1,1}, a_{p+2,2}, \ldots, a_{p+\ell,\ell}$, in which $\ell = \min\{n - p, m\}$, $p = 0, 1, 2, \ldots, n - 1$.

0.2.2 Linear transformations. Let U be an n-dimensional vector space and let V be an m-dimensional vector space, both over the same field $\mathbf{F}$; let $\mathcal{B}_U$ be a basis of U and let $\mathcal{B}_V$ be a basis of V. We may use the isomorphisms $x \to [x]_{\mathcal{B}_U}$ and $y \to [y]_{\mathcal{B}_V}$ to represent vectors in U and V as n-vectors and m-vectors over $\mathbf{F}$, respectively. A *linear transformation* is a function $T : U \to V$ such that $T(a_1x_1 + a_2x_2) = a_1T(x_1) + a_2T(x_2)$ for any scalars a_1, a_2 and vectors x_1, x_2. A matrix $A \in M_{m,n}(\mathbf{F})$ corresponds to a linear transformation $T : U \to V$ in the following way: $y = T(x)$ if and only if $[y]_{\mathcal{B}_V} = A[x]_{\mathcal{B}_U}$. The matrix A is said to *represent the linear transformation* T (relative to the bases $\mathcal{B}_U$ and $\mathcal{B}_V$); the representing matrix A depends on the bases chosen. When we study a matrix A, we realize that we are studying a linear transformation relative to a particular choice of bases, but explicit appeal to the bases is usually not necessary.

0.2.3 Vector spaces associated with a matrix or linear transformation. Any n-dimensional vector space over $\mathbf{F}$ may be identified with $\mathbf{F}^n$; we may think of

$A \in M_{m,n}(\mathbf{F})$ as a linear transformation $x \to Ax$ from $\mathbf{F}^n$ to $\mathbf{F}^m$ (and also as an array). The *domain* of this linear transformation is $\mathbf{F}^n$; its *range* is range $A = \{y \in \mathbf{F}^m : y = Ax\}$ for some $x \in \mathbf{F}^n$; its *null space* is nullspace $A = \{x \in \mathbf{F}^n : Ax = 0\}$. The range of A is a subspace of $\mathbf{F}^m$, and the null space of A is a subspace of $\mathbf{F}^n$. The dimension of nullspace A is denoted by nullity A; the dimension of range A is denoted by rank A. These numbers are related by the *rank-nullity theorem*

$$\dim(\text{range } A) + \dim(\text{nullspace } A) = \text{rank } A + \text{nullity } A = n \tag{0.2.3.1}$$

for $A \in M_{m,n}(\mathbf{F})$. The null space of A is a set of vectors in $\mathbf{F}^n$ whose entries satisfy m homogeneous linear equations.

0.2.4 Matrix operations. Matrix addition is defined entrywise for arrays of the same dimensions and is denoted by $+$ ("$A + B$"). It corresponds to addition of linear transformations (relative to the same basis), and it inherits commutativity and associativity from the scalar field. The *zero matrix* (all entries are zero) is the additive identity, and $M_{m,n}(\mathbf{F})$ is a vector space over $\mathbf{F}$. Matrix multiplication is denoted by juxtaposition ("AB") and corresponds to the composition of linear transformations. Therefore, it is defined only when $A \in M_{m,n}(\mathbf{F})$ and $B \in M_{n,q}(\mathbf{F})$. It is associative, but not always commutative. For example,

$$\begin{bmatrix} 1 & 2 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \neq \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 3 & 8 \end{bmatrix}$$

The *identity matrix*

$$I = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} \in M_n(\mathbf{F})$$

is the multiplicative identity in $M_n(\mathbf{F})$; its main diagonal entries are 1, and all other entries are 0. The identity matrix and any scalar multiple of it (a *scalar matrix*) commute with every matrix in $M_n(\mathbf{F})$; they are the only matrices that do so. Matrix multiplication is distributive over matrix addition.

The symbol 0 is used throughout the book to denote each of the following: the zero scalar of a field, the zero vector of a vector space, the zero n-vector in $\mathbf{F}^n$ (all entries equal to the zero scalar in $\mathbf{F}$), and the zero matrix in $M_{m,n}(\mathbf{F})$ (all entries equal to the zero scalar). The symbol I denotes the identity matrix of any size. If there is potential for confusion, we indicate the dimension of a zero or identity matrix with subscripts, for example, $0_{p,q}$, 0_k, or I_k.

0.2.5 The transpose, conjugate transpose, and trace. If $A = [a_{ij}] \in M_{m,n}(\mathbf{F})$, the *transpose* of A, denoted by A^T, is the matrix in $M_{n,m}(\mathbf{F})$ whose i, j entry is a_{ji}; that is, rows are exchanged for columns and vice versa. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Of course, $(A^T)^T = A$. The *conjugate transpose* (sometimes called the *adjoint* or *Hermitian adjoint*) of $A \in M_{m,n}(\mathbf{C})$, is denoted by A^* and defined by $A^* = \bar{A}^T$, in

which $\bar{A}$ is the entrywise conjugate. For example,

$$\begin{bmatrix} 1+i & 2-i \\ -3 & -2i \end{bmatrix}^* = \begin{bmatrix} 1-i & -3 \\ 2+i & 2i \end{bmatrix}$$

Both the transpose and the conjugate transpose obey the *reverse-order law*: $(AB)^* = B^*A^*$ and $(AB)^T = B^TA^T$. For the complex conjugate of a product, there is no reversing: $\overline{AB} = \bar{A}\bar{B}$. If x, y are real or complex vectors of the same size, then y^*x is a scalar and its conjugate transpose and complex conjugate are the same: $(y^*x)^* = \overline{y^*x} = x^*y = y^T\bar{x}$.

Many important classes of matrices are defined by identities involving the transpose or conjugate transpose. For example, $A \in M_n(\mathbf{F})$ is said to be *symmetric* if $A^T = A$, *skew symmetric* if $A^T = -A$, and *orthogonal* if $A^TA = I$; $A \in M_n(\mathbf{C})$ is said to be *Hermitian* if $A^* = A$, *skew Hermitian* if $A^* = -A$, *essentially Hermitian* if $e^{i\theta}A$ is Hermitian for some $\theta \in \mathbf{R}$, *unitary* if $A^*A = I$, and *normal* if $A^*A = AA^*$.

Each $A \in M_n(\mathbf{F})$ can be written in exactly one way as $A = S(A) + C(A)$, in which $S(A)$ is symmetric and $C(A)$ is skew symmetric: $S(A) = \frac{1}{2}(A + A^T)$ is the *symmetric part* of A; $C(A) = \frac{1}{2}(A - A^T)$ is the *skew-symmetric part* of A.

Each $A \in M_{m,n}(\mathbf{C})$ can be written in exactly one way as $A = B + iC$, in which $B, C \in M_{m,n}(\mathbf{R})$: $B = \frac{1}{2}(A + \bar{A})$ is the *real part* of A; $C = \frac{1}{2i}(A - \bar{A})$ is the *imaginary part* of A.

Each $A \in M_n(\mathbf{C})$ can be written in exactly one way as $A = H(A) + iK(A)$, in which $H(A)$ and $K(A)$ are Hermitian: $H(A) = \frac{1}{2}(A + A^*)$ is the *Hermitian part* of A; $iK(A) = \frac{1}{2}(A - A^*)$ is the *skew-Hermitian part* of A. The representation $A = H(A) + iK(A)$ of a complex or real matrix is its *Toeplitz decomposition*.

The *trace* of $A = [a_{ij}] \in M_{m,n}(\mathbf{F})$ is the sum of its main diagonal entries: $\operatorname{tr} A = a_{11} + \cdots + a_{qq}$, in which $q = \min\{m, n\}$. For any $A = [a_{ij}] \in M_{m,n}(\mathbf{C})$, $\operatorname{tr} AA^* = \operatorname{tr} A^*A = \sum_{i,j} |a_{ij}|^2$, so

$$\operatorname{tr} AA^* = 0 \text{ if and only if } A = 0 \tag{0.2.5.1}$$

A vector $x \in \mathbf{F}^n$ is *isotropic* if $x^Tx = 0$. For example, $[1\ i]^T \in \mathbf{C}^2$ is a nonzero isotropic vector. There are no nonzero isotropic vectors in $\mathbf{R}^n$.

0.2.6 Metamechanics of matrix multiplication. In addition to the conventional definition of matrix-vector and matrix-matrix multiplication, several alternative viewpoints can be useful.

1. If $A \in M_{m,n}(\mathbf{F})$, $x \in \mathbf{F}^n$, and $y \in \mathbf{F}^m$, then the (column) vector Ax is a linear combination of the columns of A; the coefficients of the linear combination are the entries of x. The row vector y^TA is a linear combination of the rows of A; the coefficients of the linear combination are the entries of y.
2. If b_j is the jth column of B and a_i^T is the ith row of A, then the jth column of AB is Ab_j and the ith row of AB is a_i^TB.

To paraphrase, in the matrix product AB, *left multiplication by A multiplies the columns of B* and *right multiplication by B multiplies the rows of A*. See (0.9.1) for an important special case of this observation when one of the factors is a diagonal matrix.

Suppose that $A \in M_{m,p}(\mathbf{F})$ and $B \in M_{n,q}(\mathbf{F})$. Let a_k be the kth column of A and let b_k be the kth column of B. Then

3. If $m = n$, then $A^T B = \left[a_i^T b_j\right]$: the i, j entry of $A^T B$ is the scalar $a_i^T b_j$.
4. If $p = q$, then $AB^T = \sum_{k=1}^n a_k b_k^T$: each summand is an m-by-n matrix, the *outer product* of a_k and b_k.

0.2.7 Column space and row space of a matrix. The range of $A \in M_{m,n}(\mathbf{F})$ is also called its *column space* because Ax is a linear combination of the columns of A for any $x \in \mathbf{F}^n$ (the entries of x are the coefficients in the linear combination); range A is the span of the columns of A. Analogously, $\{y^T A : y \in \mathbf{F}^m\}$ is called the *row space* of A. If the column space of $A \in M_{m,n}(\mathbf{F})$ is contained in the column space of $B \in M_{m,k}(\mathbf{F})$, then there is some $X \in M_{k,n}(\mathbf{F})$ such that $A = BX$ (and conversely); the entries in column j of X tell how to express column j of A as a linear combination of the columns of B.

If $A \in M_{m,n}(\mathbf{F})$ and $B \in M_{m,q}(\mathbf{F})$, then

$$\text{range } A + \text{range } B = \text{range} \begin{bmatrix} A & B \end{bmatrix} \tag{0.2.7.1}$$

If $A \in M_{m,n}(\mathbf{F})$ and $B \in M_{p,n}(\mathbf{F})$, then

$$\text{nullspace } A \cap \text{nullspace } B = \text{nullspace} \begin{bmatrix} A \\ B \end{bmatrix} \tag{0.2.7.2}$$

0.2.8 The all-ones matrix and vector. In $\mathbf{F}^n$, every entry of the vector $e = e_1 + \cdots + e_n$ is 1. Every entry of the matrix $J_n = ee^T$ is 1.

0.3 Determinants

Often in mathematics, it is useful to summarize a multivariate phenomenon with a single number, and the determinant function is an example of this. Its domain is $M_n(\mathbf{F})$ (square matrices only), and it may be presented in several different ways. We denote the determinant of $A \in M_n(\mathbf{F})$ by $\det A$.

0.3.1 Laplace expansion by minors along a row or column. The determinant may be defined inductively for $A = [a_{ij}] \in M_n(\mathbf{F})$ in the following way. Assume that the determinant is defined over $M_{n-1}(\mathbf{F})$ and let $A_{ij} \in M_{n-1}(\mathbf{F})$ denote the submatrix of $A \in M_n(\mathbf{F})$ obtained by deleting row i and column j of A. Then, for any $i, j \in \{1, \ldots, n\}$, we have

$$\det A = \sum_{k=1}^n (-1)^{i+k} a_{ik} \det A_{ik} = \sum_{k=1}^n (-1)^{k+j} a_{kj} \det A_{kj} \tag{0.3.1.1}$$

The first sum is the *Laplace expansion by minors along row i*; the second sum is the *Laplace expansion by minors along column j*. This inductive presentation begins by

defining the determinant of a 1-by-1 matrix to be the value of the single entry. Thus,

$$\det[\,a_{11}] = a_{11}$$

$$\det\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{aligned} &a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{aligned}$$

and so on. Notice that $\det A^T = \det A$, $\det A^* = \overline{\det A}$ if $A \in M_n(\mathbf{C})$, and $\det I = 1$.

0.3.2 Alternating sums and permutations. A *permutation* of $\{1, \ldots, n\}$ is a one-to-one function $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$. The *identity permutation* satisfies $\sigma(i) = i$ for each $i = 1, \ldots, n$. There are $n!$ distinct permutations of $\{1, \ldots, n\}$, and the collection of all such permutations forms a group under composition of functions.

Consistent with the low-dimensional examples in (0.3.1), for $A = [a_{ij}] \in M_n(\mathbf{F})$ we have the alternative presentation

$$\det A = \sum_{\sigma} \left(\operatorname{sgn}\sigma \prod_{i=1}^{n} a_{i\sigma(i)} \right) \tag{0.3.2.1}$$

in which the sum is over all $n!$ permutations of $\{1, \ldots, n\}$ and $\operatorname{sgn}\sigma$, the "sign" or "signum" of a permutation σ, is $+1$ or -1 according to whether the minimum number of transpositions (pairwise interchanges) necessary to achieve it starting from $\{1, \ldots, n\}$ is even or odd. We say that a permutation σ is *even* if $\operatorname{sgn}\sigma = +1$; σ is *odd* if $\operatorname{sgn}\sigma = -1$.

If $\operatorname{sgn}\sigma$ in (0.3.2.1) is replaced by certain other functions of σ, one obtains *generalized matrix functions* in place of $\det A$. For example, the *permanent* of A, denoted by $\operatorname{per} A$, is obtained by replacing $\operatorname{sgn}\sigma$ by the function that is identically $+1$.

0.3.3 Elementary row and column operations. Three simple and fundamental operations on rows or columns, called *elementary row and column operations*, can be used to transform a matrix (square or not) into a simple form that facilitates such tasks as solving linear equations, determining rank, and calculating determinants and inverses of square matrices. We focus on *row operations*, which are implemented by matrices that act on the left. *Column operations* are defined and used in a similar fashion; the matrices that implement them act on the right.

Type 1: Interchange of two rows.

For $i \neq j$, interchange of rows i and j of A results from left multiplication of A by

$$\begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 0 & & \cdots & & 1 & & \\ & & & 1 & & & & & \\ & & \vdots & & \ddots & & \vdots & & \\ & & & & & 1 & & & \\ & & 1 & \cdots & \cdots & \cdots & 0 & & \\ & & & & & & & 1 & \\ & & & & & & & & \ddots \\ & & & & & & & & & 1 \end{bmatrix}$$

The two off-diagonal 1s are in the i, j and j, i positions, the two diagonal 0s are in positions i, i and j, j, and all unspecified entries are 0.

Type 2: Multiplication of a row by a nonzero scalar.

Multiplication of row i of A by a nonzero scalar c results from left multiplication of A by

$$\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & c & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix}$$

The i, i entry is c, all other main diagonal entries are 1, and all unspecified entries are 0.

Type 3: Addition of a scalar multiple of one row to another row.

For $i \neq j$, addition of c times row i of A to row j of A results from left multiplication of A by

$$\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & c & & \ddots & \\ & & & & & 1 \end{bmatrix}$$

The j, i entry is c, all main diagonal entries are 1, and all unspecified entries are 0. The displayed matrix illustrates the case in which $j > i$.

