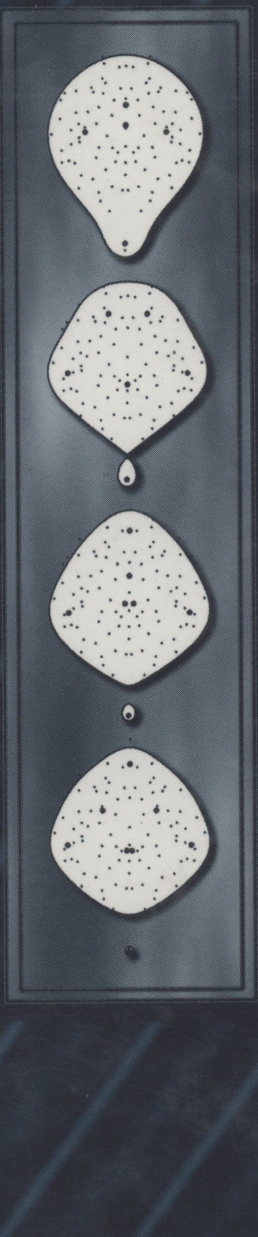

NUMERICAL LINEAR ALGEBRA

LLOYD N. TREFETHEN
DAVID BAU, III

siam

NUMERICAL LINEAR ALGEBRA

LLOYD N. TREFETHEN
Oxford University
Oxford, England

DAVID BAU, III
Microsoft Corporation
Redmond, Washington

siam
Society for Industrial and Applied Mathematics
Philadelphia

10 9 8 7 6 5 4

Library of Congress Cataloging-in-Publication Data
The Library of Congress has already catalogued the softcover edition as follows:

Trefethen, Lloyd N. (Lloyd Nicholas)
Numerical linear algebra / Lloyd N. Trefethen, David Bau III.
p. cm.
Includes bibliographical references and index.
ISBN 0-89871-361-7 (pbk.)
1. Algebras, Linear. 2. Numerical calculations. I. Bau, David.
II. Title.
QA184.T74 1997
512'.5--dc21

ISBN 0-89871-487-7 96-52458

Cover Illustration. The four curves reminiscent of water drops are polynomial lemniscates in the complex plane associated with steps 5, 6, 7, 8 of an Arnoldi iteration. The small dots are the eigenvalues of the underlying matrix A, and the large dots are the Ritz values of the Arnoldi iteration. As the iteration proceeds, the lemniscate first reaches out to engulf one of the eigenvalues λ, then pinches off and shrinks steadily to a point. The Ritz value inside it thus converges geometrically to λ. See Figure 34.3 on p. 263.

Notation

For square or rectangular matrices $A \in \mathbb{C}^{m\times n}$, $m \geq n$:

QR factorization: $A = QR$

Reduced QR factorization: $A = \hat{Q}\hat{R}$

SVD: $A = U\Sigma V^*$

Reduced SVD: $A = \hat{U}\hat{\Sigma}V^*$

For square matrices $A \in \mathbb{C}^{m\times m}$:

LU factorization: $PA = LU$

Cholesky factorization: $A = R^*R$

Eigenvalue decomposition: $A = X\Lambda X^{-1}$

Schur factorization: $A = UTU^*$

Orthogonal projector: $P = \hat{Q}\hat{Q}^*$

Householder reflector: $F = I - 2\dfrac{vv^*}{v^*v}$

QR algorithm: $A^k = \underline{Q}^{(k)}\underline{R}^{(k)}$, $A^{(k)} = (\underline{Q}^{(k)})^T A \underline{Q}^{(k)}$

Arnoldi iteration: $AQ_n = Q_{n+1}\tilde{H}_n$, $H_n = Q_n^* A Q_n$

Lanczos iteration: $AQ_n = Q_{n+1}\tilde{T}_n$, $T_n = Q_n^T A Q_n$

NUMERICAL LINEAR ALGEBRA

To our parents
Florence and Lloyd MacG. Trefethen
and
Rachel and Paul Bau

Preface

Since the early 1980s, the first author has taught a graduate course in numerical linear algebra at MIT and Cornell. The alumni of this course, now numbering in the hundreds, have been graduate students in all fields of engineering and the physical sciences. This book is an attempt to put this course on paper.

In the field of numerical linear algebra, there is already an encyclopedic treatment on the market: *Matrix Computations*, by Golub and Van Loan, now in its third edition. This book is in no way an attempt to duplicate that one. It is small, scaled to the size of one university semester. Its aim is to present fundamental ideas in as elegant a fashion as possible. We hope that every reader of this book will have access also to Golub and Van Loan for the pursuit of further details and additional topics, and for its extensive references to the research literature. Two other important recent books are those of Higham and Demmel, described in the Notes at the end (p. 329).

The field of numerical linear algebra is more beautiful, and more fundamental, than its rather dull name may suggest. More beautiful, because it is full of powerful ideas that are quite unlike those normally emphasized in a linear algebra course in a mathematics department. (At the end of the semester, students invariably comment that there is more to this subject than they ever imagined.) More fundamental, because, thanks to a trick of history, "numerical" linear algebra is really *applied* linear algebra. It is here that one finds the essential ideas that every mathematical scientist needs to work effectively with vectors and matrices. In fact, our subject is more than just vectors and matrices, for virtually everything we do carries over to functions and operators. Numerical linear algebra is really functional analysis, but with the emphasis always on practical algorithmic ideas rather than mathematical technicalities.

