

Springer Texts in Statistics

Series Editors

Richard DeVeaux
Stephen E. Fienberg
Ingram Olkin

More information about this series at <http://www.springer.com/series/417>

James E. Gentle

Matrix Algebra

Theory, Computations and Applications
in Statistics

Second Edition

 Springer

James E. Gentle
Fairfax, VA, USA

ISSN 1431-875X ISSN 2197-4136 (electronic)
Springer Texts in Statistics
ISBN 978-3-319-64866-8 ISBN 978-3-319-64867-5 (eBook)
DOI 10.1007/978-3-319-64867-5

Library of Congress Control Number: 2017952371

1st edition: © Springer Science+Business Media, LLC 2007

2nd edition: © Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer International Publishing AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To María

Preface to the Second Edition

In this second edition, I have corrected all known typos and other errors; I have (it is hoped) clarified certain passages; I have added some additional material; and I have enhanced the Index.

I have added a few more comments about vectors and matrices with complex elements, although, as before, unless stated otherwise, all vectors and matrices in this book are assumed to have real elements. I have begun to use “ $\det(A)$ ” rather than “ $|A|$ ” to represent the determinant of A , except in a few cases. I have also expressed some derivatives as the transposes of the expressions I used formerly.

I have put more conscious emphasis on “user-friendliness” in this edition. In a book, user-friendliness is primarily a function of references, both internal and external, and of the index. As an old software designer, I’ve always thought that user-friendliness is very important. To the extent that internal references were present in the first edition, the positive feedback I received from users of that edition about the friendliness of those internal references (“I liked the fact that you said ‘equation (x.xx) on page yy,’ instead of just ‘equation (x.xx)’”) encouraged me to try to make the internal references even more useful. It’s only when you’re “eating your own dog food,” that you become aware of where details matter, and in using the first edition, I realized that the choice of entries in the Index was suboptimal. I have spent significant time in organizing it, and I hope that the user will find the Index to this edition to be very useful. I think that it has been vastly improved over the Index in the first edition.

The overall organization of chapters has been preserved, but some sections have been changed. The two chapters that have been changed most are Chaps. 3 and 12. Chapter 3, on the basics of matrices, got about 30 pages longer. It is by far the longest chapter in the book, but I just didn’t see any reasonable way to break it up. In Chap. 12 of the first edition, “Software for Numerical Linear Algebra,” I discussed four software systems or languages, C/C++, Fortran, Matlab, and R, and did not express any preference for one

over another. In this edition, although I occasionally mention various languages and systems, I now limit most of my discussion to Fortran and R.

There are many reasons for my preference for these two systems. R is oriented toward statistical applications. It is open source and freely distributed. As for Fortran versus C/C++, Python, or other programming languages, I agree with the statement by Hanson and Hopkins (2013, page ix), "... Fortran is currently the best computer language for numerical software." Many people, however, still think of Fortran as the language their elders (or they themselves) used in the 1970s. (On a personal note, Richard Hanson, who passed away recently, was a member of my team that designed the IMSL C Libraries in the mid 1980s. Not only was C much cooler than Fortran at the time, but the ANSI committee working on updating the Fortran language was so fractured by competing interests that approval of the revision was repeatedly delayed. Many numerical analysts who were not concerned with coolness turned to C because it provided dynamic storage allocation and it allowed flexible argument lists, and the Fortran constructs could not be agreed upon.)

Language preferences are personal, of course, and there is a strong "coolness factor" in choice of a language. Python is currently one of the coolest languages, but I personally don't like the language for most of the stuff I do.

Although this book has separate parts on applications in statistics and computational issues as before, statistical applications have informed the choices I made throughout the book, and computational considerations have given direction to most discussions.

I thank the readers of the first edition who informed me of errors. Two people in particular made several meaningful comments and suggestions. Clark Fitzgerald not only identified several typos, he made several broad suggestions about organization and coverage that resulted in an improved text (I think). Andreas Eckner found, in addition to typos, some gaps in my logic and also suggested better lines of reasoning at some places. (Although I don't follow an itemized "theorem-proof" format, I try to give reasons for any nonobvious statements I make.) I thank Clark and Andreas especially for their comments. Any remaining typos, omissions, gaps in logic, and so on are entirely my responsibility.

Again, I thank my wife, María, to whom this book is dedicated, for everything.

I used \TeX via $\text{\LaTeX} 2_{\epsilon}$ to write the book. I did all of the typing, programming, etc., myself, so all mistakes (mistakes!) are mine. I would appreciate receiving suggestions for improvement and notification of errors. Notes on this book, including errata, are available at

<http://mason.gmu.edu/~jgentle/books/matbk/>

Fairfax County, VA, USA
July 14, 2017

James E. Gentle

Preface to the First Edition

I began this book as an update of *Numerical Linear Algebra for Applications in Statistics*, published by Springer in 1998. There was a modest amount of new material to add, but I also wanted to supply more of the reasoning behind the facts about vectors and matrices. I had used material from that text in some courses, and I had spent a considerable amount of class time proving assertions made but not proved in that book. As I embarked on this project, the character of the book began to change markedly. In the previous book, I apologized for spending 30 pages on the theory and basic facts of linear algebra before getting on to the main interest: *numerical* linear algebra. In this book, discussion of those basic facts takes up over half of the book.