The matrices of each of the three elementary row (or column) operations are just the result of applying the respective operation to the identity matrix I (on the left for a row operation; on the right for a column operation). The effect of a type 1 operation on the determinant is to multiply it by -1; the effect of a type 2 operation is to multiply it by the nonzero scalar c; a type 3 operation does not change the determinant. The determinant of a square matrix with a zero row is zero. The determinant of a square matrix is zero if and only if some subset of the rows of the matrix is linearly dependent.

0.3.4 Reduced row echelon form. To each $A = [a_{ij}] \in M_{m,n}(\mathbf{F})$ there corresponds a (unique) canonical form in $M_{m,n}(\mathbf{F})$, the *reduced row echelon form*, also known as the *Hermite normal form*.

If a row of A is nonzero, its *leading entry* is its first nonzero entry. The defining specifications of the RREF are as follows:

(a) Any zero rows occur at the bottom of the matrix.
(b) The leading entry of any nonzero row is a 1.
(c) All other entries in the column of a leading entry are zero.
(d) The leading entries occur in a stairstep pattern, left to right; that is, if row i is nonzero and a_{ik} is its leading entry, then either $i = m$, or row $i + 1$ is zero, or the leading entry in row $i + 1$ is $a_{i+1,\ell}$, in which $\ell > k$.

For example,

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & \pi \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is in RREF.

If $R \in M_{m,n}(\mathbf{F})$ is the RREF of A, then $R = EA$, in which the nonsingular matrix $E \in M_m(\mathbf{F})$ is a product of type 1, type 2, and type 3 elementary matrices corresponding to the sequence of elementary row operations performed to reduce A to RREF.

The determinant of $A \in M_n(\mathbf{F})$ is nonzero if and only if its RREF is I_n. The value of $\det A$ may be calculated by recording the effects on the determinant of each of the elementary operations that lead to the RREF.

For the system of linear equations $Ax = b$, with $A \in M_{m,n}(\mathbf{F})$ and $b \in \mathbf{F}^m$ given and $x \in \mathbf{F}^n$ unknown, the set of solutions is unchanged if the same sequence of elementary row operations is performed on both A and b. The solutions of $Ax = b$ are revealed by inspection of the RREF of $[A\ b]$. Since the RREF is unique, for given $A_1, A_2 \in M_{m,n}$ and given $b_1, b_2 \in \mathbf{F}^m$, the systems of linear equations $A_1x = b_1$ and $A_2x = b_2$ have the same set of solutions if and only if $[A_1\ b_1]$ and $[A_2\ b_2]$ have the same RREF.

0.3.5 Multiplicativity. A key property of the determinant function is that it is multiplicative: For $A, B \in M_n(\mathbf{F})$

$$\det AB = \det A \det B$$

This may be proved using elementary operations that row-reduce both A and B.

0.3.6 Functional characterization of the determinant. If we think of the determinant as a function of each row (or column) of a matrix separately with the others fixed, the Laplace expansion (0.3.1.1) reveals that the determinant is a linear function of the entries in any one given row (column). We summarize this property by saying that the function $A \to \det A$ is *multilinear* in the rows (columns) of A.

The determinant function $A \to \det A$ is the unique function $f : M_n(\mathbf{F}) \to \mathbf{F}$ that is

(a) Multilinear in the rows of its argument
(b) Alternating: any type 1 operation on A changes the sign of $f(A)$
(c) Normalized: $f(I) = 1$

The permanent function is also multilinear (as are other generalized matrix functions), and it is normalized, but it is not alternating.

0.4 Rank

0.4.1 Definition. If $A \in M_{m,n}(\mathbf{F})$, $\operatorname{rank} A = \dim \operatorname{range} A$ is the length of a longest linearly independent list of columns of A. There can be more than one linearly independent list of columns whose length equals the rank. It is a remarkable fact that $\operatorname{rank} A^T = \operatorname{rank} A$. Therefore, an equivalent definition of rank is the length of a longest linearly independent list of rows of A: row rank = column rank.

0.4.2 Rank and linear systems. Let $A \in M_{m,n}(\mathbf{F})$ and $b \in \mathbf{F}^n$ be given. The linear system $Ax = b$ may have no solution, exactly one solution, or infinitely many solutions; these are the only possibilities. If there is at least one solution, the linear system is *consistent*; if there is no solution, the linear system is *inconsistent*. The linear system $Ax = b$ is consistent if and only if $\operatorname{rank}[A\ b] = \operatorname{rank} A$. The matrix $[A\ b] \in M_{m,n+1}(\mathbf{F})$ is the *augmented matrix*. To say that the augmented matrix and the *coefficient matrix* A of a linear system have the same rank is just to say that b is a linear combination of the columns of A. In this case, appending b to the columns of A does not increase the rank. A *solution* of the linear system $Ax = b$ is a vector x whose entries are the coefficients in a representation of b as a linear combination of the columns of A.

0.4.3 RREF and rank. Elementary operations do not change the rank of a matrix, and thus $\operatorname{rank} A$ is the same as the rank of the RREF of A, which is just the number of nonzero rows in the RREF. As a practical matter, however, numerical calculation of the rank by calculation of the RREF is unwise. Round-off errors in intermediate numerical calculations can make zero rows of the RREF appear to be nonzero, thereby affecting perception of the rank.

0.4.4 Characterizations of rank. The following statements about a given matrix $A \in M_{m,n}(\mathbf{F})$ are equivalent; each can be useful in a different context. Note that in (b) and (c) the key issue is linear independence of *lists* of columns or rows of a matrix:

(a) $\operatorname{rank} A = k$.
(b) k, and no more than k, rows of A are linearly independent.
(c) k, and no more than k, columns of A are linearly independent.
(d) Some k-by-k submatrix of A has nonzero determinant, and every $(k+1)$-by-$(k+1)$ submatrix of A has zero determinant.
(e) $\dim(\operatorname{range} A) = k$.

(f) There are k, but no more than k, linearly independent vectors $b_1, \ldots, b_k$ such that the linear system $Ax = b_j$ is consistent for each $j = 1, \ldots, k$.
(g) $k = n - \dim(\text{nullspace } A)$ (the *rank-nullity theorem*).
(h) $k = \min\{p : A = XY^T \text{ for some } X \in M_{m,p}(\mathbf{F}), Y \in M_{n,p}(\mathbf{F})\}$.
(i) $k = \min\{p : A = x_1 y_1^T + \cdots + x_p y_p^T\}$ for some $x_1, \ldots, x_p \in \mathbf{F}^m$, $y_1, \ldots, y_p \in \mathbf{F}^n$.

0.4.5 Rank inequalities. Some fundamental inequalities involving rank are:

(a) If $A \in M_{m,n}(\mathbf{F})$, then $\operatorname{rank} A \leq \min\{m, n\}$.
(b) If one or more rows and/or columns are deleted from a matrix, the rank of the resulting submatrix is not greater than the rank of the original matrix.
(c) Sylvester inequality: If $A \in M_{m,k}(\mathbf{F})$ and $B \in M_{k,n}(\mathbf{F})$, then

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

(d) The rank-sum inequality: If $A, B \in M_{m,n}(\mathbf{F})$, then

$$|\operatorname{rank} A - \operatorname{rank} B| \leq \operatorname{rank}(A + B) \leq \operatorname{rank} A + \operatorname{rank} B \tag{0.4.5.1}$$

with equality in the second inequality if and only if $(\text{range } A) \cap (\text{range } B) = \{0\}$ and $(\text{range } A^T) \cap (\text{range } B^T) = \{0\}$. If $\operatorname{rank} B = 1$ then

$$|\operatorname{rank}(A + B) - \operatorname{rank} A| \leq 1 \tag{0.4.5.2}$$

in particular, changing one entry of a matrix can change its rank by at most 1.
(e) Frobenius inequality: If $A \in M_{m,k}(\mathbf{F})$, $B \in M_{k,p}(\mathbf{F})$, and $C \in M_{p,n}(\mathbf{F})$, then

$$\operatorname{rank} AB + \operatorname{rank} BC \leq \operatorname{rank} B + \operatorname{rank} ABC$$

with equality if and only if there are matrices X and Y such that $B = BCX + YAB$.

0.4.6 Rank equalities. Some fundamental equalities involving rank are:

(a) If $A \in M_{m,n}(\mathbf{C})$, then $\operatorname{rank} A^* = \operatorname{rank} A^T = \operatorname{rank} \bar{A} = \operatorname{rank} A$.
(b) If $A \in M_m(\mathbf{F})$ and $C \in M_n(\mathbf{F})$ are nonsingular and $B \in M_{m,n}(\mathbf{F})$, then $\operatorname{rank} AB = \operatorname{rank} B = \operatorname{rank} BC = \operatorname{rank} ABC$; that is, left or right multiplication by a nonsingular matrix leaves rank unchanged.
(c) If $A, B \in M_{m,n}(\mathbf{F})$, then $\operatorname{rank} A = \operatorname{rank} B$ if and only if there exist a nonsingular $X \in M_m(\mathbf{F})$ and a nonsingular $Y \in M_n(\mathbf{F})$ such that $B = XAY$.
(d) If $A \in M_{m,n}(\mathbf{C})$, then $\operatorname{rank} A^*A = \operatorname{rank} A$.
(e) *Full-rank factorization*: If $A \in M_{m,n}(\mathbf{F})$, then $\operatorname{rank} A = k$ if and only if $A = XY^T$ for some $X \in M_{m,k}(\mathbf{F})$ and $Y \in M_{n,k}(\mathbf{F})$ that each have independent columns. The equivalent factorization $A = XBY^T$ for some nonsingular $B \in M_k(\mathbf{F})$ can also be useful. In particular, $\operatorname{rank} A = 1$ if and only if $A = xy^T$ for some nonzero vectors $x \in \mathbf{F}^m$ and $y \in \mathbf{F}^n$.
(f) If $A \in M_{m,n}(\mathbf{F})$, then $\operatorname{rank} A = k$ if and only if there exist nonsingular matrices $S \in M_m(\mathbf{F})$ and $T \in M_n(\mathbf{F})$ such that $A = S\begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}T$.

(g) Let $A \in M_{m,n}(\mathbf{F})$. If $X \in M_{n,k}(\mathbf{F})$ and $Y \in M_{m,k}(\mathbf{F})$, and if $W = Y^T AX$ is nonsingular, then

$$\operatorname{rank}(A - AXW^{-1}Y^T A) = \operatorname{rank} A - \operatorname{rank} AXW^{-1}Y^T A \tag{0.4.6.1}$$

When $k = 1$, this is *Wedderburn's rank-one reduction formula*: If $x \in \mathbf{F}^n$ and $y \in \mathbf{F}^m$, and if $\omega = y^T Ax \neq 0$, then

$$\operatorname{rank}\left(A - \omega^{-1}Axy^T A\right) = \operatorname{rank} A - 1 \tag{0.4.6.2}$$

Conversely, if $\sigma \in \mathbf{F}$, $u \in \mathbf{F}^n$, $v \in \mathbf{F}^m$, and $\operatorname{rank}\left(A - \sigma uv^T\right) < \operatorname{rank} A$, then $\operatorname{rank}\left(A - \sigma uv^T\right) = \operatorname{rank} A - 1$ and there are $x \in \mathbf{F}^n$ and $y \in \mathbf{F}^m$ such that $u = Ax$, $v = A^T y$, $y^T Ax \neq 0$, and $\sigma = (y^T Ax)^{-1}$.

0.5 Nonsingularity

A linear transformation or matrix is said to be *nonsingular* if it produces the output 0 only for the input 0. Otherwise, it is *singular*. If $A \in M_{m,n}(\mathbf{F})$ and $m < n$, then A is necessarily singular. An $A \in M_n(\mathbf{F})$ is *invertible* if there is a matrix $A^{-1} \in M_n(\mathbf{F})$ (the *inverse* of A) such that $A^{-1}A = I$. If $A \in M_n$ and $A^{-1}A = I$, then $AA^{-1} = I$; that is, A^{-1} is a *left inverse* if and only if it is a *right inverse*; A^{-1} is unique whenever it exists.

It is useful to be able to call on a variety of criteria for a square matrix to be nonsingular. The following are equivalent for a given $A \in M_n(\mathbf{F})$:

(a) A is nonsingular.
(b) A^{-1} exists.
(c) $\operatorname{rank} A = n$.
(d) The rows of A are linearly independent.
(e) The columns of A are linearly independent.
(f) $\det A \neq 0$.
(g) The dimension of the range of A is n.
(h) The dimension of the null space of A is 0.
(i) $Ax = b$ is consistent for each $b \in \mathbf{F}^n$.
(j) If $Ax = b$ is consistent, then the solution is unique.
(k) $Ax = b$ has a unique solution for each $b \in \mathbf{F}^n$.
(l) The only solution to $Ax = 0$ is $x = 0$.
(m) 0 is not an eigenvalue of A (see Chapter 1).

The conditions (g) and (h) are equivalent for a linear transformation $T : V \to V$ on a finite dimensional vector space V; that is, $Tx = y$ has a solution x for every $y \in V$ if and only if the only x such that $Tx = 0$ is $x = 0$ if and only if $Tx = y$ has a *unique* solution x for every $y \in V$.

The nonsingular matrices in $M_n(\mathbf{F})$ form a group, the *general linear group*, often denoted by $GL(n, \mathbf{F})$.

If $A \in M_n(\mathbf{F})$ is nonsingular, then $((A^{-1})^T A^T)^T = A(A^{-1}) = I$, so $(A^{-1})^T A^T = I$, which means that $(A^{-1})^T = (A^T)^{-1}$. It is convenient to write either $(A^{-1})^T$ or $(A^T)^{-1}$ as A^{-T}. If $A \in M_n(\mathbf{C})$ is nonsingular, then $(A^{-1})^* = (A^*)^{-1}$, and we may safely write either as A^{-*}.

0.6 The Euclidean inner product and norm

0.6.1 Definitions. The scalar $\langle x, y\rangle = y^*x$ is the *Euclidean inner product* (*standard inner product, usual inner product, scalar product, dot product*) of $x, y \in \mathbf{C}^n$. The *Euclidean norm* (*usual norm, Euclidean length*) function on $\mathbf{C}^n$ is the real-valued function $\|x\|_2 = \langle x, x\rangle^{1/2} = (x^*x)^{1/2}$; two important properties of this function are that $\|x\|_2 > 0$ for all nonzero $x \in \mathbf{C}^n$ and $\|\alpha x\|_2 = |\alpha|\,\|x\|_2$ for all $x \in \mathbf{C}^n$ and all $\alpha \in \mathbf{C}$.

The function $\langle\cdot,\cdot\rangle : \mathbf{C}^n \times \mathbf{C}^n \to \mathbf{C}$ is linear in the first argument and *conjugate linear* in the second; that is, $\langle \alpha x_1 + \beta x_2, y\rangle = \alpha\langle x_1, y\rangle + \beta\langle x_2, y\rangle$ and $\langle x, \alpha y_1 + \beta y_2\rangle = \bar{\alpha}\langle x, y_1\rangle + \bar{\beta}\langle x, y_2\rangle$ for all $\alpha, \beta \in \mathbf{C}$ and $y_1, y_2 \in \mathbf{C}^n$. If V is a real or complex vector space and $f : V \times V \to \mathbf{C}$ is a function that is linear in its first argument and conjugate linear in its second argument, we say that f is *sesquilinear* on V; f is a *semi-inner product* on V if it is sesquilinear on V and $f(x, x) \geq 0$ for every $x \in V$; f is an *inner product* on V if it is sesquilinear on V and $f(x, x) > 0$ for every nonzero $x \in V$. An *inner product space* is a pair (V, f) in which V is a real or complex vector space and f is an inner product on V.

