Preface

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 the present 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 Chapters 1 through 7, covers most of the material in linear algebra needed by statisticians. (The word "matrix" in the title of the present 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 Chapter 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 Chapter 6, we address the more mechanical aspects of the problem of fitting equations to data. Applications in statistical data analysis are discussed in Chapter 9. In those applications, we need to make statements (that is, 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, Chapter 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 270:

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

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 270 — 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 lead 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 Chapter 4 of Part I. Certain types of matrices are more common in statistics, and Chapter 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 Chapter 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 Chapter 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 Chapter 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 10.3 on page 400 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 Chapter 12.

While Chapter 10 deals with general issues in numerical analysis, Chapter 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 IMSL™ 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, beginning on page **??**, or Appendix A, beginning on page 479.

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. 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/linear-algebra" 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.