The orientation and perspective of this book remains *numerical linear algebra for applications in statistics*. Computational considerations inform the narrative. There is an emphasis on the areas of matrix analysis that are important for statisticians, and the kinds of matrices encountered in statistical applications receive special attention.

This book is divided into three parts plus a set of appendices. The three parts correspond generally to the three areas of the book's subtitle—theory, computations, and applications—although the parts are in a different order, and there is no firm separation of the topics.

Part I, consisting of Chaps. 1 through 7, covers most of the material in linear algebra needed by statisticians. (The word “matrix” in the title of this book may suggest a somewhat more limited domain than “linear algebra”; but I use the former term only because it seems to be more commonly used by statisticians and is used more or less synonymously with the latter term.)

The first four chapters cover the basics of vectors and matrices, concentrating on topics that are particularly relevant for statistical applications. In Chap. 4, it is assumed that the reader is generally familiar with the basics of partial differentiation of scalar functions. Chapters 5 through 7 begin to take on more of an applications flavor, as well as beginning to give more consideration to computational methods. Although the details of the computations

are not covered in those chapters, the topics addressed are oriented more toward computational algorithms. Chapter 5 covers methods for decomposing matrices into useful factors.

Chapter 6 addresses applications of matrices in setting up and solving linear systems, including overdetermined systems. We should not confuse statistical inference with fitting equations to data, although the latter task is a component of the former activity. In Chap. 6, we address the more mechanical aspects of the problem of fitting equations to data. Applications in statistical data analysis are discussed in Chap. 9. In those applications, we need to make statements (i.e., assumptions) about relevant probability distributions.

Chapter 7 discusses methods for extracting eigenvalues and eigenvectors. There are many important details of algorithms for eigenanalysis, but they are beyond the scope of this book. As with other chapters in Part I, Chap. 7 makes some reference to statistical applications, but it focuses on the mathematical and mechanical aspects of the problem.

Although the first part is on “theory,” the presentation is informal; neither definitions nor facts are highlighted by such words as “definition,” “theorem,” “lemma,” and so forth. It is assumed that the reader follows the natural development. Most of the facts have simple proofs, and most proofs are given naturally in the text. No “Proof” and “Q.E.D.” or “■” appear to indicate beginning and end; again, it is assumed that the reader is engaged in the development. For example, on page 341:

If A is nonsingular and symmetric, then A^{-1} is also symmetric because
 $(A^{-1})^T = (A^T)^{-1} = A^{-1}$.

The first part of that sentence could have been stated as a theorem and given a number, and the last part of the sentence could have been introduced as the proof, with reference to some previous theorem that the inverse and transposition operations can be interchanged. (This had already been shown before page 341—in an unnumbered theorem of course!)

None of the proofs are original (at least, I don’t think they are), but in most cases, I do not know the original source or even the source where I first saw them. I would guess that many go back to C. F. Gauss. Most, whether they are as old as Gauss or not, have appeared somewhere in the work of C. R. Rao. Some lengthier proofs are only given in outline, but references are given for the details. Very useful sources of details of the proofs are Harville (1997), especially for facts relating to applications in linear models, and Horn and Johnson (1991), for more general topics, especially those relating to stochastic matrices. The older books by Gantmacher (1959) provide extensive coverage and often rather novel proofs. These two volumes have been brought back into print by the American Mathematical Society.

I also sometimes make simple assumptions without stating them explicitly. For example, I may write “for all i ” when i is used as an index to a vector. I hope it is clear that “for all i ” means only “for i that correspond to indices

of the vector.” Also, my use of an expression generally implies existence. For example, if “ AB ” is used to represent a matrix product, it implies that “ A and B are conformable for the multiplication AB .” Occasionally, I remind the reader that I am taking such shortcuts.

The material in Part I, as in the entire book, was built up recursively. In the first pass, I began with some definitions and followed those with some facts that are useful in applications. In the second pass, I went back and added definitions and additional facts that led to the results stated in the first pass. The supporting material was added as close to the point where it was needed as practical and as necessary to form a logical flow. Facts motivated by additional applications were also included in the second pass. In subsequent passes, I continued to add supporting material as necessary and to address the linear algebra for additional areas of application. I sought a bare-bones presentation that gets across what I considered to be the theory necessary for most applications in the data sciences. The material chosen for inclusion is motivated by applications.

Throughout the book, some attention is given to numerical methods for computing the various quantities discussed. This is in keeping with my belief that statistical computing should be dispersed throughout the statistics curriculum and statistical literature generally. Thus, unlike in other books on matrix “theory,” I describe the “modified” Gram-Schmidt method, rather than just the “classical” GS. (I put “modified” and “classical” in quotes because, to me, GS *is* MGS. History is interesting, but in computational matters, I do not care to dwell on the methods of the past.) Also, condition numbers of matrices are introduced in the “theory” part of the book, rather than just in the “computational” part. Condition numbers also relate to fundamental properties of the model and the data.