0.6.2 Orthogonality and orthonormality. Two vectors $x, y \in \mathbf{C}^n$ are *orthogonal* if $\langle x, y\rangle = 0$. In $\mathbf{R}^2$ and $\mathbf{R}^3$, "orthogonal" has the conventional geometric interpretation of "perpendicular." A list of vectors $x_1, \ldots, x_m \in \mathbf{C}^n$ is said to be *orthogonal* if $\langle x_i, x_j\rangle = 0$ for all distinct $i, j \in \{1, \ldots, m\}$. An orthogonal list of nonzero vectors is linearly independent. A vector whose Euclidean norm is 1 is said to be *normalized* (a *unit vector*). For any nonzero $x \in \mathbf{C}^n$, $x/\|x\|_2$ is a unit vector. An orthogonal list of vectors is an *orthonormal* list if each of its elements is a unit vector. An orthonormal list of vectors is linearly independent. Each of these concepts has a straightforward generalization to the context of an inner product space.

0.6.3 The Cauchy–Schwarz inequality. The *Cauchy–Schwarz inequality* states that

$$|\langle x, y\rangle| \leq \|x\|_2\,\|y\|_2$$

for all $x, y \in \mathbf{C}^n$, with equality if and only if one of the vectors is a scalar multiple of the other. The *angle* θ between two nonzero real vectors $x, y \in \mathbf{R}^n$ is defined by

$$\cos\theta = \frac{\langle x, y\rangle}{\|x\|_2\,\|y\|_2}, \quad 0 \leq \theta \leq \pi \tag{0.6.3.1}$$

0.6.4 Gram–Schmidt orthonormalization. Any finite independent list of vectors in an inner product space may be replaced by an orthonormal list with the same span. This replacement may be carried out in many ways, but there is a systematic way to do so that has a useful special property. The *Gram–Schmidt process* starts with a list of vectors $v_1, \ldots, v_n$ and (if the given list is linearly independent) produces an orthonormal list of vectors $z_1, \ldots, z_n$ such that $\operatorname{span}\{z_1, \ldots, z_k\} = \operatorname{span}\{x_1, \ldots, x_k\}$ for each $k = 1, \ldots, n$. The vectors z_i may be calculated in turn as follows: Let $y_1 = x_1$ and normalize it: $z_1 = y_1/\|y_1\|_2$. Let $y_2 = x_2 - \langle x_2, z_1\rangle z_1$ (y_2 is orthogonal to z_1) and normalize it: $z_2 = y_2/\|y_2\|_2$. Once $z_1, \ldots, z_{k-1}$ have been determined, the vector

$$y_k = x_k - \langle x_k, z_{k-1}\rangle z_{k-1} - \langle x_k, z_{k-2}\rangle z_{k-2} - \cdots - \langle x_k, z_1\rangle z_1$$

is orthogonal to $z_1, \ldots, z_{k-1}$; normalize it: $z_k = y_k / \|y_k\|_2$. Continue until $k = n$. If we denote $Z = [z_1 \ \ldots \ z_n]$ and $X = [x_1 \ \ldots \ x_n]$, the Gram–Schmidt process gives a factorization $X = ZR$, in which the square matrix $R = [r_{ij}]$ is nonsingular and upper triangular; that is, $r_{ij} = 0$ whenever $i > j$.

If the vectors $x_1, \ldots, x_k$ are orthonormal and the vectors $x_1, \ldots, x_k, x_{k+1}, \ldots, x_n$ are linearly independent, applying the Gram–Schmidt process to the latter list produces the list $x_1, \ldots, x_k, z_{k+1}, \ldots, z_n$ of orthonormal vectors.

The Gram–Schmidt process may be applied to any finite list of vectors, independent or not. If $x_1, \ldots, x_n$ are linearly dependent, the Gram–Schmidt process produces a vector $y_k = 0$ for the least value of k for which x_k is a linear combination of $x_1, \ldots, x_{k-1}$.

0.6.5 Orthonormal bases. An *orthonormal basis* of an inner product space is a basis whose elements constitute an orthonormal list. Since any finite ordered basis may be transformed with the Gram–Schmidt process to an orthonormal basis, any finite-dimensional inner product space has an orthonormal basis, and any orthonormal list may be extended to an orthonormal basis. Such a basis is pleasant to work with, since the cross terms in inner product calculations all vanish.

0.6.6 Orthogonal complements. Given any subset $S \subset \mathbf{C}^n$, the *orthogonal complement* of S is the set $S^\perp = \{x \in \mathbf{C}^n : x^* y = 0 \text{ for all } y \in S\}$. Even if S is not a subspace, $S^\perp$ is always a subspace. We have $(S^\perp)^\perp = \operatorname{span} S$, and $(S^\perp)^\perp = S$ if S is a subspace. It is always the case that $\dim S^\perp + \dim(S^\perp)^\perp = n$. If S_1 and S_2 are subspaces, then $(S_1 + S_2)^\perp = S_1^\perp \cap S_2^\perp$.

For a given $A \in M_{m,n}$, range A is the orthogonal complement of nullspace A^*. Therefore, for a given $b \in \mathbf{C}^m$, the linear system $Ax = b$ has a solution (not necessarily unique) if and only if $b^* z = 0$ for every $z \in \mathbf{C}^m$ such that $A^* z = 0$. This equivalence is sometimes stated as the *Fredholm alternative* (*theorem of the alternative*) — exactly one of the following two statements is true: *Either* (1) $Ax = b$ has a solution *or* (2) $A^* y = 0$ has a solution such that $y^* b \neq 0$.

If $A \in M_{m,n}$ and $B \in M_{m,q}$, if $X \in M_{m,r}$ and $Y \in M_{m,s}$, and if range $X =$ nullspace A^* and range $Y =$ nullspace B^*, then we have the following companion to (0.2.7.1) and (0.2.7.2):

$$\operatorname{range} A \cap \operatorname{range} B = \operatorname{nullspace} \begin{bmatrix} X^* \\ Y^* \end{bmatrix} \tag{0.6.6.1}$$

0.7 Partitioned sets and matrices

A *partition* of a set $\mathcal{S}$ is a collection of subsets of $\mathcal{S}$ such that each element of $\mathcal{S}$ is a member of one and only one of the subsets. For example, a partition of the set $\{1, 2, \ldots, n\}$ is a collection of subsets $\alpha_1, \ldots, \alpha_t$ (called *index sets*) such that each integer between 1 and n is in one and only one of the index sets. A *sequential partition* of $\{1, 2, \ldots, n\}$ is a partition in which the index sets have the special form $\alpha_1 = \{1, \ldots, i_1\}, \alpha_2 = \{i_1 + 1, \ldots, i_2\}, \ldots, \alpha_t = \{i_{t-1} + 1, \ldots, n\}$.

A *partition* of a matrix is a decomposition of the matrix into submatrices such that each entry of the original matrix is in one and only one of the submatrices. Partitioning of matrices is often a convenient device for perception of useful structure. For example, partitioning $B = [b_1 \ \dots \ b_n] \in M_n(\mathbf{F})$ according to its columns reveals the presentation $AB = [Ab_1 \ \dots \ Ab_n]$ of the matrix product, partitioned according to the columns of AB.

0.7.1 Submatrices. Let $A \in M_{m,n}(\mathbf{F})$. For index sets $\alpha \subseteq \{1, \dots, m\}$ and $\beta \subseteq \{1, \dots, n\}$, we denote by $A[\alpha, \beta]$ the (sub)matrix of entries that lie in the rows of A indexed by α and the columns indexed by β. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} [\{1, 3\}, \{1, 2, 3\}] = \begin{bmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}$$

If $\alpha = \beta$, the submatrix $A[\alpha] = A[\alpha, \alpha]$ is a *principal submatrix* of A. An n-by-n matrix has $\binom{n}{k}$ distinct principal submatrices of size k.

For $A \in M_n(\mathbf{F})$ and $k \in \{1. \dots, n\}$, $A[\{1, \dots, k\}]$ is a *leading principal submatrix* and $A[\{k, \dots, n\}]$ is a *trailing principal submatrix.*

It is often convenient to indicate a submatrix or principal submatrix via deletion, rather than inclusion, of rows or columns. This may be accomplished by complementing the index sets. Let $\alpha^c = \{1, \dots, m\} \backslash \alpha$ and $\beta^c = \{1, \dots, n\} \backslash \beta$ denote the index sets complementary to α and β, respectively. Then $A[\alpha^c, \beta^c]$ is the submatrix obtained by *deleting* the rows indexed by α and the columns indexed by β. For example, the submatrix $A[\alpha, \varnothing^c]$ contains the rows of A indexed by α; $A[\varnothing^c, \beta]$ contains the columns of A indexed by β.

The determinant of an r-by-r submatrix of A is called a *minor*; if we wish to indicate the size of the submatrix, we call its determinant a *minor of size* r. If the r-by-r submatrix is a *principal* submatrix, then its determinant is a *principal minor* (*of size* r*; if the submatrix is a leading principal matrix, then its determinant is a leading principal minor*; if the submatrix is a trailing principal submatrix, then its determinant is a *trailing principal minor*. By convention, the empty principal minor is 1; that is, $\det A[\emptyset] = 1$.

A signed minor, such as those appearing in the Laplace expansion (0.3.1.1) $[(-1)^{i+j} \det A_{ij}]$ is called a *cofactor*; if we wish to indicate the size of the submatrix, we call its signed determinant a *cofactor of size* r.

0.7.2 Partitions, block matrices, and multiplication. If $\alpha_1, \dots, \alpha_t$ constitute a partition of $\{1, \dots, m\}$ and $\beta_1, \dots, \beta_s$ constitute a partition of $\{1, \dots, n\}$, then the matrices $A\left[\alpha_i, \beta_j\right]$ form a partition of the matrix $A \in M_{m,n}(\mathbf{F})$, $1 \le i \le t$, $1 \le j \le s$. If $A \in M_{m,n}(\mathbf{F})$ and $B \in M_{n,p}(\mathbf{F})$ are partitioned so that the two partitions of $\{1, \dots, n\}$ coincide, the two matrix partitions are said to be *conformal*. In this event,

$$(AB)\left[\alpha_i, \gamma_j\right] = \sum_{k=1}^{s} A\left[\alpha_i, \beta_k\right] B\left[\beta_k, \gamma_j\right] \tag{0.7.2.1}$$

in which the respective collections of submatrices $A[\alpha_i, \beta_k]$ and $B[\beta_k, \gamma_j]$ are conformal partitions of A and B, respectively. The left-hand side of (0.7.2.1) is a submatrix

of the product AB (calculated in the usual way), and each summand on the right-hand side is a standard matrix product. Thus, multiplication of conformally partitioned matrices mimics usual matrix multiplication. The sum of two partitioned matrices $A, B \in M_{m,n}(\mathbf{F})$ of the same size has a similarly pleasant representation if the partitions of their rows (respectively, of their columns) are the same:

$$(A+B)\left[\alpha_i, \beta_j\right] = A\left[\alpha_i, \beta_j\right] + B\left[\alpha_i, \beta_j\right]$$

If a matrix is partitioned by sequential partitions of its rows and columns, the resulting partitioned matrix is called a *block matrix*. For example, if the rows and columns of $A \in M_n(\mathbf{F})$ are partitioned by the same sequential partition $\alpha_1 = \{1, \ldots, k\}, \alpha_2 = \{k+1, \ldots, n\}$, the resulting block matrix is

$$A = \begin{bmatrix} A[\alpha_1, \alpha_1] & A[\alpha_1, \alpha_2] \\ A[\alpha_2, \alpha_1] & A[\alpha_2, \alpha_2] \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

in which the *blocks* are $A_{ij} = A[\alpha_i, \alpha_j]$. Computations with block matrices are employed throughout the book; 2-by-2 block matrices are the most important and useful.

0.7.3 The inverse of a partitioned matrix. It can be useful to know the corresponding blocks in the inverse of a partitioned nonsingular matrix A, that is, to present the inverse of a partitioned matrix in conformally partitioned form. This may be done in a variety of apparently different, but equivalent, ways — assuming that certain submatrices of $A \in M_n(\mathbf{F})$ and A^{-1} are also nonsingular. For simplicity, let A be partitioned as a 2-by-2 block matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

with $A_{ii} \in M_{n_i}(\mathbf{F})$, $i = 1, 2$, and $n_1 + n_2 = n$. A useful expression for the correspondingly partitioned presentation of A^{-1} is

$$\begin{bmatrix} \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} & A_{11}^{-1}A_{12}\left(A_{21}A_{11}^{-1}A_{12} - A_{22}\right)^{-1} \\ A_{22}^{-1}A_{21}\left(A_{12}A_{22}^{-1}A_{21} - A_{11}\right)^{-1} & \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} \end{bmatrix} \tag{0.7.3.1}$$

assuming that all the relevant inverses exist. This expression for A^{-1} may be verified by doing a partitioned multiplication by A and then simplifying. In general index set notation, we may write

$$A^{-1}[\alpha] = \left(A[\alpha] - A\left[\alpha, \alpha^c\right] A\left[\alpha^c\right]^{-1} A\left[\alpha^c, \alpha\right]\right)^{-1}$$

and

$$\begin{aligned} A^{-1}\left[\alpha, \alpha^c\right] &= A[\alpha]^{-1} A\left[\alpha, \alpha^c\right]\left(A\left[\alpha^c, \alpha\right] A[\alpha]^{-1} A\left[\alpha, \alpha^c\right] - A\left[\alpha^c\right]\right)^{-1} \\ &= \left(A\left[\alpha, \alpha^c\right] A\left[\alpha^c\right]^{-1} A\left[\alpha^c, \alpha\right] - A[\alpha]\right)^{-1} A\left[\alpha, \alpha^c\right] A\left[\alpha^c\right]^{-1} \end{aligned}$$

again assuming that the relevant inverses exist. There is an intimate relationship between these representations and the Schur complement; see (0.8.5). Notice that $A^{-1}[\alpha]$ is a submatrix of A^{-1}, while $A[\alpha]^{-1}$ is the inverse of a submatrix of A; these two objects are not, in general, the same.

0.7.4 The Sherman–Morrison–Woodbury formula. Suppose that a nonsingular matrix $A \in M_n(\mathbf{F})$ has a known inverse A^{-1} and consider $B = A + XRY$, in which X is n-by-r, Y is r-by-n, and R is r-by-r and nonsingular. If B and $R^{-1} + YA^{-1}X$ are nonsingular, then

$$B^{-1} = A^{-1} - A^{-1}X(R^{-1} + YA^{-1}X)^{-1}YA^{-1} \tag{0.7.4.1}$$

If r is much smaller than n, then R and $R^{-1} + YA^{-1}X$ may be much easier to invert than B. For example, if $x, y \in \mathbf{F}^n$ are nonzero vectors, $X = x$, $Y = y^T$, $y^T A^{-1}x \neq -1$, and $R = [1]$, then (0.7.4.1) becomes a formula for the inverse of a rank-1 adjustment to A:

$$\left(A + xy^T\right)^{-1} = A^{-1} - \left(1 + y^T A^{-1}x\right)^{-1} A^{-1}xy^T A^{-1} \tag{0.7.4.2}$$

In particular, if $B = I + xy^T$ for $x, y \in \mathbf{F}^n$ and $y^T x \neq -1$, then $B^{-1} = I - (1 + y^T x)^{-1}xy^T$.

0.7.5 Complementary nullities. Suppose that $A \in M_n(\mathbf{F})$ is nonsingular, let α and β be nonempty subsets of $\{1, \ldots, n\}$, and write $|\alpha| = r$ and $|\beta| = s$ for the cardinalities of α and β. The *law of complementary nullities* is

$$\text{nullity}\,(A\,[\alpha, \beta]) = \text{nullity}\left(A^{-1}\left[\beta^c, \alpha^c\right]\right) \tag{0.7.5.1}$$

which is equivalent to the rank identity

$$\text{rank}\,(A\,[\alpha, \beta]) = \text{rank}\left(A^{-1}\left[\beta^c, \alpha^c\right]\right) + r + s - n \tag{0.7.5.2}$$

Since we can permute rows and columns to place first the r rows indexed by α and the s columns indexed by β, it suffices to consider the presentations

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ and } A^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

in which A_{11} and B_{11}^T are r-by-s and A_{22} and B_{22}^T are $(n-r)$-by-$(n-s)$. Then (0.7.5.1) says that nullity A_{11} = nullity B_{22}.