The difference between an expression and a computing method is emphasized. For example, often we may write the solution to the linear system $Ax = b$ as $A^{-1}b$. Although this is the solution (so long as A is square and of full rank), solving the linear system does not involve computing A^{-1} . We may write $A^{-1}b$, but we know we can compute the solution without inverting the matrix.

“This is an instance of a principle that we will encounter repeatedly:
*the form of a mathematical expression and the way the expression
 should be evaluated in actual practice may be quite different.*”

(The statement in quotes appears word for word in several places in the book.)

Standard textbooks on “matrices for statistical applications” emphasize their uses in the analysis of traditional linear models. This is a large and important field in which real matrices are of interest, and the important kinds of real matrices include symmetric, positive definite, projection, and generalized inverse matrices. This area of application also motivates much of the discussion in this book. In other areas of statistics, however, there are different matrices

of interest, including similarity and dissimilarity matrices, stochastic matrices, rotation matrices, and matrices arising from graph-theoretic approaches to data analysis. These matrices have applications in clustering, data mining, stochastic processes, and graphics; therefore, I describe these matrices and their special properties. I also discuss the geometry of matrix algebra. This provides a better intuition of the operations. Homogeneous coordinates and special operations in \mathbb{R}^3 are covered because of their geometrical applications in statistical graphics.

Part II addresses selected applications in data analysis. Applications are referred to frequently in Part I, and of course, the choice of topics for coverage was motivated by applications. The difference in Part II is in its orientation.

Only “selected” applications in data analysis are addressed; there are applications of matrix algebra in almost all areas of statistics, including the theory of estimation, which is touched upon in Chap. 4 of Part I. Certain types of matrices are more common in statistics, and Chap. 8 discusses in more detail some of the important types of matrices that arise in data analysis and statistical modeling. Chapter 9 addresses selected applications in data analysis. The material of Chap. 9 has no obvious definition that could be covered in a single chapter (or a single part or even a single book), so I have chosen to discuss briefly a wide range of areas. Most of the sections and even subsections of Chap. 9 are on topics to which entire books are devoted; however, I do not believe that any single book addresses all of them.

Part III covers some of the important details of numerical computations, with an emphasis on those for linear algebra. I believe these topics constitute the most important material for an introductory course in numerical analysis for statisticians and should be covered in every such course.

Except for specific computational techniques for optimization, random number generation, and perhaps symbolic computation, Part III provides the basic material for a course in statistical computing. All statisticians should have a passing familiarity with the principles.

Chapter 10 provides some basic information on how data are stored and manipulated in a computer. Some of this material is rather tedious, but it is important to have a general understanding of computer arithmetic before considering computations for linear algebra. Some readers may skip or just skim Chap. 10, but the reader should be aware that the way the computer stores numbers and performs computations has far-reaching consequences. Computer arithmetic differs from ordinary arithmetic in many ways; for example, computer arithmetic lacks associativity of addition and multiplication, and series often converge even when they are not supposed to. (On the computer, a straightforward evaluation of $\sum_{x=1}^{\infty} x$ converges!)

I emphasize the differences between the abstract number system \mathbb{R} , called the reals, and the computer number system \mathbb{F} , the floating-point numbers unfortunately also often called “real.” Table 10.4 on page 492 summarizes some of these differences. All statisticians should be aware of the effects of these differences. I also discuss the differences between \mathbb{Z} , the abstract number

system called the integers, and the computer number system \mathbb{I} , the fixed-point numbers. (Appendix A provides definitions for this and other notation that I use.)

Chapter 10 also covers some of the fundamentals of algorithms, such as iterations, recursion, and convergence. It also discusses software development. Software issues are revisited in Chap. 12.

While Chap. 10 deals with general issues in numerical analysis, Chap. 11 addresses specific issues in numerical methods for computations in linear algebra.

Chapter 12 provides a brief introduction to software available for computations with linear systems. Some specific systems mentioned include the IMSLTM libraries for Fortran and C, Octave or MATLAB[®] (or Matlab[®]), and R or S-PLUS[®] (or S-Plus[®]). All of these systems are easy to use, and the best way to learn them is to begin using them for simple problems. I do not use any particular software system in the book, but in some exercises, and particularly in Part III, I do assume the ability to program in either Fortran or C and the availability of either R or S-Plus, Octave or Matlab, and Maple[®] or Mathematica[®]. My own preferences for software systems are Fortran and R, and occasionally, these preferences manifest themselves in the text.

Appendix A collects the notation used in this book. It is generally “standard” notation, but one thing the reader must become accustomed to is the lack of notational distinction between a vector and a scalar. All vectors are “column” vectors, although I usually write them as horizontal lists of their elements. (Whether vectors are “row” vectors or “column” vectors is generally only relevant for how we write expressions involving vector/matrix multiplication or partitions of matrices.)

I write algorithms in various ways, sometimes in a form that looks similar to Fortran or C and sometimes as a list of numbered steps. I believe all of the descriptions used are straightforward and unambiguous.

This book could serve as a basic reference either for courses in statistical computing or for courses in linear models or multivariate analysis. When the book is used as a reference, rather than looking for “definition” or “theorem,” the user should look for items set off with bullets or look for numbered equations or else should use the Index or Appendix A, beginning on page 589.

The prerequisites for this text are minimal. Obviously, some background in mathematics is necessary. Some background in statistics or data analysis and some level of scientific computer literacy are also required. References to rather advanced mathematical topics are made in a number of places in the text. To some extent, this is because many sections evolved from class notes that I developed for various courses that I have taught. All of these courses were at the graduate level in the computational and statistical sciences, but they have had wide ranges in mathematical level. I have carefully reread the sections that refer to groups, fields, measure theory, and so on and am convinced that if the reader does not know much about these topics, the material is still understandable but if the reader is familiar with these topics, the references

add to that reader's appreciation of the material. In many places, I refer to computer programming, and some of the exercises require some programming. A careful coverage of Part III requires background in numerical programming.

In regard to the use of the book as a text, most of the book evolved in one way or another for my own use in the classroom. I must quickly admit, however, that I have never used this whole book as a text for any single course. I have used Part III in the form of printed notes as the primary text for a course in the "foundations of computational science" taken by graduate students in the natural sciences (including a few statistics students, but dominated by physics students). I have provided several sections from Parts I and II in online PDF files as supplementary material for a two-semester course in mathematical statistics at the "baby measure theory" level (using Shao, 2003). Likewise, for my courses in computational statistics and statistical visualization, I have provided many sections, either as supplementary material or as the primary text, in online PDF files or printed notes. I have not taught a regular "applied statistics" course in almost 30 years, but if I did, I am sure that I would draw heavily from Parts I and II for courses in regression or multivariate analysis. If I ever taught a course in "matrices for statistics" (I don't even know if such courses exist), this book would be my primary text because I think it covers most of the things statisticians need to know about matrix theory and computations.

Some exercises are Monte Carlo studies. I do not discuss Monte Carlo methods in this text, so the reader lacking background in that area may need to consult another reference in order to work those exercises. The exercises should be considered an integral part of the book. For some exercises, the required software can be obtained from either `statlib` or `netlib` (see the bibliography). Exercises in any of the chapters, not just in Part III, may require computations or computer programming.

Penultimately, I must make some statement about the relationship of this book to some other books on similar topics. A much important statistical theory and many methods make use of matrix theory, and many statisticians have contributed to the advancement of matrix theory from its very early days. Widely used books with derivatives of the words "statistics" and "matrices/linearalgebra" in their titles include Basilevsky (1983), Graybill (1983), Harville (1997), Schott (2004), and Searle (1982). All of these are useful books. The computational orientation of this book is probably the main difference between it and these other books. Also, some of these other books only address topics of use in linear models, whereas this book also discusses matrices useful in graph theory, stochastic processes, and other areas of application. (If the applications are only in linear models, most matrices of interest are symmetric and all eigenvalues can be considered to be real.) Other differences among all of these books, of course, involve the authors' choices of secondary topics and the ordering of the presentation.

Acknowledgments

I thank John Kimmel of Springer for his encouragement and advice on this book and other books on which he has worked with me. I especially thank Ken Berk for his extensive and insightful comments on a draft of this book. I thank my student Li Li for reading through various drafts of some of the chapters and pointing out typos or making helpful suggestions. I thank the anonymous reviewers of this edition for their comments and suggestions. I also thank the many readers of my previous book on numerical linear algebra who informed me of errors and who otherwise provided comments or suggestions for improving the exposition. Whatever strengths this book may have can be attributed in large part to these people, named or otherwise. The weaknesses can only be attributed to my own ignorance or hardheadedness.

I thank my wife, María, to whom this book is dedicated, for everything.

I used $\text{T}_{\text{E}}\text{X}$ via $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X} 2_{\varepsilon}$ to write the book. I did all of the typing, programming, etc., myself, so all mistakes are mine. I would appreciate receiving suggestions for improvement and notification of errors.