The underlying principle here is very simple. Suppose that the nullity of A_{11} is k. If $k \geq 1$, let the columns of $X \in M_{s,k}(\mathbf{F})$ be a basis for the null space of A_{11}. Since A is nonsingular,

$$A\begin{bmatrix} X \\ 0 \end{bmatrix} = \begin{bmatrix} A_{11}X \\ A_{21}X \end{bmatrix} = \begin{bmatrix} 0 \\ A_{21}X \end{bmatrix}$$

has full rank, so $A_{21}X$ has k independent columns. But

$$\begin{bmatrix} B_{12}\,(A_{21}X) \\ B_{22}\,(A_{21}X) \end{bmatrix} = A^{-1}\begin{bmatrix} 0 \\ A_{21}X \end{bmatrix} = A^{-1}A\begin{bmatrix} X \\ 0 \end{bmatrix} = \begin{bmatrix} X \\ 0 \end{bmatrix}$$

so $B_{22}(A_{21}X) = 0$ and hence nullity $B_{22} \geq k$ = nullity A_{11}, a statement that is trivially correct if $k = 0$. A similar argument starting with B_{22} shows that nullity $A_{11} \geq$ nullity B_{22}. For a different approach, see (3.5.P13).

Of course, (0.7.5.1) also tells us that nullity A_{12} = nullity B_{12}, nullity A_{21} = nullity B_{21}, and nullity A_{22} = nullity B_{11}. If $r + s = n$, then rank A_{11} = rank B_{22} and rank A_{22} = rank B_{11}, while if $n = 2r = 2s$, then we also have rank A_{12} = rank B_{12} and

rank A_{21} = rank B_{21}. Finally, (0.7.5.2) tells us that the rank of an r-by-s submatrix of an n-by-n nonsingular matrix is at least $r + s - n$.

0.7.6 Rank in a partitioned matrix and rank-principal matrices. Partition $A \in M_n(\mathbf{F})$ as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad A_{11} \in M_r(\mathbf{F}), \mathbf{A}_{22} \in M_{n-r}(\mathbf{F})$$

If A_{11} is nonsingular, then of course rank $\begin{bmatrix} A_{11} & A_{12} \end{bmatrix} = r$ and rank $\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} = r$. Remarkably, the converse is true:

$$\text{if rank } A = \text{rank}[A_{11}\ A_{12}] = \text{rank} \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}, \text{ then } A_{11} \text{ is nonsingular} \tag{0.7.6.1}$$

This follows from (0.4.6(c)): If A_{11} is singular, then rank $A_{11} = k < r$, and there are nonsingular $S, T \in M_r(\mathbf{F})$ such that

$$SA_{11}T = \begin{bmatrix} I_k & 0 \\ 0 & 0_{r-k} \end{bmatrix}$$

Therefore,

$$\hat{A} = \begin{bmatrix} S & 0 \\ 0 & I_{n-r} \end{bmatrix} A \begin{bmatrix} T & 0 \\ 0 & I_{n-r} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} I_k & 0 \\ 0 & 0_{r-k} \end{bmatrix} & SA_{12} \\ A_{21}T & A_{22} \end{bmatrix}$$

has rank r, as do its first block row and column. Because the rth row of the first block column of $\hat{A}$ is zero, there must be some column in SA_{12} whose rth entry is not zero, which means that $\hat{A}$ has at least $r + 1$ independent columns. This contradicts rank $\hat{A}$ = rank $A = r$, so A_{11} must be nonsingular.

Let $A \in M_{m,n}(\mathbf{F})$ and suppose that rank $A = r > 0$. Let $A = XY^T$ be a full-rank factorization with $X, Y \in M_{m,r}(\mathbf{F})$; see (0.4.6c). Let $\alpha, \beta \subseteq \{1, \ldots, m\}$ and $\gamma, \delta \subseteq \{1, \ldots, n\}$ be index sets of cardinality r. Then $A[\alpha, \gamma] = X[\alpha, \varnothing^c]Y[\gamma, \varnothing^c]^T \in M_r(\mathbf{F})$, which is nonsingular whenever rank $X[\alpha, \varnothing^c]$ = rank $Y[\gamma, \varnothing^c] = r$. The multiplicativity property (0.3.5) ensures that

$$\det A[\alpha, \gamma] \det A[\beta, \delta] = \det A[\alpha, \delta] \det A[\beta, \gamma] \tag{0.7.6.2}$$

Suppose that $A \in M_n(\mathbf{F})$ and rank $A = r$. We say that A is *rank principal* if it has a nonsingular r-by-r *principal* submatrix. It follows from (0.7.6.1) that if there is some index set $\alpha \subset \{1, \ldots, n\}$ such that

$$\text{rank } A = \text{rank } A\left[\alpha, \varnothing^c\right] = \text{rank } A\left[\varnothing^c, \alpha\right] \tag{0.7.6.3}$$

(that is, if there are r linearly independent rows of A such that the corresponding r columns are linearly independent), then A is rank principal; moreover, $A[\alpha]$ is nonsingular.

If $A \in M_n(\mathbf{F})$ is symmetric or skew symmetric, or if $A \in M_n(\mathbf{C})$ is Hermitian or skew Hermitian, then rank $A[\alpha, \varnothing^c]$ = rank $A[\varnothing^c, \alpha]$ for every index set α, so A satisfies (0.7.6.3) and is therefore rank principal.

0.7.7 Commutativity, anticommutativity, and block diagonal matrices. Two matrices $A, B \in M_n(\mathbf{F})$ are said to *commute* if $AB = BA$. Commutativity is not typical, but one important instance is encountered frequently. Suppose that $\Lambda = [\Lambda_{ij}]_{i,j=1}^s \in M_n(\mathbf{F})$ is a block matrix in which $\Lambda_{ij} = 0$ if $i \neq j$; $\Lambda_{ii} = \lambda_i I_{n_i}$ for some $\lambda_i \in \mathbf{F}$ for each $i = 1, \ldots, s$; and $\lambda_i \neq \lambda_j$ if $i \neq j$. Partition $B = [B_{ij}]_{i,j=1}^s \in M_n(\mathbf{F})$ conformally with Λ. Then $\Lambda B = B\Lambda$ if and only if $\lambda_i B_{ij} = B_{ij}\lambda_j$ for each $i, j = 1, \ldots, s$, that is, $(\lambda_i - \lambda_j)B_{ij} = 0$ for each $i, j = 1, \ldots, s$. These identities are satisfied if and only if $B_{ij} = 0$ whenever $i \neq j$. Thus, Λ *commutes with* B *if and only if* B *is block diagonal conformal with* Λ; see (0.9.2).

Two matrices $A, B \in M_n(\mathbf{F})$ are said to *anticommute* if $AB = -BA$. For example, the matrices $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ anticommute.

0.7.8 The vec mapping. Partition a matrix $A \in M_{m,n}(\mathbf{F})$ according to its columns: $A = [a_1 \ \ldots \ a_n]$. The mapping vec : $M_{m,n}(\mathbf{F}) \to \mathbf{F}^{mn}$ is

$$\operatorname{vec} A = [a_1^T \ \ldots \ a_n^T]^T$$

that is, vec A is the vector obtained by stacking the columns of A, left to right. The vec operator can be a convenient tool in problems involving matrix equations.

0.8 Determinants again

Some additional facts about and identities for the determinant are useful for reference.

0.8.1 Compound matrices. Let $A \in M_{m,n}(\mathbf{F})$. Let $\alpha \subseteq \{1, \ldots, m\}$ and $\beta \subseteq \{1, \ldots, n\}$ be index sets of cardinality $r \leq \min\{m, n\}$ elements. The $\binom{m}{r}$-by-$\binom{n}{r}$ matrix whose α, β, entry is $\det A[\alpha, \beta]$ is called the rth *compound matrix* of A and is denoted by $C_r(A)$. In forming the rows and column of $C_r(A)$, we arrange index sets lexicographically, that is, $\{1, 2, 4\}$ before $\{1, 2, 5\}$ before $\{1, 3, 4\}$, and so on. For example, if

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix} \tag{0.8.1.0}$$

then $C_2(A) =$

$$\begin{bmatrix} \det\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} & \det\begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix} & \det\begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix} \\ \det\begin{bmatrix} 1 & 2 \\ 7 & 8 \end{bmatrix} & \det\begin{bmatrix} 1 & 3 \\ 7 & 10 \end{bmatrix} & \det\begin{bmatrix} 2 & 3 \\ 8 & 10 \end{bmatrix} \\ \det\begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix} & \det\begin{bmatrix} 4 & 6 \\ 7 & 10 \end{bmatrix} & \det\begin{bmatrix} 5 & 6 \\ 8 & 10 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} -3 & -6 & -3 \\ -6 & -11 & -4 \\ -3 & -2 & 2 \end{bmatrix}$$

If $A \in M_{m,k}(\mathbf{F})$, $B \in M_{k,n}(\mathbf{F})$, and $r \leq \min\{m, k, n\}$, it follows from the Cauchy–Binet formula (0.8.7) that

$$C_r(AB) = C_r(A)C_r(B) \tag{0.8.1.1}$$

which is the *multiplicativity property* of the rth compound matrix.

We *define* $C_0(A) = 1$. We have $C_1(A) = A$; if $A \in M_n(\mathbf{F})$, then $C_n(A) = \det A$.

If $A \in M_{m,k}(\mathbf{F})$ and $t \in \mathbf{F}$, then $C_r(tA) = t^r C_r(A)$
If $1 \le r \le n$, then $C_r(I_n) = I_{\binom{n}{r}} \in M_{\binom{n}{r}}$
If $A \in M_n$ is nonsingular and $1 \le r \le n$, then $C_r(A)^{-1} = C_r(A^{-1})$
If $A \in M_n$ and $1 \le r \le n$, then $\det C_r(A) = (\det A)^{\binom{n-1}{r-1}}$
If $A \in M_{m,n}(\mathbf{F})$ and $r = \operatorname{rank} A$, then $\operatorname{rank} C_r(A) = 1$
If $A \in M_{m,n}(\mathbf{F})$ and $1 \le r \le \min\{m, n\}$, then $C_r(A^T) = C_r(A)^T$
If $A \in M_{m,n}(\mathbf{C})$ and $1 \le r \le \min\{m, n\}$, then $C_r(A^*) = C_r(A)^*$

If $\Delta = [d_{ij}] \in M_n(\mathbf{F})$ is upper (respectively, lower) triangular (see (0.9.3)), then $C_r(\Delta)$ is upper (respectively, lower) triangular; its main diagonal entries are the $\binom{n}{r}$ possible products of r entries chosen from the list $d_{11}, \ldots, d_{nn}$, that is, they are the $\binom{n}{r}$ scalars $d_{i_1 i_1} \cdots d_{i_r i_r}$ such that $1 \le i_1 < \cdots < i_r \le n$, arranged lexicographically. Consequently, if $D = \operatorname{diag}(d_1, \ldots, d_n) \in M_n(\mathbf{F})$ is diagonal, then so is $C_r(D)$; its main diagonal entries are the $\binom{n}{r}$ possible products of r entries chosen from the list $d_1, \ldots, d_n$, that is, they are the $\binom{n}{r}$ scalars $d_{i_1} \cdots d_{i_r}$ such that $1 \le i_1 < \cdots < i_r \le n$, arranged lexicographically. See chapter 6 of (Fiedler, 1986) for a detailed discussion of compound matrices.

0.8.2 The adjugate and the inverse. If $A \in M_n(\mathbf{F})$ and $n \ge 2$, the transposed matrix of cofactors of A

$$\operatorname{adj} A = \left[(-1)^{i+j} \det A\left[\{j\}^c, \{i\}^c\right]\right] \tag{0.8.2.0}$$

is the *adjugate* of A; it is also called the *classical adjoint* of A. For example, $\operatorname{adj}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

A calculation using the Laplace expansion for the determinant reveals the basic property of the adjugate:

$$(\operatorname{adj} A)\, A = A\, (\operatorname{adj} A) = (\det A)\, I \tag{0.8.2.1}$$

Thus, $\operatorname{adj} A$ is nonsingular if A is nonsingular, and $\det(\operatorname{adj} A) = (\det A)^{n-1}$.

If A is nonsingular, then

$$\operatorname{adj} A = (\det A)\, A^{-1}, \quad \text{that is,} \quad A^{-1} = (\det A)^{-1} \operatorname{adj} A \tag{0.8.2.2}$$

For example, $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = (ad - bc)^{-1}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ if $ad \ne bc$. In particular, $\operatorname{adj}(A^{-1}) = A/\det A = (\operatorname{adj} A)^{-1}$.

If A is singular and $\operatorname{rank} A \le n - 2$, then every minor of A of size $n - 1$ is zero, so $\operatorname{adj} A = 0$.

If A is singular and $\operatorname{rank} A = n - 1$, then some minor of A of size $n - 1$ is nonzero, so $\operatorname{adj} A \ne 0$ and $\operatorname{rank} \operatorname{adj} A \ge 1$. Moreover, some list of $n - 1$ columns of A is linearly independent, so the identity $(\operatorname{adj} A)\, A = (\det A)\, I = 0$ ensures that the null space of $\operatorname{adj} A$ has dimension at least $n - 1$ and hence $\operatorname{rank} \operatorname{adj} A \le 1$. We conclude that $\operatorname{rank} \operatorname{adj} A = 1$. The full-rank factorization (0.4.6(e)) ensures that $\operatorname{adj} A = \alpha x y^T$ for some nonzero $\alpha \in \mathbf{F}$ and nonzero $x, y \in \mathbf{F}^n$ that are determined as follows: Compute

$$(Ax)y^T = A(\operatorname{adj} A) = 0 = (\operatorname{adj} A)A = x(y^T A)$$

and conclude that $Ax = 0$ and $y^T A = 0$, that is, x (respectively, y) is determined up to a nonzero scalar factor as a nonzero element of the one-dimensional null space of A (respectively, A^T).

The function $A \to \operatorname{adj} A$ is continuous on M_n (each entry of $\operatorname{adj} A$ is a multinomial in the entries of A) and every matrix in M_n is a limit of nonsingular matrices, so properties of the adjugate can be deduced from continuity and properties of the inverse function. For example, if $A, B \in M_n$ are nonsingular, then $\operatorname{adj}(AB) = (\det AB)(AB)^{-1} = (\det A)(\det B)B^{-1}A^{-1} = (\det B)B^{-1}(\det A)A^{-1} = (\operatorname{adj} B)(\operatorname{adj} A)$. Continuity then ensures that

$$\operatorname{adj}(AB) = (\operatorname{adj} B)(\operatorname{adj} A) \text{ for all } A, B \in M_n \tag{0.8.2.3}$$

For any $c \in \mathbf{F}$ and any $A \in M_n(\mathbf{F})$, $\operatorname{adj}(cA) = c^{n-1} \operatorname{adj} A$. In particular, $\operatorname{adj}(cI) = c^{n-1} I$ and $\operatorname{adj} 0 = 0$.

If A is nonsingular, then

$$\begin{aligned} \operatorname{adj}(\operatorname{adj} A) &= \operatorname{adj}((\det A)A^{-1}) = (\det A)^{n-1} \operatorname{adj} A^{-1} \\ &= (\det A)^{n-1} (A/\det A) = (\det A)^{n-2} A \end{aligned}$$

so continuity ensures that

$$\operatorname{adj}(\operatorname{adj} A) = (\det A)^{n-2} A \text{ for all } A \in M_n \tag{0.8.2.4}$$

If $A + B$ is nonsingular, then $A(A+B)^{-1}B = B(A+B)^{-1}A$, so continuity ensures that

$$A \operatorname{adj}(A+B) B = B \operatorname{adj}(A+B) A \text{ for all } A, B \in M_n \tag{0.8.2.5}$$

Let $A, B \in M_n$ and suppose that A commutes with B. If A is nonsingular, then $BA^{-1} = A^{-1}ABA^{-1} = A^{-1}BAA^{-1} = A^{-1}B$, so A^{-1} commutes with B. But $BA^{-1} = (\det A)^{-1} B \operatorname{adj} A$ and $A^{-1}B = (\det A)^{-1}(\operatorname{adj} A)B$, so $\operatorname{adj} A$ commutes with B. Continuity ensures that $\operatorname{adj} A$ commutes with B whenever A commutes with B, even if A is singular.

If $A = [a_{ij}]$ is upper triangular, then $\operatorname{adj} A = [b_{ij}]$ is upper triangular and each $b_{ii} = \prod_{j \neq i} a_{jj}$; if A is diagonal, then so is $\operatorname{adj} A$.

The adjugate is the transpose of the gradient of $\det A$:

$$(\operatorname{adj} A) = \left[\frac{\partial}{\partial a_{ij}} \det A \right]^T \tag{0.8.2.6}$$

If A is nonsingular, it follows from (0.8.2.6) that

$$\left[\frac{\partial}{\partial a_{ij}} \det A \right]^T = (\det A)\, A^{-1} \tag{0.8.2.7}$$

If $A \in M_n$ is nonsingular, then $\operatorname{adj} A^T = (\det A^T)A^{-T} = (\det A)A^{-T} = ((\det A)A^{-1})^T = (\operatorname{adj} A)^T$. Continuity ensures that

$$\operatorname{adj} A^T = (\operatorname{adj} A)^T \text{ for all } A \in M_n(\mathbf{F}) \tag{0.8.2.8}$$

A similar argument shows that

$$\operatorname{adj} A^* = (\operatorname{adj} A)^* \text{ for all } A \in M_n \tag{0.8.2.9}$$

Let $A = [a_1 \ \dots \ a_n] \in M_n(\mathbf{F})$ be partitioned according to its columns and let $b \in \mathbf{F}^n$. Define

$$\left(A \underset{i}{\leftarrow} b\right) = [a_1 \ \dots \ a_{i-1} \ b \ a_{i+1} \ \dots \ a_n]$$

that is, $(A \underset{i}{\leftarrow} b)$ denotes the matrix whose ith column is b and whose remaining columns coincide with those of A. Examination of the Laplace expansion (0.3.1.1) of $\det\left(A \underset{i}{\leftarrow} b\right)$ by minors along column i reveals that it is the ith entry of the vector $(\operatorname{adj} A)\, b$, that is,

$$\left[\det\left(A \underset{i}{\leftarrow} b\right)\right]_{i=1}^{n} = (\operatorname{adj} A)\, b \tag{0.8.2.10}$$

Applying this vector identity to each column of $C = [c_1 \ \dots \ c_n] \in M_n(\mathbf{F})$ gives the matrix identity

$$\left[\det\left(A \underset{i}{\leftarrow} c_j\right)\right]_{i,j=1}^{n} = (\operatorname{adj} A)\, C \tag{0.8.2.11}$$

0.8.3 Cramer's rule. Cramer's rule is a useful way to present analytically a particular entry of the solution to $Ax = b$ when $A \in M_n(\mathbf{F})$ is nonsingular. The identity

$$A\left[\det\left(A \underset{i}{\leftarrow} b\right)\right]_{i=1}^{n} = A\,(\operatorname{adj} A)\, b = (\det A)\, b$$

follows from (0.8.2.9). If $\det A \neq 0$, we obtain Cramer's rule

$$x_i = \frac{\det(A \underset{i}{\leftarrow} b)}{\det A}$$

for the ith entry x_i of the solution vector x. Cramer's rule also follows directly from multiplicativity of the determinant. The system $Ax = b$ may be rewritten as

$$A(I \underset{i}{\leftarrow} x) = A \underset{i}{\leftarrow} b$$

and taking determinants of both sides (using multiplicativity) gives

$$(\det A)\ \det(I \underset{i}{\leftarrow} x) = \det(A \underset{i}{\leftarrow} b)$$

But $\det(I \underset{i}{\leftarrow} x) = x_i$, and the formula follows.

0.8.4 Minors of the inverse. *Jacobi's identity* generalizes the adjugate formula for the inverse of a nonsingular matrix and relates the minors of A^{-1} to those of $A \in M_n(\mathbf{F})$:

$$\det A^{-1}\left[\alpha^c, \beta^c\right] = (-1)^{p(\alpha,\beta)} \frac{\det A\,[\beta, \alpha]}{\det A} \tag{0.8.4.1}$$

in which $p(\alpha, \beta) = \sum_{i\in\alpha} i + \sum_{j\in\beta} j$. Our universal convention is that $\det A[\varnothing] = 1$. For principal submatrices, Jacobi's identity assumes the simple form

$$\det A^{-1}[\alpha^c] = \frac{\det A\,[\alpha]}{\det A} \tag{0.8.4.2}$$

0.8.5 Schur complements and determinantal formulae. Let $A = [a_{ij}] \in M_n(\mathbf{F})$ be given and suppose that $\alpha \subseteq \{1, \dots, n\}$ is an index set such that $A[\alpha]$ is nonsingular. An important formula for $\det A$, based on the 2-partition of A using α and α^c, is

$$\det A = \det A\,[\alpha] \det\left(A\left[\alpha^c\right] - A\left[\alpha^c, \alpha\right] A\,[\alpha]^{-1}\, A\left[\alpha, \alpha^c\right]\right) \tag{0.8.5.1}$$

which generalizes the familiar formula for the determinant of a 2-by-2 matrix. The special matrix

$$A/A[\alpha] = A\left[\alpha^c\right] - A\left[\alpha^c, \alpha\right] A[\alpha]^{-1} A\left[\alpha, \alpha^c\right] \tag{0.8.5.2}$$

which also appears in the partitioned form for the inverse in (0.7.3.1), is called the *Schur complement of* $A[\alpha]$ *in* A. When convenient, we take $\alpha = \{1, \ldots, k\}$ and write A as a 2-by-2 block matrix $A = [A_{ij}]$ with $A_{11} = A[\alpha]$, $A_{22} = A[\alpha^c]$, $A_{12} = A[\alpha, \alpha^c]$, and $A_{21} = A[\alpha^c, \alpha]$. The formula (0.8.5.1) may be verified by computing the determinant of both sides of the identity

$$\begin{aligned} &\begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix} \end{aligned} \tag{0.8.5.3}$$

which contains a wealth of information about the Schur complement $S = [s_{ij}] = A/A_{11} = A_{22} - A_{21}A_{11}^{-1}A_{12}$:

(a) The Schur complement S arises (uniquely) in the lower right corner if linear combinations of the first k rows (respectively, columns) of A are added to the last $n-k$ rows (respectively, columns) in such a way as to produce a zero block in the lower left (respectively, upper right) corner; this is *block Gaussian elimination*, and it is (uniquely) possible because A_{11} is nonsingular. Any submatrix of A that includes A_{11} as a principal submatrix has the same determinant before and after the block eliminations that produce the block diagonal form in (0.8.5.3). Thus, for any index set $\beta = \{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n-k\}$, if we construct the shifted index set $\tilde{\beta} = \{i_1 + k, \ldots, i_m + k\}$, then $\det A[\alpha \cup \tilde{\beta}, \alpha \cup \tilde{\gamma}]$ (before) $= \det(A_{11} \oplus S[\beta, \gamma])$ (after), so

$$\det S[\beta, \gamma] = \det A\left[\alpha \cup \tilde{\beta}, \alpha \cup \tilde{\gamma}\right] / \det A[\alpha] \tag{0.8.5.4}$$

For example, if $\beta = \{i\}$ and $\gamma = \{j\}$, then with $\alpha = \{1, \ldots, k\}$, we have

$$\begin{aligned} \det S[\beta, \gamma] &= s_{ij} \\ &= \det A[\{1, \ldots k, k+i\}, \{1, \ldots, k, k+j\}] / \det A_{11} \end{aligned} \tag{0.8.5.5}$$

so all the entries of S are ratios of minors of A.

(b) $\operatorname{rank} A = \operatorname{rank} A_{11} + \operatorname{rank} S \geq \operatorname{rank} A_{11}$, and $\operatorname{rank} A = \operatorname{rank} A_{11}$ if and only if $A_{22} = A_{21}A_{11}^{-1}A_{12}$.

(c) A is nonsingular if and only if S is nonsingular, since $\det A = \det A_{11} \det S$. If A is nonsingular, then $\det S = \det A / \det A_{11}$.

Suppose that A is nonsingular. Then inverting both sides of (0.8.5.3) gives a presentation of the inverse different from that in (0.7.3.1):

$$A^{-1} = \begin{bmatrix} A_{11} + A_{11}^{-1}A_{12}S^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}S^{-1} \\ -S^{-1}A_{21}A_{11}^{-1} & S^{-1} \end{bmatrix} \tag{0.8.5.6}$$

Among other things, this tells us that $A^{-1}[\{k+1, \ldots, n\}] = S^{-1}$, so

$$\det A^{-1}[\{k+1, \ldots, n\}] = \det A_{11} / \det A \tag{0.8.5.7}$$

This is a form of Jacobi's identity (0.8.4.1). Another form results from using the adjugate to write the inverse, which gives

$$\det((\operatorname{adj} A)[\{k+1,\ldots,n\}]) = (\det A)^{n-k-1}\det A_{11} \tag{0.8.5.8}$$

When α^c consists of a single element, the Schur complement of $A[\alpha]$ in A is a scalar and (0.8.5.1) reduces to the identity

$$\det A = A\left[\alpha^c\right]\det A[\alpha] - A\left[\alpha^c,\alpha\right](\operatorname{adj} A[\alpha])\,A\left[\alpha,\alpha^c\right] \tag{0.8.5.9}$$

which is valid even if $A[\alpha]$ is singular. For example, if $\alpha = \{1,\ldots,n-1\}$, then $\alpha^c = \{n\}$ and A is presented as a *bordered matrix*

$$A = \begin{bmatrix} \tilde{A} & x \\ y^T & a \end{bmatrix}$$

with $a \in \mathbf{F}$, $x, y \in \mathbf{F}^{n-1}$, and $\tilde{A} \in M_{n-1}(\mathbf{F})$; (0.8.5.9) is the *Cauchy expansion* of the determinant of a bordered matrix

$$\det\begin{bmatrix} \tilde{A} & x \\ y^T & a \end{bmatrix} = a\det\tilde{A} - y^T\left(\operatorname{adj}\tilde{A}\right)x \tag{0.8.5.10}$$

The Cauchy expansion (0.8.5.10) involves signed minors of A of size $n-2$ (the entries of $\operatorname{adj}\tilde{A}$) and a bilinear form in the entries of a row *and* column; the Laplace expansion (0.3.1.1) involves signed minors of A of size $n-1$ and a linear form in the entries of a row *or* column. If $a \neq 0$, we can use the Schur complement of $[a]$ in A to express

$$\det\begin{bmatrix} \tilde{A} & x \\ y^T & a \end{bmatrix} = a\det(\tilde{A} - a^{-1}xy^T)$$

Equating the right-hand side of this identity to that of (0.8.5.10) and setting $a = -1$ gives *Cauchy's formula for the determinant of a rank-one perturbation*

$$\det\left(\tilde{A} + xy^T\right) = \det\tilde{A} + y^T\left(\operatorname{adj}\tilde{A}\right)x \tag{0.8.5.11}$$

The uniqueness property of the Schur complement discussed in (a) can be used to derive an identity involving a Schur complement within a Schur complement. Suppose that the nonsingular k-by-k block A_{11} is partitioned as a 2-by-2 block matrix $A_{11} = [\mathcal{A}_{ij}]$ in which the upper left ℓ-by-ℓ block $\mathcal{A}_{11}$ is nonsingular. Write $A_{21} = [\mathcal{A}_1\ \mathcal{A}_2]$, in which $\mathcal{A}_1$ is $(n-k)$-by-ℓ, and write $A_{12}^T = [\mathcal{B}_1^T\ \mathcal{B}_2^T]$, in which $\mathcal{B}_1$ is ℓ-by-$(n-k)$; this gives the refined partition

$$A = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} & \mathcal{B}_1 \\ \mathcal{A}_{21} & \mathcal{A}_{22} & \mathcal{B}_2 \\ \mathcal{A}_1 & \mathcal{A}_2 & A_{22} \end{bmatrix}$$

Now add linear combinations of the first ℓ rows of A to the next $k-\ell$ rows to reduce $\mathcal{A}_{21}$ to a zero block. The result is

$$A' = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} & \mathcal{B}_1 \\ 0 & A_{11}/\mathcal{A}_{11} & \mathcal{B}_2' \\ \mathcal{A}_1 & \mathcal{A}_2 & A_{22} \end{bmatrix}$$

in which we have identified the resulting 2,2 block of A' as the (necessarily nonsingular) Schur complement of $\mathcal{A}_{11}$ in A_{11}. Now add linear combinations of the first k rows of

A' to the last $n-k$ rows to reduce $[\ \mathcal{A}_1\ \mathcal{A}_2]$ to a zero block. The result is

$$A'' = \begin{bmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} & \mathcal{B}_1 \\ 0 & A_{11}/\mathcal{A}_{11} & \mathcal{B}_2' \\ 0 & 0 & A/A_{11} \end{bmatrix}$$

in which we have identified the resulting 3,3 block of A'' as the Schur complement of A_{11} in A. The lower right 2-by-2 block of A'' must be $A/\mathcal{A}_{11}$, the Schur complement of $\mathcal{A}_{11}$ in A. Moreover, the lower right block of $A/\mathcal{A}_{11}$ must be the Schur complement of $A_{11}/\mathcal{A}_{11}$ in $A/\mathcal{A}_{11}$. This observation is the *quotient property of Schur complements*:

$$A/A_{11} = (A/\mathcal{A}_{11})\,/\,(A_{11}/\mathcal{A}_{11}) \tag{0.8.5.12}$$

If the four blocks A_{ij} in (0.8.5.3) are square and the same size, and if A_{11} commutes with A_{21}, then

$$\begin{aligned} \det A &= \det A_{11} \det S = \det(A_{11}S) \\ &= \det(A_{11}A_{22} - A_{11}A_{21}A_{11}^{-1}A_{12}) = \det(A_{11}A_{22} - A_{21}A_{12}) \end{aligned}$$

If A_{11} commutes with A_{12}, the same conclusion follows from computing $\det A = (\det S)(\det A) = \det(SA_{11})$. By continuity, the identity

$$\det A = \det(A_{11}A_{22} - A_{21}A_{12}) \tag{0.8.5.13}$$

is valid whenever A_{11} commutes with either A_{21} or A_{12}, even if it is singular. If A_{22} commutes with either A_{12} or A_{21}, a similar argument using the Schur complement of A_{22} shows that

$$\det A = \det(A_{11}A_{22} - A_{12}A_{21}) \tag{0.8.5.14}$$

if A_{22} commutes with either A_{21} or A_{12}.

0.8.6 Determinantal identities of Sylvester and Kronecker. We consider two consequences of (0.8.5.4). If we set

$$B = \left[b_{ij}\right] = [\det A\,[\{1,\ldots k, k+i\}\,,\{1,\ldots,k,k+j\}]]_{i,j=1}^{n-k}$$

then each entry of B is the determinant of a bordered matrix of the form (0.8.5.10): $\tilde{A}$ is A_{11}, x is the jth column of A_{12}, y^T is the ith row of A_{21}, and a is the $i,\ j$ entry of A_{22}. The identity (0.8.5.5) tells us that $B = (\det A_{11})S$, so

$$\begin{aligned} \det B &= (\det A_{11})^{n-k} \det S \\ &= (\det A_{11})^{n-k}\,(\det A/\det A_{11}) = (\det A_{11})^{n-k-1} \det A \end{aligned}$$

This observation about B is *Sylvester's identity for bordered determinants*:

$$\det B = (\det A\,[\alpha])^{n-k-1} \det A \tag{0.8.6.1}$$

in which $B = [\det A[\alpha \cup \{i\}\,,\alpha \cup \{j\}]]$ and $i,\ j$ are indices *not* contained in α.