Fairfax County, VA, USA
June 12, 2007

James E. Gentle

Contents

Preface to the Second Edition	vii
--	-----

Preface to the First Edition	ix
---	----

Part I Linear Algebra

1 Basic Vector/Matrix Structure and Notation	3
1.1 Vectors	4
1.2 Arrays	5
1.3 Matrices	5
1.3.1 Subvectors and Submatrices	8
1.4 Representation of Data	8
2 Vectors and Vector Spaces	11
2.1 Operations on Vectors	11
2.1.1 Linear Combinations and Linear Independence	12
2.1.2 Vector Spaces and Spaces of Vectors	13
2.1.3 Basis Sets for Vector Spaces	21
2.1.4 Inner Products	23
2.1.5 Norms	25
2.1.6 Normalized Vectors	31
2.1.7 Metrics and Distances	32
2.1.8 Orthogonal Vectors and Orthogonal Vector Spaces	33
2.1.9 The “One Vector”	34
2.2 Cartesian Coordinates and Geometrical Properties of Vectors .	35
2.2.1 Cartesian Geometry	36
2.2.2 Projections	36
2.2.3 Angles Between Vectors	37
2.2.4 Orthogonalization Transformations: Gram-Schmidt	38
2.2.5 Orthonormal Basis Sets	40

2.2.6	Approximation of Vectors	41
2.2.7	Flats, Affine Spaces, and Hyperplanes	43
2.2.8	Cones	43
2.2.9	Cross Products in \mathbb{R}^3	46
2.3	Centered Vectors and Variances and Covariances of Vectors	48
2.3.1	The Mean and Centered Vectors	48
2.3.2	The Standard Deviation, the Variance, and Scaled Vectors	49
2.3.3	Covariances and Correlations Between Vectors	50
	Exercises	52
3	Basic Properties of Matrices	55
3.1	Basic Definitions and Notation	55
3.1.1	Multiplication of a Matrix by a Scalar	56
3.1.2	Diagonal Elements: $\text{diag}(\cdot)$ and $\text{vecdiag}(\cdot)$	56
3.1.3	Diagonal, Hollow, and Diagonally Dominant Matrices	57
3.1.4	Matrices with Special Patterns of Zeroes	58
3.1.5	Matrix Shaping Operators	59
3.1.6	Partitioned Matrices	61
3.1.7	Matrix Addition	63
3.1.8	Scalar-Valued Operators on Square Matrices: The Trace	65
3.1.9	Scalar-Valued Operators on Square Matrices: The Determinant	66
3.2	Multiplication of Matrices and Multiplication of Vectors and Matrices	75
3.2.1	Matrix Multiplication (Cayley)	75
3.2.2	Multiplication of Matrices with Special Patterns	78
3.2.3	Elementary Operations on Matrices	80
3.2.4	The Trace of a Cayley Product That Is Square	88
3.2.5	The Determinant of a Cayley Product of Square Matrices	88
3.2.6	Multiplication of Matrices and Vectors	89
3.2.7	Outer Products	90
3.2.8	Bilinear and Quadratic Forms: Definiteness	91
3.2.9	Anisometric Spaces	93
3.2.10	Other Kinds of Matrix Multiplication	94
3.3	Matrix Rank and the Inverse of a Matrix	99
3.3.1	Row Rank and Column Rank	100
3.3.2	Full Rank Matrices	101
3.3.3	Rank of Elementary Operator Matrices and Matrix Products Involving Them	101
3.3.4	The Rank of Partitioned Matrices, Products of Matrices, and Sums of Matrices	102

3.3.5	Full Rank Partitioning	104
3.3.6	Full Rank Matrices and Matrix Inverses	105
3.3.7	Full Rank Factorization	109
3.3.8	Equivalent Matrices	110
3.3.9	Multiplication by Full Rank Matrices	112
3.3.10	Gramian Matrices: Products of the Form $A^T A$	115
3.3.11	A Lower Bound on the Rank of a Matrix Product	117
3.3.12	Determinants of Inverses	117
3.3.13	Inverses of Products and Sums of Nonsingular Matrices	118
3.3.14	Inverses of Matrices with Special Forms	120
3.3.15	Determining the Rank of a Matrix	121
3.4	More on Partitioned Square Matrices:	
	The Schur Complement	121
3.4.1	Inverses of Partitioned Matrices	122
3.4.2	Determinants of Partitioned Matrices	122
3.5	Linear Systems of Equations	123
3.5.1	Solutions of Linear Systems	123
3.5.2	Null Space: The Orthogonal Complement	126
3.6	Generalized Inverses	127
3.6.1	Immediate Properties of Generalized Inverses	127
3.6.2	Special Generalized Inverses: The Moore-Penrose Inverse	127
3.6.3	Generalized Inverses of Products and Sums of Matrices	130
3.6.4	Generalized Inverses of Partitioned Matrices	131
3.7	Orthogonality	131
3.7.1	Orthogonal Matrices: Definition and Simple Properties	132
3.7.2	Orthogonal and Orthonormal Columns	133
3.7.3	The Orthogonal Group	133
3.7.4	Conjugacy	134
3.8	Eigenanalysis: Canonical Factorizations	134
3.8.1	Eigenvalues and Eigenvectors Are Remarkable	135
3.8.2	Left Eigenvectors	135
3.8.3	Basic Properties of Eigenvalues and Eigenvectors	136
3.8.4	The Characteristic Polynomial	138
3.8.5	The Spectrum	141
3.8.6	Similarity Transformations	146
3.8.7	Schur Factorization	147
3.8.8	Similar Canonical Factorization: Diagonalizable Matrices	148
3.8.9	Properties of Diagonalizable Matrices	152
3.8.10	Eigenanalysis of Symmetric Matrices	153