If $A_{22} = 0$, then each entry of B is the determinant of a bordered matrix of the form (0.8.5.10) with $a = 0$. In this case, the Schur complement $A/A_{11} = -A_{21}A_{11}^{-1}A_{12}$ has rank at most k, so the determinant of every $(k+1)$-by-$(k+1)$ submatrix of B is zero; this observation about B is *Kronecker's theorem for bordered determinants*.

0.8.7 The Cauchy–Binet formula. This useful formula can be remembered because of its similarity in appearance to the formula for matrix multiplication. This is no accident, since it is equivalent to multiplicativity of the compound matrix (0.8.1.1). Let $A \in M_{m,k}(\mathbf{F})$, $B \in M_{k,n}(\mathbf{F})$, and $C = AB$. Furthermore, let $1 \le r \le \min\{m, k, n\}$, and let $\alpha \subseteq \{1, \ldots, m\}$ and $\beta \subseteq \{1, \ldots, n\}$ be index sets, each of cardinality r. An expression for the α, β minor of C is

$$\det C\,[\alpha, \beta] = \sum_{\gamma} \det\, A\,[\alpha, \gamma] \det B\,[\gamma, \beta]$$

in which the sum is taken over all index sets $\gamma \subseteq \{1, \ldots, k\}$ of cardinality r.

0.8.8 Relations among minors. Let $A \in M_{m,n}(\mathbf{F})$ be given and let a fixed index set $\alpha \subseteq \{1, \ldots, m\}$ of cardinality k be given. The minors $\det\, A\,[\alpha, \omega]$, as $\omega \subseteq \{1, \ldots, n\}$ runs over *ordered* index sets of cardinality k, are not algebraically independent since there are more minors than there are distinct entries among the submatrices. Quadratic relations are known among these minors. Let $i_1, i_2, \ldots, i_k \in \{1, \ldots, n\}$ be k distinct indices, not necessarily in natural order, and let $A[\alpha; i_1, \ldots, i_k]$ denote the matrix whose rows are indicated by α and whose jth column is column i_j of $A[\alpha, \{1, \ldots, n\}]$. The difference between this and our previous notation is that columns might not occur in natural order as in $A(\{1, 3\}; 4, 2)$, whose first column has the 1, 4 and 3, 4 entries of A. We then have the relations

$$\begin{aligned}
&\det A\,[\alpha; i_1, \ldots, i_k] \det\, A\,[\alpha; j_1, \ldots, j_k] \\
&= \sum_{t=1}^{k} \det\, A\,[\alpha; i_1, \ldots, i_{s-1}, j_t, i_{s+1}, \ldots, i_k] \det\, A\,[\alpha; j_1, \ldots, j_{t-1}, i_s, j_{t+1}, \ldots, j_k]
\end{aligned}$$

for each $s = 1, \ldots, k$ and all sequences of distinct indices $i_1, \ldots, i_k \in \{1, \ldots, n\}$ and $j_1, \ldots, j_k \in \{1, \ldots, n\}$.

0.8.9 The Laplace expansion theorem. The Laplace expansion (0.3.1.1) by minors along a given row or column is included in a natural family of expressions for the determinant. Let $A \in M_n(\mathbf{F})$, let $k \in \{1, \ldots, n\}$ be given, and let $\beta \subseteq \{1, \ldots, n\}$ be any given index set of cardinality k. Then

$$\begin{aligned}
\det A &= \sum_{\alpha} (-1)^{p(\alpha,\beta)} \det A\,[\alpha, \beta] \det A\left[\alpha^c, \beta^c\right] \\
&= \sum_{\alpha} (-1)^{p(\alpha,\beta)} \det A\,[\beta, \alpha] \det A\left[\beta^c, \alpha^c\right]
\end{aligned}$$

in which the sums are over all index sets $\alpha \subseteq \{1, \ldots, n\}$ of cardinality k, and $p(\alpha, \beta) = \sum_{i\in\alpha} i + \sum_{j\in\beta} j$. Choosing $k = 1$ and $\beta = \{i\}$ or $\{j\}$ gives the expansions in (0.3.1.1).

0.8.10 Derivative of the determinant. Let $A(t) = [a_1(t)\ \ldots\ a_n(t)] = [a_{ij}(t)]$ be an n-by-n complex matrix whose entries are differentiable functions of t and define $A'(t) = [a'_{ij}(t)]$. It follows from multilinearity of the determinant (0.3.6(a)) and the

definition of the derivative that

$$\frac{d}{dt}\det A(t) = \sum_{j=1}^{n} \det\left(A(t) \underset{j}{\leftarrow} a_j'(t)\right) = \sum_{j=1}^{n}\sum_{i=1}^{n}((\operatorname{adj} A(t))^T)_{ij}a_{ij}'(t)$$
$$= \operatorname{tr}((\operatorname{adj} A(t))A'(t)) \tag{0.8.10.1}$$

For example, if $A \in M_n$ and $A(t) = tI - A$, then $A'(t) = I$ and

$$\frac{d}{dt}\det(tI - A) = \operatorname{tr}((\operatorname{adj} A(t))I) = \operatorname{tr}\operatorname{adj}(tI - A) \tag{0.8.10.2}$$

0.8.11 Dodgson's identity. Let $A \in M_n(\mathbf{F})$. Define $a = \det A[\{n\}^c]$, $b = A[\{n\}^c, \{1\}^c]$, $c = A[\{1\}^c, \{n\}^c]$, $d = \det A[\{1\}^c]$, and $e = \det A[[\{1, n\}^c]$. If $e \neq 0$, then $\det A = (ad - bc)/e$.

0.8.12 Adjugates and compounds. Let $A, B \in M_n(\mathbf{F})$. Let $\alpha \subseteq \{1, \ldots, m\}$ and $\beta \subseteq \{1, \ldots, n\}$ be index sets of cardinality $r \leq n$. The α, β entry of the *rth adjugate matrix* $\operatorname{adj}_r(A) \in M_{\binom{n}{r}}(\mathbf{F})$ is

$$(-1)^{p(\alpha,\beta)} \det A[\beta^c, \alpha^c] \tag{0.8.12.1}$$

in which $p(\alpha, \beta) = \sum_{i\in\alpha} i + \sum_{j\in\beta} j$. The rows and columns of $\operatorname{adj}_r(A)$ are formed by arranging the index sets lexicographically, just as for the rth compound matrix. For example, using the matrix A in (0.8.1.0), we have

$$\operatorname{adj}_2(A) = \begin{bmatrix} 10 & -6 & 3 \\ -8 & 5 & -2 \\ 7 & -4 & 1 \end{bmatrix}$$

The *multiplicativity property* of the rth adjunct matrix is

$$\operatorname{adj}_r(AB) = \operatorname{adj}_r(B)\operatorname{adj}_r(A) \tag{0.8.12.2}$$

We *define* $\operatorname{adj}_n(A) = 1$. We have $\operatorname{adj}_0(A) = \det A$ and $\operatorname{adj}_1(A) = A$. The rth adjugate and rth compound matrices are related by the identity

$$\operatorname{adj}_r(A)C_r(A) = C_r(A)\operatorname{adj}_r(A) = (\det A)I_{\binom{n}{r}}$$

of which the identities in (0.8.9) are special cases. In particular, $C_r(A)^{-1} = (\det A)^{-1}\operatorname{adj}_r(A)$ if A is nonsingular.

The determinant of a sum of matrices can be expressed using the rth adjugate and rth compound matrices:

$$\det(sA + tB) = \sum_{k=0}^{n} s^k t^{n-k} \operatorname{tr}(\operatorname{adj}_k(A)C_r(B)) \tag{0.8.12.3}$$

In particular, $\det(A + I) = \sum_{k=0}^{n} \operatorname{tr}\operatorname{adj}_k(A) = \sum_{k=0}^{n} \operatorname{tr} C_r(B)$.

0.9 Special types of matrices

Certain matrices of special form arise frequently and have important properties. Some of these are cataloged here for reference and terminology.

0.9.1 Diagonal matrices. A matrix $D = [d_{ij}] \in M_{n,m}(\mathbf{F})$ is *diagonal* if $d_{ij} = 0$ whenever $j \neq i$. If all the diagonal entries of a diagonal matrix are positive (nonnegative) real numbers, we refer to it as a *positive* (*nonnegative*) *diagonal matrix*. The term *positive diagonal matrix* means that the matrix is diagonal and has positive diagonal entries; it does not refer to a general matrix with positive diagonal entries. The identity matrix $I \in M_n$ is a positive diagonal matrix. A square diagonal matrix D is a *scalar matrix* if its diagonal entries are all equal, that is, $D = \alpha I$ for some $\alpha \in \mathbf{F}$. Left or right multiplication of a matrix by a scalar matrix has the same effect as multiplying it by the corresponding scalar.

If $A = [a_{ij}] \in M_{n,m}(\mathbf{F})$ and $q = \min\{m, n\}$, then $\operatorname{diag} A = [a_{11}, \ldots, a_{qq}]^T \in \mathbf{F}^q$ denotes the vector of diagonal entries of A (0.2.1). Conversely, if $x \in \mathbf{F}^q$ and if m and n are positive integers such that $\min\{m, n\} = q$, then $\operatorname{diag} x \in M_{n,m}(\mathbf{F})$ denotes the n-by-m diagonal matrix A such that $\operatorname{diag} A = x$; for $\operatorname{diag} x$ to be well-defined, both m and n must be specified. For any $a_1, \ldots, a_n \in \mathbf{F}$, $\operatorname{diag}(a_1, \ldots, a_n)$ always denotes the matrix $A = [a_{ij}] \in M_n(\mathbf{F})$ such that $a_{ii} = a_i$ for each $i = 1, \ldots, n$ and $a_{ij} = 0$ if $i \neq j$.

Suppose that $D = [d_{ij}]$, $E = [e_{ij}] \in M_n(\mathbf{F})$ are diagonal and let $A = [a_{ij}] \in M_n(\mathbf{F})$ be given. Then (a) $\det D = \prod_{i=1}^n d_{ii}$; (b) D is nonsingular if and only if all $d_{ii} \neq 0$; (c) *left* multiplication of A by D multiplies the *rows* of A by the diagonal entries of D (the ith row of DA is d_{ii} times the ith row of A); (d) *right* multiplication of A by D multiplies the *columns* of A by the diagonal entries of D, that is, the jth column of AD is d_{jj} times the jth column of A; (e) $DA = AD$ if and only if $a_{ij} = 0$ whenever $d_{ii} \neq d_{jj}$; (f) if all the diagonal entries of D are distinct and $DA = AD$, then A is diagonal; (g) for any positive integer k, $D^k = \operatorname{diag}(d_{11}^k, \ldots, d_{nn}^k)$; and (h) any two diagonal matrices D and E of the same size commute: $DE = \operatorname{diag}(d_{11}e_{11}, \ldots, d_{nn}e_{nn}) = ED$.

0.9.2 Block diagonal matrices and direct sums. A matrix $A \in M_n(\mathbf{F})$ of the form

$$A = \begin{bmatrix} A_{11} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & A_{kk} \end{bmatrix}$$

in which $A_{ii} \in M_{n_i}(\mathbf{F})$, $i = 1, \ldots, k$, $\sum_{i=1}^k n_i = n$, and all blocks above and below the block diagonal are zero blocks, is called *block diagonal*. It is convenient to write such a matrix as

$$A = A_{11} \oplus A_{22} \oplus \cdots \oplus A_{kk} = \bigoplus_{i=1}^k A_{ii}$$

This is the *direct sum* of the matrices $A_{11}, \ldots, A_{kk}$. Many properties of block diagonal matrices generalize those of diagonal matrices. For example, $\det(\oplus_{i=1}^k A_{ii}) = \prod_{i=1}^k \det A_{ii}$, so that $A = \oplus A_{ii}$ is nonsingular if and only if each A_{ii} is nonsingular,

$i = 1, \ldots, k$. Furthermore, two direct sums $A = \oplus_{i=1}^{k} A_{ii}$ and $B = \oplus_{i=1}^{k} B_{ii}$, in which each A_{ii} is the same size as B_{ii}, commute if and only if each pair A_{ii} and B_{ii} commutes, $i = 1, \ldots, k$. Also, $\operatorname{rank}(\oplus_{i=1}^{k} A_{ii}) = \sum_{i=1}^{k} \operatorname{rank} A_{ii}$.

If $A \in M_n$ and $B \in M_m$ are nonsingular, then $(A \oplus B)^{-1} = A^{-1} \oplus B^{-1}$ and $(\det(A \oplus B))(A \oplus B)^{-1} = (\det A)(\det B)(A^{-1} \oplus B^{-1}) = ((\det B)(\det A)A^{-1} \oplus (\det A)(\det B)B^{-1})$, so a continuity argument ensures that

$$\operatorname{adj}(A \oplus B) = (\det B) \operatorname{adj} A \oplus (\det A) \operatorname{adj} B \tag{0.9.2.1}$$

0.9.3 Triangular matrices. A matrix $T = [t_{ij}] \in M_{n,m}(\mathbf{F})$ is *upper triangular* if $t_{ij} = 0$ whenever $i > j$. If $t_{ij} = 0$ whenever $i \geq j$, then T is said to be *strictly upper triangular.* Analogously, T is *lower triangular* (or *strictly lower triangular*) if its transpose is upper triangular (or strictly upper triangular). A *triangular* matrix is either lower or upper triangular; a *strictly triangular matrix* is either strictly upper triangular or strictly lower triangular. A *unit triangular matrix* is a triangular matrix (upper or lower) that has ones on its main diagonal. Sometimes the terms *right* (in place of *upper*) and *left* (in place of *lower*) are used to describe triangular matrices.

Let $T \in M_{n,m}(\mathbf{F})$ be given. If T is upper triangular, then $T = [R \;\; T_2]$ if $n \leq m$, whereas $T = \begin{bmatrix} R \\ 0 \end{bmatrix}$ if $n \geq m$; $R \in M_{\min\{n,m\}}(\mathbf{F})$ is upper triangular and T_2 is arbitrary (empty if $n = m$). If T is lower triangular, then $T = [L \;\; 0]$ if $n \leq m$, whereas $T = \begin{bmatrix} L \\ T_2 \end{bmatrix}$ if $n \geq m$; $L \in M_{\min\{n,m\}}(\mathbf{F})$ is lower triangular and T_2 is arbitrary (empty if $n = m$).

A square triangular matrix shares with a square diagonal matrix the property that its determinant is the product of its diagonal entries. Square triangular matrices need not commute with other square triangular matrices of the same size. However, if $T \in M_n$ is triangular, has distinct diagonal entries, and commutes with $B \in M_n$, then B must be triangular of the same type as T (2.4.5.1).

For each $i = 1, \ldots, n$, left multiplication of $A \in M_n(\mathbf{F})$ by a lower triangular matrix L ($A \to LA$) replaces the ith row of A by a linear combination of the first through ith rows of A. The result of performing a finite number of type 3 row operations on A (0.3.3) is a matrix LA, in which L is a unit lower triangular matrix. Corresponding statements may be made about column operations and right multiplication by an upper triangular matrix.

The rank of a triangular matrix is at least, and can be greater than, the number of nonzero entries on the main diagonal. If a square triangular matrix is nonsingular, its inverse is a triangular matrix of the same type. A product of square triangular matrices of the same size and type is a triangular matrix of the same type; each i, i diagonal entry of such a matrix product is the product of the i, i entries of the factors.