3.8.11	Positive Definite and Nonnegative Definite Matrices . . .	159
3.8.12	Generalized Eigenvalues and Eigenvectors	160
3.8.13	Singular Values and the Singular Value Decomposition (SVD)	161
3.9	Matrix Norms	164
3.9.1	Matrix Norms Induced from Vector Norms	165
3.9.2	The Frobenius Norm—The “Usual” Norm	167
3.9.3	Other Matrix Norms	169
3.9.4	Matrix Norm Inequalities	170
3.9.5	The Spectral Radius	171
3.9.6	Convergence of a Matrix Power Series	171
3.10	Approximation of Matrices	175
3.10.1	Measures of the Difference Between Two Matrices	175
3.10.2	Best Approximation with a Matrix of Given Rank	176
	Exercises	178
4	Vector/Matrix Derivatives and Integrals	185
4.1	Functions of Vectors and Matrices	186
4.2	Basics of Differentiation	186
4.2.1	Continuity	188
4.2.2	Notation and Properties	188
4.2.3	Differentials	190
4.3	Types of Differentiation	190
4.3.1	Differentiation with Respect to a Scalar	190
4.3.2	Differentiation with Respect to a Vector	191
4.3.3	Differentiation with Respect to a Matrix	196
4.4	Optimization of Scalar-Valued Functions	198
4.4.1	Stationary Points of Functions	200
4.4.2	Newton’s Method	200
4.4.3	Least Squares	202
4.4.4	Maximum Likelihood	206
4.4.5	Optimization of Functions with Constraints	208
4.4.6	Optimization Without Differentiation	213
4.5	Integration and Expectation: Applications to Probability Distributions	214
4.5.1	Multidimensional Integrals and Integrals Involving Vectors and Matrices	215
4.5.2	Integration Combined with Other Operations	216
4.5.3	Random Variables and Probability Distributions	217
	Exercises	222
5	Matrix Transformations and Factorizations	227
5.1	Factorizations	227
5.2	Computational Methods: Direct and Iterative	228

- 5.3 Linear Geometric Transformations 229
 - 5.3.1 Invariance Properties of Linear Transformations 229
 - 5.3.2 Transformations by Orthogonal Matrices 230
 - 5.3.3 Rotations 231
 - 5.3.4 Reflections 233
 - 5.3.5 Translations: Homogeneous Coordinates 234
- 5.4 Householder Transformations (Reflections) 235
 - 5.4.1 Zeroing All Elements But One in a Vector 236
 - 5.4.2 Computational Considerations 237
- 5.5 Givens Transformations (Rotations) 238
 - 5.5.1 Zeroing One Element in a Vector 239
 - 5.5.2 Givens Rotations That Preserve Symmetry 240
 - 5.5.3 Givens Rotations to Transform to Other Values 240
 - 5.5.4 Fast Givens Rotations 241
- 5.6 Factorization of Matrices 241
- 5.7 LU and LDU Factorizations 242
 - 5.7.1 Properties: Existence 243
 - 5.7.2 Pivoting 246
 - 5.7.3 Use of Inner Products 247
 - 5.7.4 Properties: Uniqueness 247
 - 5.7.5 Properties of the LDU Factorization of a Square Matrix 248
- 5.8 QR Factorization 248
 - 5.8.1 Related Matrix Factorizations 249
 - 5.8.2 Matrices of Full Column Rank 249
 - 5.8.3 Relation to the Moore-Penrose Inverse for Matrices of Full Column Rank 250
 - 5.8.4 Nonfull Rank Matrices 251
 - 5.8.5 Relation to the Moore-Penrose Inverse 251
 - 5.8.6 Determining the Rank of a Matrix 252
 - 5.8.7 Formation of the QR Factorization 252
 - 5.8.8 Householder Reflections to Form the QR Factorization 252
 - 5.8.9 Givens Rotations to Form the QR Factorization 253
 - 5.8.10 Gram-Schmidt Transformations to Form the QR Factorization 254
- 5.9 Factorizations of Nonnegative Definite Matrices 254
 - 5.9.1 Square Roots 254
 - 5.9.2 Cholesky Factorization 255
 - 5.9.3 Factorizations of a Gramian Matrix 258
- 5.10 Approximate Matrix Factorization 259
 - 5.10.1 Nonnegative Matrix Factorization 259
 - 5.10.2 Incomplete Factorizations 260
- Exercises 261

6	Solution of Linear Systems	265
6.1	Condition of Matrices	266
6.1.1	Condition Number	267
6.1.2	Improving the Condition Number	272
6.1.3	Numerical Accuracy	273
6.2	Direct Methods for Consistent Systems	274
6.2.1	Gaussian Elimination and Matrix Factorizations	274
6.2.2	Choice of Direct Method	279
6.3	Iterative Methods for Consistent Systems	279
6.3.1	The Gauss-Seidel Method with Successive Overrelaxation	279
6.3.2	Conjugate Gradient Methods for Symmetric Positive Definite Systems	281
6.3.3	Multigrid Methods	286
6.4	Iterative Refinement	286
6.5	Updating a Solution to a Consistent System	287
6.6	Overdetermined Systems: Least Squares	289
6.6.1	Least Squares Solution of an Overdetermined System	290
6.6.2	Least Squares with a Full Rank Coefficient Matrix	292
6.6.3	Least Squares with a Coefficient Matrix Not of Full Rank	293
6.6.4	Weighted Least Squares	295
6.6.5	Updating a Least Squares Solution of an Overdetermined System	295
6.7	Other Solutions of Overdetermined Systems	296
6.7.1	Solutions that Minimize Other Norms of the Residuals	297
6.7.2	Regularized Solutions	300
6.7.3	Minimizing Orthogonal Distances	301
	Exercises	305
7	Evaluation of Eigenvalues and Eigenvectors	307
7.1	General Computational Methods	308
7.1.1	Numerical Condition of an Eigenvalue Problem	308
7.1.2	Eigenvalues from Eigenvectors and Vice Versa	310
7.1.3	Deflation	310
7.1.4	Preconditioning	312
7.1.5	Shifting	312
7.2	Power Method	313
7.2.1	Inverse Power Method	315
7.3	Jacobi Method	315
7.4	QR Method	318

7.5 Krylov Methods 321
 7.6 Generalized Eigenvalues 321
 7.7 Singular Value Decomposition 322
 Exercises 324

Part II Applications in Data Analysis

8 Special Matrices and Operations Useful in Modeling and Data Analysis 329

8.1 Data Matrices and Association Matrices 330

8.1.1 Flat Files 330

8.1.2 Graphs and Other Data Structures 331

8.1.3 Term-by-Document Matrices 338

8.1.4 Probability Distribution Models 339

8.1.5 Derived Association Matrices 340

8.2 Symmetric Matrices and Other Unitarily Diagonalizable Matrices 340

8.2.1 Some Important Properties of Symmetric Matrices 340

8.2.2 Approximation of Symmetric Matrices and an Important Inequality 341

8.2.3 Normal Matrices 345

8.3 Nonnegative Definite Matrices: Cholesky Factorization 346

8.3.1 Eigenvalues of Nonnegative Definite Matrices 347

8.3.2 The Square Root and the Cholesky Factorization 347

8.3.3 The Convex Cone of Nonnegative Definite Matrices 348

8.4 Positive Definite Matrices 348

8.4.1 Leading Principal Submatrices of Positive Definite Matrices 350

8.4.2 The Convex Cone of Positive Definite Matrices 351

8.4.3 Inequalities Involving Positive Definite Matrices 351

8.5 Idempotent and Projection Matrices 352

8.5.1 Idempotent Matrices 353

8.5.2 Projection Matrices: Symmetric Idempotent Matrices 358

8.6 Special Matrices Occurring in Data Analysis 359

8.6.1 Gramian Matrices 360

8.6.2 Projection and Smoothing Matrices 362

8.6.3 Centered Matrices and Variance-Covariance Matrices 365

8.6.4 The Generalized Variance 368

8.6.5 Similarity Matrices 370

8.6.6 Dissimilarity Matrices 371

8.7	Nonnegative and Positive Matrices	372
8.7.1	The Convex Cones of Nonnegative and Positive Matrices	373
8.7.2	Properties of Square Positive Matrices	373
8.7.3	Irreducible Square Nonnegative Matrices	375
8.7.4	Stochastic Matrices	379
8.7.5	Leslie Matrices	380
8.8	Other Matrices with Special Structures	380
8.8.1	Helmert Matrices	381
8.8.2	Vandermonde Matrices	382
8.8.3	Hadamard Matrices and Orthogonal Arrays	382
8.8.4	Toeplitz Matrices	384
8.8.5	Circulant Matrices	386
8.8.6	Fourier Matrices and the Discrete Fourier Transform	387
8.8.7	Hankel Matrices	390
8.8.8	Cauchy Matrices	391
8.8.9	Matrices Useful in Graph Theory	392
8.8.10	Z-Matrices and M-Matrices	396
	Exercises	396
9	Selected Applications in Statistics	399
9.1	Structure in Data and Statistical Data Analysis	399
9.2	Multivariate Probability Distributions	400
9.2.1	Basic Definitions and Properties	400
9.2.2	The Multivariate Normal Distribution	401
9.2.3	Derived Distributions and Cochran's Theorem	401
9.3	Linear Models	403
9.3.1	Fitting the Model	405
9.3.2	Linear Models and Least Squares	408
9.3.3	Statistical Inference	410
9.3.4	The Normal Equations and the Sweep Operator	414
9.3.5	Linear Least Squares Subject to Linear Equality Constraints	415
9.3.6	Weighted Least Squares	416
9.3.7	Updating Linear Regression Statistics	417
9.3.8	Linear Smoothing	419
9.3.9	Multivariate Linear Models	420
9.4	Principal Components	424
9.4.1	Principal Components of a Random Vector	424
9.4.2	Principal Components of Data	425
9.5	Condition of Models and Data	428
9.5.1	Ill-Conditioning in Statistical Applications	429
9.5.2	Variable Selection	429
9.5.3	Principal Components Regression	430