0.9.4 Block triangular matrices. A matrix $A \in M_n(\mathbf{F})$ of the form

$$A = \begin{bmatrix} A_{11} & \bigstar & \bigstar \\ & \ddots & \bigstar \\ \mathbf{0} & & A_{kk} \end{bmatrix} \tag{0.9.4.1}$$

in which $A_{ii} \in M_{n_i}(\mathbf{F})$, $i = 1, \ldots, k$, $\Sigma_{i=1}^{k} n_i = n$, and all blocks below the block diagonal are zero, is *block upper triangular*; it is strictly block upper triangular if, in

addition, all the diagonal blocks are zero blocks. A matrix is *block lower triangular* if its transpose is block upper triangular; it is *strictly block lower triangular* if its transpose is strictly block upper triangular. We say that a matrix is *block triangular* if it is either block lower triangular or block upper triangular; a matrix is both block lower triangular and block upper triangular if and only if it is block diagonal.

A block upper triangular matrix in which all the diagonal blocks are 1-by-1 or 2-by-2 is said to be *upper quasitriangular*. A matrix is *lower quasitriangular* if its transpose is upper quasitriangular; it is *quasitriangular* if it is either upper quasitriangular or lower quasitriangular. A matrix that is both upper quasitriangular and lower quasitriangular is said to be *quasidiagonal*.

Consider the square block triangular matrix A in (0.9.4.1). We have $\det A = \det A_{11} \cdots \det A_{kk}$ and $\operatorname{rank} A \geq \operatorname{rank} A_{11} + \cdots + \operatorname{rank} A_{kk}$. If A is nonsingular (that is, if A_{ii} is nonsingular for all $i = 1, \ldots, k$), then A^{-1} is a block triangular matrix partitioned conformally to A whose diagonal blocks are $A_{11}^{-1}, \ldots, A_{kk}^{-1}$.

If $A \in M_n(\mathbf{F})$ is upper triangular, then $[A[\alpha_i, \alpha_j]]_{i,j=1}^t$ is block upper triangular for *any* sequential partition $\alpha_1, \ldots, \alpha_t$ of $\{1, \ldots, n\}$ (0.7.2).

0.9.5 Permutation matrices. A square matrix P is a *permutation matrix* if exactly one entry in each row and column is equal to 1 and all other entries are 0. Multiplication by such matrices effects a permutation of the rows or columns of the matrix multiplied. For example,

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

illustrates how a permutation matrix produces a permutation of the rows (entries) of a vector: it sends the first entry to the second position, sends the second entry to the first position, and leaves the third entry in the third position. Left multiplication of a matrix $A \in M_{m,n}$ by an m-by-m permutation matrix P permutes the rows of A, while right multiplication of A by an n-by-n permutation matrix P permutes the columns of A. The matrix that carries out a type 1 elementary operation (0.3.3) is an example of a special type of permutation matrix called a *transposition*. Any permutation matrix is a product of transpositions.

The determinant of a permutation matrix is ± 1, so permutation matrices are nonsingular. Although permutation matrices need not commute, the product of two permutation matrices is again a permutation matrix. Since the identity is a permutation matrix and $P^T = P^{-1}$ for every permutation matrix P, the set of n-by-n permutation matrices is a subgroup of $GL(n, C)$ with cardinality $n!$.

Since right multiplication by $P^T = P^{-1}$ permutes columns in the same way that left multiplication by P permutes rows, the transformation $A \to PAP^T$ permutes the rows and columns (and hence also the main diagonal entries) of $A \in M_n$ in the same way. In the context of linear equations with coefficient matrix A, this transformation amounts to renumbering the variables and the equations in the same way. A matrix $A \in M_n$ such that PAP^T is triangular for some permutation matrix P is called *essentially triangular*; these matrices have much in common with triangular matrices.

If $\Lambda \in M_n$ is diagonal and $P \in M_n$ is a permutation matrix, then $P\Lambda P^T$ is a diagonal matrix.

The n-by-n *reversal matrix* is the permutation matrix

$$K_n = \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix} = \left[\kappa_{ij}\right] \in M_n \tag{0.9.5.1}$$

in which $\kappa_{i,n-i+1} = 1$ for $i = 1, \ldots, n$ and all other entries are zero. The rows of $K_n A$ are the rows of A presented in reverse order; the columns of AK_n are the columns of A presented in reverse order. The reversal matrix is sometimes called the *sip matrix* (*standard involutory permutation*), the *backward identity*, or the *exchange matrix*.

For any n-by-n matrix $A = [a_{ij}]$, the entries $a_{i,n-i+1}$ for $i = 1, \ldots, n$ comprise its *counterdiagonal* (sometimes called the *secondary diagonal, backward diagonal*, *cross diagonal*, *dexter-diagonal*, or *antidiagonal*).

A *generalized permutation matrix* is a matrix of the form $G = PD$, in which $P, D \in M_n$, P is a permutation matrix, and D is a nonsingular diagonal matrix. The set of n-by-n generalized permutation matrices is a subgroup of $GL(n, C)$.

0.9.6 Circulant matrices. A matrix $A \in M_n(\mathbf{F})$ of the form

$$A = \begin{bmatrix} a_1 & a_2 & & \cdots & a_n \\ a_n & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_n & a_1 & \cdots & a_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_2 & a_3 & \cdots & a_n & a_1 \end{bmatrix} \tag{0.9.6.1}$$

is a *circulant matrix*. Each row is the previous row cycled forward one step; the entries in each row are a cyclic permutation of those in the first. The n-by-n permutation matrix

$$C_n = \begin{bmatrix} 0 & 1 & 0 & & \cdots & 0 \\ \vdots & 0 & 1 & & & \vdots \\ & & \ddots & & \ddots & 0 \\ 0 & & & & & 1 \\ 1 & 0 & & & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 0 & I_{n-1} \\ 1 & 0_{1,n-1} \end{bmatrix} \tag{0.9.6.2}$$

is the *basic circulant permutation* matrix. A matrix $A \in M_n(\mathbf{F})$ can be written in the form

$$A = \sum_{k=0}^{n-1} a_{k+1} C_n^k \tag{0.9.6.3}$$

(a polynomial in the matrix C_n) if and only if it is a circulant. We have $C_n^0 = I = C_n^n$, and the coefficients $a_1, \ldots, a_n$ are the entries of the first row of A. This representation reveals that the circulant matrices of size n are a commutative algebra: linear combinations and products of circulants are circulants; the inverse of a nonsingular circulant is a circulant; any two circulants of the same size commute.

0.9.7 Toeplitz matrices. A matrix $A = \left[a_{ij}\right] \in M_{n+1}(\mathbf{F})$ of the form

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & \cdots & a_n \\ a_{-1} & a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_{-2} & a_{-1} & a_0 & a_1 & \cdots & a_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & a_1 \\ a_{-n} & a_{-n+1} & \cdots & \cdots & a_{-1} & a_0 \end{bmatrix}$$

is a *Toeplitz matrix*. The entry a_{ij} is equal to a_{j-i} for some given sequence $a_{-n}, a_{-n+1}, \ldots, a_{-1}, a_0, a_1, a_2, \ldots, a_{n-1}, a_n \in \mathbf{C}$. The entries of A are constant down the diagonals parallel to the main diagonal. The Toeplitz matrices

$$B = \begin{bmatrix} 0 & 1 & & \mathbf{0} \\ & 0 & \ddots & \\ & & \ddots & 1 \\ \mathbf{0} & & & 0 \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} 0 & & & \mathbf{0} \\ 1 & 0 & & \\ & \ddots & \ddots & \\ \mathbf{0} & & 1 & 0 \end{bmatrix}$$

are called the *backward shift* and *forward shift* because of their effect on the elements of the standard basis $\{e_1, \ldots, e_{n+1}\}$. Moreover, $F = B^T$ and $B = F^T$. A matrix $A \in M_{n+1}$ can be written in the form

$$A = \sum_{k=1}^{n} a_{-k} F^k + \sum_{k=0}^{n} a_k B^k \tag{0.9.7.1}$$

if and only if it is a Toeplitz matrix. Toeplitz matrices arise naturally in problems involving trigonometric moments.

Using a reversal matrix K of appropriate size (0.9.5.1), notice that the forward and backward shift matrices are related: $F = KBK = B^T$ and $B = KFK = F^T$. The representation (0.9.7.1) ensures that $KA = A^T K$ for any Toeplitz matrix A, that is, $A^T = KAK = KAK^{-1}$.

An upper triangular Toeplitz matrix $A \in M_{n+1}(\mathbf{F})$ can be represented as a polynomial in B:

$$A = a_0 I + a_1 B + \cdots + a_n B^n$$

This representation (and the fact that $B^{n+1} = 0$) makes it clear why the upper triangular Toeplitz matrices of size n are a commutative algebra: Linear combinations and products of upper triangular Toeplitz matrices are upper triangular Toeplitz matrices; A is nonsingular if and only if $a_0 \neq 0$, in which case $A^{-1} = b_0 I + b_1 B + \cdots + b_n B^n$ is also an upper triangular Toeplitz matrix with $b_0 = a_0^{-1}$ and $b_k = a_0^{-1}(\sum_{m=0}^{k-1} a_{k-m} b_m)$ for $k = 1, \ldots, n$. Any two upper triangular Toeplitz matrices of the same size commute.

0.9.8 Hankel matrices. A matrix $A \in M_{n+1}(\mathbf{F})$ of the form

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_n \\ a_1 & a_2 & \cdots & \iddots & a_{n+1} \\ a_2 & & \iddots & & \vdots \\ \vdots & a_n & & & a_{2n-1} \\ a_n & a_{n+1} & \cdots & a_{2n-1} & a_{2n} \end{bmatrix}$$

is a *Hankel matrix* . Each entry a_{ij} is equal to a_{i+j-2} for some given sequence $a_0, a_1, a_2, \ldots, a_{2n-1}, a_{2n}$. The entries of A are constant along the diagonals perpendicular to the main diagonal. Hankel matrices arise naturally in problems involving power moments. Using a reversal matrix K of appropriate size (0.9.5.1), notice that KA and AK are Hankel matrices for any Toeplitz matrix A; KH and HK are Toeplitz matrices for any Hankel matrix H. Since $K = K^T = K^{-1}$ and Hankel matrices are symmetric, this means that any Toeplitz matrix is a product of two symmetric matrices with special structure: a reversal matrix and a Hankel matrix.

0.9.9 Hessenberg matrices. A matrix $A = [a_{ij}] \in M_n(\mathbf{F})$ is said to be in *upper Hessenberg form* or to be an *upper Hessenberg matrix* if $a_{ij} = 0$ for all $i > j + 1$:

$$A = \begin{bmatrix} a_{11} & & & & \bigstar \\ a_{21} & a_{22} & & & \\ & a_{32} & \ddots & & \\ & & \ddots & \ddots & \\ 0 & & & a_{n,n-1} & a_{nn} \end{bmatrix}$$

An upper Hessenberg matrix A is said to be *unreduced* if all its subdiagonal entries are nonzero, that is, if $a_{i+1,i} \neq 0$ for all $i = 1, \ldots, n-1$; the rank of such a matrix is at least $n-1$ since its first $n-1$ columns are independent.

Let $A \in M_n(\mathbf{F})$ be unreduced upper Hessenberg. Then $A - \lambda I$ is unreduced upper Hessenberg for all $\lambda \in \mathbf{F}$, so $\operatorname{rank}(A - \lambda I) \geq n - 1$ for all $\lambda \in \mathbf{F}$.

A matrix $A \in M_n(\mathbf{F})$ is *lower Hessenberg* if A^T is upper Hessenberg.

0.9.10 Tridiagonal, bidiagonal, and other structured matrices. A matrix $A = [a_{ij}] \in M_n(\mathbf{F})$ that is *both* upper and lower Hessenberg is called *tridiagonal*, that is, A is tridiagonal if $a_{ij} = 0$ whenever $|i - j| > 1$:

$$A = \begin{bmatrix} a_1 & b_1 & & 0 \\ c_1 & a_2 & \ddots & \\ & \ddots & \ddots & b_{n-1} \\ 0 & & c_{n-1} & a_n \end{bmatrix} \tag{0.9.10.1}$$

The determinant of A can be calculated inductively starting with $\det A_1 = a_1$, $\det A_2 = a_1 a_2 - b_1 c_1$, and then computing a sequence of 2-by-2 matrix products

$$\begin{bmatrix} \det A_{k+1} & 0 \\ \det A_k & 0 \end{bmatrix} = \begin{bmatrix} a_{k+1} & -b_k c_k \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \det A_k & 0 \\ \det A_{k-1} & 0 \end{bmatrix}, \quad k = 2, \ldots, n-1$$

A *Jacobi matrix* is a real symmetric tridiagonal matrix with positive subdiagonal entries.

An *upper bidiagonal matrix* $A \in M_n(\mathbf{F})$ is a tridiagonal matrix (0.9.10.1) in which $c_1 = \cdots = c_{n-1} = 0$. A matrix $A \in M_n(\mathbf{F})$ is *lower bidiagonal* if A^T is upper bidiagonal.

A *block tridiagonal* or *block bidiagonal* matrix has a block structure like the pattern in (0.9.10.1); the diagonal blocks are square and the sizes of the superdiagonal and subdiagonal blocks are determined by the sizes of their nearest diagonal blocks.

A matrix $A = [a_{ij}] \in M_n(\mathbf{F})$ is *persymmetric* if $a_{ij} = a_{n+1-j,n+1-i}$ for all $i, j = 1, \ldots, n$; that is, a persymmetric matrix is symmetric with respect to the counterdiagonal. An alternative, and very useful, characterization is that A is persymmetric if $K_n A = A^T K_n$, in which K_n is the reversal matrix (0.9.5.1). If A is persymmetric and invertible, then A^{-1} is also persymmetric since $K_n A^{-1} = (AK_n)^{-1} = (K_n A^T)^{-1} = A^{-T} K_n$. Toeplitz matrices are persymmetric. We say that $A \in M_n(\mathbf{F})$ is *skew persymmetric* if $K_n A = -A^T K_n$; the inverse of a nonsingular skew-persymmetric matrix is skew persymmetric.

A complex matrix $A \in M_n$ such that $K_n A = A^* K_n$ is *perhermitian*; A is *skew perhermitian* if $K_n A = -A^* K_n$. The inverse of a nonsingular perhermitian (respectively, skew perhermitian) matrix is perhermitian (respectively, skew perhermitian).

A matrix $A = [a_{ij}] \in M_n(\mathbf{F})$ is *centrosymmetric* if $a_{ij} = a_{n+1-i,n+1-j}$ for all $i, j = 1, \ldots, n$. Equivalently, A is centrosymmetric if $K_n A = AK_n$; A is *skew centrosymmetric* if $K_n A = -AK_n$. A centrosymmetric matrix is symmetric about its geometric center, as illustrated by the example

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 6 & 7 & 8 & 9 \\ -1 & -2 & -3 & -2 & -1 \\ 9 & 8 & 7 & 6 & 0 \\ 5 & 4 & 3 & 2 & 1 \end{bmatrix}$$

If A is nonsingular and centrosymmetric (respectively, skew centrosymmetric), then A^{-1} is also centrosymmetric (respectively, skew centrosymmetric) since $KA^{-1} = (AK_n)^{-1} = (K_n A)^{-1} = A^{-1} K_n$. If A and B are centrosymmetric, then AB is centrosymmetric since $K_n AB = AK_n B = ABK_n$. If A and B are skew centrosymmetric, then AB is centrosymmetric.

A centrosymmetric matrix $A \in M_n(\mathbf{F})$ has a special block structure. If $n = 2m$, then

$$A = \begin{bmatrix} B & K_m C K_m \\ C & K_m B K_m \end{bmatrix}, \quad B, C \in M_m(\mathbf{F}) \tag{0.9.10.2}$$

If $n = 2m + 1$, then

$$A = \begin{bmatrix} B & K_m y & K_m C K_m \\ x^T & \alpha & x^T K_m \\ C & y & K_m B K_m \end{bmatrix}, \quad B, C \in M_m(\mathbf{F}),\ x, y \in \mathbf{F}^m,\ \alpha \in \mathbf{F} \tag{0.9.10.3}$$

A complex matrix $A \in M_n$ such that $K_n A = \bar{A} K_n$ is *centrohermitian*; it is *skew centrohermitian* if $K_n A = -\bar{A} K_n$. The inverse of a nonsingular centrohermitian

(respectively, skew centrohermitian) matrix is centrohermitian (respectively, skew centrohermitian). A product of centrohermitian matrices is centrohermitian.