- 9.5.4 Shrinkage Estimation 431
- 9.5.5 Statistical Inference about the Rank of a Matrix 433
- 9.5.6 Incomplete Data 437
- 9.6 Optimal Design 440
 - 9.6.1 D-Optimal Designs 441
- 9.7 Multivariate Random Number Generation 443
 - 9.7.1 The Multivariate Normal Distribution..... 443
 - 9.7.2 Random Correlation Matrices 444
- 9.8 Stochastic Processes 445
 - 9.8.1 Markov Chains 445
 - 9.8.2 Markovian Population Models..... 448
 - 9.8.3 Autoregressive Processes 449
- Exercises 452

Part III Numerical Methods and Software

- 10 Numerical Methods** 461
 - 10.1 Digital Representation of Numeric Data 466
 - 10.1.1 The Fixed-Point Number System 466
 - 10.1.2 The Floating-Point Model for Real Numbers 468
 - 10.1.3 Language Constructs for Representing Numeric Data..... 476
 - 10.1.4 Other Variations in the Representation of Data; Portability of Data 482
 - 10.2 Computer Operations on Numeric Data 483
 - 10.2.1 Fixed-Point Operations 485
 - 10.2.2 Floating-Point Operations 485
 - 10.2.3 Language Constructs for Operations on Numeric Data..... 491
 - 10.2.4 Software Methods for Extending the Precision 493
 - 10.2.5 Exact Computations 495
 - 10.3 Numerical Algorithms and Analysis 496
 - 10.3.1 Algorithms and Programs 496
 - 10.3.2 Error in Numerical Computations 496
 - 10.3.3 Efficiency 504
 - 10.3.4 Iterations and Convergence 510
 - 10.3.5 Other Computational Techniques 513
 - Exercises 516
- 11 Numerical Linear Algebra** 523
 - 11.1 Computer Storage of Vectors and Matrices..... 523
 - 11.1.1 Storage Modes 524
 - 11.1.2 Strides 524
 - 11.1.3 Sparsity 524

11.2	General Computational Considerations for Vectors and Matrices	525
11.2.1	Relative Magnitudes of Operands	525
11.2.2	Iterative Methods	527
11.2.3	Assessing Computational Errors	528
11.3	Multiplication of Vectors and Matrices	529
11.3.1	Strassen's Algorithm	531
11.3.2	Matrix Multiplication Using MapReduce	533
11.4	Other Matrix Computations	533
11.4.1	Rank Determination	534
11.4.2	Computing the Determinant	535
11.4.3	Computing the Condition Number	535
	Exercises	537
12	Software for Numerical Linear Algebra	539
12.1	General Considerations	539
12.1.1	Software Development and Open Source Software	540
12.1.2	Collaborative Research and Version Control	541
12.1.3	Finding Software	541
12.1.4	Software Design	541
12.1.5	Software Development, Maintenance, and Testing	550
12.1.6	Reproducible Research	553
12.2	Software Libraries	555
12.2.1	BLAS	555
12.2.2	Level 2 and Level 3 BLAS, LAPACK, and Related Libraries	557
12.2.3	Libraries for High Performance Computing	559
12.2.4	The IMSL Libraries	562
12.3	General Purpose Languages	564
12.3.1	Programming Considerations	566
12.3.2	Modern Fortran	568
12.3.3	C and C++	570
12.3.4	Python	571
12.4	Interactive Systems for Array Manipulation	572
12.4.1	R	572
12.4.2	MATLAB and Octave	580
	Exercises	582

Appendices and Back Matter

Notation and Definitions	589
A.1 General Notation	589
A.2 Computer Number Systems	591
A.3 General Mathematical Functions and Operators	592
A.3.1 Special Functions	594

A.4	Linear Spaces and Matrices	595
A.4.1	Norms and Inner Products	597
A.4.2	Matrix Shaping Notation	598
A.4.3	Notation for Rows or Columns of Matrices	600
A.4.4	Notation Relating to Matrix Determinants	600
A.4.5	Matrix-Vector Differentiation	600
A.4.6	Special Vectors and Matrices	601
A.4.7	Elementary Operator Matrices	601
A.5	Models and Data	602
Solutions and Hints for Selected Exercises		603
Bibliography		619
Index		633

Author Biography

James E. Gentle, PhD, is University Professor of Computational Statistics at George Mason University. He is a Fellow of the American Statistical Association (ASA) and of the American Association for the Advancement of Science. Professor Gentle has held national offices in the ASA and has served as editor and associate editor of journals of the ASA as well as for other journals in statistics and computing. He is the author of “Random Number Generation and Monte Carlo Methods” (Springer, 2003) and “Computational Statistics” (Springer, 2009).