0.9.11 Vandermonde matrices and Lagrange interpolation. A *Vandermonde matrix* $A \in M_n(\mathbf{F})$ has the form

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix} \tag{0.9.11.1}$$

in which $x_1, \ldots, x_n \in \mathbf{F}$; that is, $A = [a_{ij}]$ with $a_{ij} = x_i^{j-1}$. It is a fact that

$$\det A = \prod_{\substack{i,j=1 \\ i>j}}^{n} (x_i - x_j) \tag{0.9.11.2}$$

so a Vandermonde matrix is nonsingular if and only if the parameters $x_1, \ldots, x_n$ are distinct.

If $x_1, \ldots, x_n$ are distinct, the entries of the inverse $A^{-1} = [\alpha_{ij}]$ of the Vandermonde matrix (0.9.11.1) are

$$\alpha_{ij} = (-1)^{i-1} \frac{S_{n-i}(x_1, \ldots, \hat{x}_j, \ldots, x_n)}{\prod_{k \neq j} (x_k - x_j)}, \quad i, j = 1, \ldots, n$$

in which $S_0 = 1$, and if $m > 0$, then $S_m(x_1, \ldots, \hat{x}_j, \ldots, x_n)$ is the mth elementary symmetric function of the $n-1$ variables x_k, $k = 1, \ldots, n$, $k \neq j$; see (1.2.14).

The Vandermonde matrix arises in the *interpolation problem* of finding a polynomial $p(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1 x + a_0$ of degree at most $n-1$ with coefficients from $\mathbf{F}$ such that

$$\begin{aligned} p(x_1) &= a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_{n-1} x_1^{n-1} = y_1 \\ p(x_2) &= a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_{n-1} x_2^{n-1} = y_2 \\ \vdots \quad \vdots & \qquad\qquad \vdots \qquad\qquad \vdots \; \vdots \\ p(x_n) &= a_0 + a_1 x_n + a_2 x_n^2 + \cdots + a_{n-1} x_n^{n-1} = y_n \end{aligned} \tag{0.9.11.3}$$

in which $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ are given elements of $\mathbf{F}$. The interpolation conditions (0.9.11.3) are a system of n equations for the n unknown coefficients $a_0, \ldots, a_{n-1}$, and they have the form $Aa = y$, in which $a = [a_0 \ \ldots \ a_{n-1}]^T \in \mathbf{F}^n$, $y = [y_1 \ \ldots \ y_n]^T \in \mathbf{F}^n$, and $A \in M_n(\mathbf{F})$ is the Vandermonde matrix (0.9.11.1). This interpolation problem always has a solution if the points $x_1, x_2, \ldots, x_n$ are distinct, since A is nonsingular in this event.

If the points $x_1, \ldots, x_n$ are distinct, the coefficients of the interpolating polynomial could in principle be obtained by solving the system (0.9.11.3), but it is usually more useful to represent the interpolating polynomial $p(x)$ as a linear combination of the

Lagrange interpolating polynomials

$$L_i(x) = \frac{\prod_{j \neq i}(x - x_j)}{\prod_{j \neq i}(x_i - x_j)}, \qquad i = 1, \ldots, n$$

Each polynomial $L_i(x)$ has degree $n - 1$ and has the property that $L_i(x_k) = 0$ if $k \neq i$, but $L_i(x_i) = 1$. *Lagrange's interpolation formula*

$$p(x) = y_1 L_1(x) + \cdots + y_n L_n(x) \tag{0.9.11.4}$$

provides a polynomial of degree at most $n - 1$ that satisfies the equations (0.9.11.3).

0.9.12 Cauchy matrices. A *Cauchy matrix* $A \in M_n(\mathbf{F})$ is a matrix of the form $A = [(a_i + b_j)^{-1}]_{i,j=1}^n$, in which $a_1, \ldots, a_n, b_1, \ldots, b_n$ are scalars such that $a_i + b_j \neq 0$ for all $i, j = 1, \ldots, n$. It is a fact that

$$\det A = \frac{\prod_{1 \leq i < j \leq n}(a_j - a_i)(b_j - b_i)}{\prod_{1 \leq i \leq j \leq n}(a_i + b_j)} \tag{0.9.12.1}$$

so A is nonsingular if and only if $a_i \neq a_j$ and $b_i \neq b_j$ for all $i \neq j$. A *Hilbert matrix* $H_n = [(i + j - 1)^{-1}]_{i,j=1}^n$ is a Cauchy matrix that is also a Hankel matrix. It is a fact that

$$\det H_n = \frac{(1!2! \cdots (n-1)!)^4}{1!2! \cdots (2n-1)!} \tag{0.9.12.2}$$

so a Hilbert matrix is always nonsingular. The entries of its inverse $H_n^{-1} = [h_{ij}]_{i,j=1}^n$ are

$$h_{ij} = \frac{(-1)^{i+j}(n + i - 1)!(n + j - 1)!}{((i-1)!(j-1)!)^2(n-i)!(n-j)!(i+j-1)} \tag{0.9.12.3}$$

0.9.13 Involution, nilpotent, projection, coninvolution. A matrix $A \in M_n(\mathbf{F})$ is

- an *involution* if $A^2 = I$, that is, if $A = A^{-1}$ (the term *involutory* is also used)
- *nilpotent* if $A^k = 0$ for some positive integer k; the least such k is the *index of nilpotence* of A
- a *projection* if $A^2 = A$ (the term *idempotent* is also used)

Now suppose that $\mathbf{F} = \mathbf{C}$. A matrix $A \in M_n$ is

- a *Hermitian projection* if $A^* = A$ and $A^2 = A$ (the term *orthogonal projection* is also used; see (4.1.P19))
- a *coninvolution* if $A\bar{A} = I$, that is, if $\bar{A} = A^{-1}$ (the term *coninvolutory* is also used)

0.10 Change of basis

Let V be an n-dimensional vector space over the field $\mathbf{F}$, and let the list $\mathcal{B}_1 = v_1, v_2, \ldots, v_n$ be a basis for V. Any vector $x \in V$ can be represented as $x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$ because $\mathcal{B}_1$ spans V. If there were some other representation of $x = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n$ in the same basis, then

$$0 = x - x = (\alpha_1 - \beta_1)v_1 + (\alpha_2 - \beta_2)v_2 + \cdots + (\alpha_n - \beta_n)v_n$$

from which it follows that all $\alpha_i - \beta_i = 0$ because the list $\mathcal{B}_1$ is independent. Given the basis $\mathcal{B}_1$, the linear mapping

$$x \to [x]_{\mathcal{B}_1} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \qquad \text{in which } x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

from V to $\mathbf{F}^n$ is well-defined, one-to-one, and onto. The scalars α_i are the *coordinates* of x with respect to the basis $\mathcal{B}_1$, and the column vector $[x]_{\mathcal{B}_1}$ is the unique $\mathcal{B}_1$-*coordinate representation of* x.

Let $T : V \to V$ be a given linear transformation. The action of T on any $x \in V$ is determined once one knows the n vectors $Tv_1, Tv_2, \ldots, Tv_n$, because any $x \in V$ has a unique representation $x = \alpha_1 v_1 + \cdots + \alpha_n v_n$ and $Tx = T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = T(\alpha_1 v_1) + \cdots + T(\alpha_n v_n) = \alpha_1 T v_1 + \cdots + \alpha_n T v_n$ by linearity. Thus, the value of Tx is determined once $[x]_{\mathcal{B}_1}$ is known.

Let $\mathcal{B}_2 = \{w_1, w_2, \ldots, w_n\}$ also be a basis for V (either different from or the same as $\mathcal{B}_1$) and suppose that the $\mathcal{B}_2$-coordinate representation of Tv_j is

$$\left[T v_j\right]_{\mathcal{B}_2} = \begin{bmatrix} t_{1j} \\ \vdots \\ t_{nj} \end{bmatrix}, \qquad j = 1, 2, \ldots, n$$

Then, for any $x \in V$, we have

$$\begin{aligned} [Tx]_{\mathcal{B}_2} &= \left[\sum_{j=1}^{n} \alpha_j T v_j\right]_{\mathcal{B}_2} = \sum_{j=1}^{n} \alpha_j \left[T v_j\right]_{\mathcal{B}_2} \\ &= \sum_{j=1}^{n} \alpha_j \begin{bmatrix} t_{1j} \\ \vdots \\ t_{nj} \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \end{aligned}$$

The n-by-n array $[t_{ij}]$ depends on T and on the choice of the bases $\mathcal{B}_1$ and $\mathcal{B}_2$, but it does not depend on x. We define the $\mathcal{B}_1$-$\mathcal{B}_2$ *basis representation of* T to be

$$_{\mathcal{B}_2}[T]_{\mathcal{B}_1} = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nn} \end{bmatrix} = \left[[T v_1]_{\mathcal{B}_2} \ldots [T v_n]_{\mathcal{B}_2}\right]$$

We have just shown that $[Tx]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[T]_{\mathcal{B}_1}[x]_{\mathcal{B}_1}$ for any $x \in V$. In the important special case $\mathcal{B}_2 = \mathcal{B}_1$, we have ${}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}$, which is called the $\mathcal{B}_1$ *basis representation of* T.

Consider the identity linear transformation $I : V \to V$ defined by $Ix = x$ for all x. Then

$$[x]_{\mathcal{B}_2} = [Ix]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}[x]_{\mathcal{B}_1} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}[Ix]_{\mathcal{B}_1} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}[x]_{\mathcal{B}_2}$$

for all $x \in V$. By successively choosing $x = w_1, w_2, \ldots, w_n$, this identity permits us to identify each column of ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}$ and shows that

$${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} = I_n$$

If we do the same calculation starting with $[x]_{\mathcal{B}_1} = [Ix]_{\mathcal{B}_1} = \cdots$, we find that

$${}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}\,{}_{\mathcal{B}_2}[I]_{\mathcal{B}_1} = I_n$$

Thus, every matrix of the form ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ is invertible and ${}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}$ is its inverse. Conversely, every invertible matrix $S = [s_1\ s_2\ \ldots\ s_n] \in M_n(\mathbf{F})$ has the form ${}_{\mathcal{B}_1}[I]_{\mathcal{B}}$ for some basis $\mathcal{B}$. We may take $\mathcal{B}$ to be the vectors $\{\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n\}$ defined by $[\tilde{s}_i]_{\mathcal{B}_1} = s_i$, $i = 1, 2, \ldots, n$. The list $\mathcal{B}$ is independent because S is invertible.

Notice that

$${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1} = \big[[Iv_1]_{\mathcal{B}_2} \ldots [Iv_n]_{\mathcal{B}_2}\big] = \big[[v_1]_{\mathcal{B}_2} \ldots [v_n]_{\mathcal{B}_2}\big]$$

so ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ describes how the elements of the basis $\mathcal{B}_1$ are formed from elements of the basis $\mathcal{B}_2$. Now let $x \in V$ and compute

$$\begin{aligned}
{}_{\mathcal{B}_1}[T]_{\mathcal{B}_2}[x]_{\mathcal{B}_2} = [Tx]_{\mathcal{B}_2} &= [I(Tx)]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}[Tx]_{\mathcal{B}_1} \\
&= {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}[x]_{\mathcal{B}_1} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}[Ix]_{\mathcal{B}_1} \\
&= {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}[x]_{\mathcal{B}_2}
\end{aligned}$$

By choosing $x = w_1, w_2, \ldots, w_n$ successively, we conclude that

$${}_{\mathcal{B}_2}[T]_{\mathcal{B}_2} = {}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[T]_{\mathcal{B}_1}\,{}_{\mathcal{B}_1}[I]_{\mathcal{B}_2} \tag{0.10.1.1}$$

This identity shows how the $\mathcal{B}_1$ basis representation of T changes if the basis is changed to $\mathcal{B}_2$. For this reason, the matrix ${}_{\mathcal{B}_2}[I]_{\mathcal{B}_1}$ is called the $\mathcal{B}_1 - \mathcal{B}_2$ *change of basis matrix*.

Any matrix $A \in M_n(\mathbf{F})$ is a basis representation of some linear transformation $T : V \to V$, for if $\mathcal{B}$ is any basis of V, we can determine Tx by $[Tx]_{\mathcal{B}} = A[x]_{\mathcal{B}}$. For this T, a computation reveals that ${}_{\mathcal{B}}[T]_{\mathcal{B}} = A$.

0.11 Equivalence relations

Let S be a given set and let Γ be a given subset of $S \times S = \{(a, b) : a \in S \text{ and } b \in S\}$. Then Γ defines a *relation* on S in the following way: We say that a is *related to* b, written $a \sim b$, if $(a, b) \in \Gamma$. A relation on S is said to be an *equivalence relation* if it is (a) *reflexive* ($a \sim a$ for every $a \in S$), (b) *symmetric* ($a \sim b$ whenever $b \sim a$), and (c) *transitive* ($a \sim c$ whenever $a \sim b$ and $b \sim c$). An equivalence relation on S gives a disjoint partition of S in a natural way: If we define the *equivalence class* of any $a \in S$ by $S_a = \{b \in S : b \sim a\}$, then $S = \cup_{a \in S} S_a$, and for each $a, b \in S$, either $S_a = S_b$ (if $a \sim b$) or $S_a \cap S_b = \varnothing$ (if $a \not\sim b$). Conversely, any disjoint partition of S can be used to define an equivalence relation on S.

The following table lists several equivalence relations that arise in matrix analysis. The factors D_1, D_2, S, T, L, and R are square and nonsingular; U and V are unitary; L is lower triangular; R is upper triangular; D_1 and D_2 are diagonal; and A and B need not be square for equivalence, unitary equivalence, triangular equivalence, or diagonal equivalence.

Equivalence Relation $\sim$	$A \sim B$
congruence	$A = SBS^T$
unitary congruence	$A = UBU^T$
*congruence	$A = SBS^*$
consimilarity	$A = SB\bar{S}^{-1}$
equivalence	$A = SBT$
unitary equivalence	$A = UBV$
diagonal equivalence	$A = D_1BD_2$
similarity	$A = SBS^{-1}$
unitary similarity	$A = UBU^*$
triangular equivalence	$A = LBR$

Whenever an interesting equivalence relation arises in matrix analysis, it can be useful to identify a set of distinguished representatives of the equivalence classes (a *canonical form* or *normal form* for the equivalence relation). Alternatively, we often want to have effective criteria (*invariants*) that can be used to decide if two given matrices belong to the same equivalence class.

Abstractly, a *canonical form for an equivalence relation* $\sim$ on a set S is a subset C of S such that $S = \cup_{a \in C} S_a$ and $S_a \cap S_b = \varnothing$ whenever $a, b \in C$ and $a \neq b$; the *canonical form of an element* $a \in S$ is the unique element $c \in C$ such that $a \in S_c$.

For a given equivalence relation in matrix analysis, it is important to make an artful and simple choice of canonical form, and one sometimes does this in more than one way to tailor the canonical form to a specific purpose. For example, the Jordan and Weyr canonical forms are different canonical forms for similarity; the Jordan canonical form works well in problems involving powers of matrices, while the Weyr canonical form works well in problems involving commutativity.

An *invariant* for an equivalence relation $\sim$ on S is a function f on S such that $f(a) = f(b)$ whenever $a \sim b$. A family of invariants $\mathcal{F}$ for an equivalence relation $\sim$ on S is said to be *complete* if $f(a) = f(b)$ for all $f \in \mathcal{F}$ if and only if $a \sim b$; a complete family of invariants is often called a *complete system of invariants*. For example, the singular values of a matrix are a complete system of invariants for unitary equivalence.