The book is divided into forty lectures. We have tried to build each lecture around one or two central ideas, emphasizing the unity between topics and never getting lost in details. In many places our treatment is nonstandard. This is not the place to list all of these points (see the Notes), but we will mention one unusual aspect of this book. We have departed from the customary practice by not starting with Gaussian elimination. That algorithm is atypical of numerical linear algebra, exceptionally difficult to analyze, yet at the same time tediously familiar to every student entering a course like this. Instead, we begin with the QR factorization, which is more important, less complicated, and a fresher idea to most students. The QR factorization is the thread that connects most of the algorithms of numerical linear algebra, including methods for least squares, eigenvalue, and singular value problems, as well as iterative methods for all of these and also for systems of equations. Since the 1970s, iterative methods have moved to center stage in scientific computing, and to them we devote the last part of the book.

We hope the reader will come to share our view that if any other mathematical topic is as fundamental to the mathematical sciences as calculus and differential equations, it is numerical linear algebra.

Acknowledgments

We could not have written this book without help from many people. We must begin by thanking the hundreds of graduate students at MIT (Math 335) and Cornell (CS 621) whose enthusiasm and advice over a period of ten years guided the choice of topics and the style of presentation. About seventy of these students at Cornell worked from drafts of the book itself and contributed numerous suggestions. The number of typos caught by Keith Sollers alone was astonishing.

Most of Trefethen's own graduate students during the period of writing read the text from beginning to end—sometimes on short notice and under a gun. Thanks for numerous constructive suggestions go to Jeff Baggett, Toby Driscoll, Vicki Howle, Gudbjorn Jonsson, Kim Toh, and Divakar Viswanath. It is a privilege to have students, then colleagues, like these.

Working with the publications staff at SIAM has been a pleasure; there can be few organizations that match SIAM's combination of flexibility and professionalism. We are grateful to the half-dozen SIAM editorial, production, and design staff whose combined efforts have made this book attractive, and in particular, to Beth Gallagher, whose contributions begin with first-rate copy editing but go a long way beyond.

No institution on earth is more supportive of numerical linear algebra—or produces more books on the subject!—than the Computer Science Department at Cornell. The other three department faculty members with interests in this area are Tom Coleman, Charlie Van Loan, and Steve Vavasis, and we would like to thank them for making Cornell such an attractive center of scientific computing. Vavasis read a draft of the book in its entirety and made many valuable suggestions, and Van Loan was the one who brought Trefethen to Cornell in the first place. Among our non-numerical colleagues, we thank Dexter Kozen for providing the model on which this book was based: *The Design and Analysis of Algorithms*, also in the form of forty brief lectures. Among the department's support staff, we have depended especially on the professionalism, hard work, and good spirits of Rebekah Personius.

Outside Cornell, though a frequent and welcome visitor, another colleague who provided extensive suggestions on the text was Anne Greenbaum, one of the deepest thinkers about numerical linear algebra whom we know.

From September 1995 to December 1996, a number of our colleagues taught courses from drafts of this book and contributed their own and their students' suggestions. Among these were Gene Golub (Stanford), Bob Lynch (Purdue), Suely Oliveira (Texas A & M), Michael Overton (New York University), Haesun Park and Ahmed Sameh (University of Minnesota), Irwin Pressmann (Carleton University), Bob Russell and Manfred Trummer (Simon Fraser University), Peter Schmid (University of Washington), Daniel Szyld (Temple University), and Hong Zhang and Bill Moss (Clemson University). The record-breakers in the group were Lynch and Overton, each of whom provided long lists of detailed suggestions. Though eager to dot the last i, we found these contributions too sensible to ignore, and there are now hundreds of places in the book where the exposition is better because of Lynch or Overton.

Most important of all, when it comes to substantive help in making this a better book, we owe a debt that cannot be repaid (he refuses to consider it) to Nick Higham of the University of Manchester, whose creativity and scholarly attention to detail have inspired numerical analysts from half his age to twice it. At short notice and with characteristic good will, Higham read a draft of this book carefully and contributed many pages of technical suggestions, some of which changed the book significantly.

For decades, numerical linear algebra has been a model of a friendly and socially cohesive field. Trefethen would like in particular to acknowledge the three "father figures" whose classroom lectures first attracted him to the subject: Gene Golub, Cleve Moler, and Jim Wilkinson.

Still, it takes more than numerical linear algebra to make life worth living. For this, the first author thanks Anne, Emma (5), and Jacob (3) Trefethen, and the second thanks Heidi Yeh.

Contents

Part I

Fundamentals

Lecture 1. Matrix-Vector Multiplication

You already know the formula for matrix-vector multiplication. Nevertheless, the purpose of this first lecture is to describe a way of interpreting such products that may be less familiar. If $b = Ax$, then *b is a linear combination of the columns of A.*

Familiar Definitions

Let x be an n-dimensional column vector and let A be an $m \times n$ matrix (m rows, n columns). Then the matrix-vector product $b = Ax$ is the m-dimensional column vector defined as follows:

$$b_i = \sum_{j=1}^{n} a_{ij} x_j, \qquad i = 1, \dots, m. \tag{1.1}$$

Here b_i denotes the ith entry of b, a_{ij} denotes the i,j entry of A (ith row, jth column), and x_j denotes the jth entry of x. For simplicity, we assume in all but a few lectures of this book that quantities such as these belong to $\mathbb{C}$, the field of complex numbers. The space of m-vectors is $\mathbb{C}^m$, and the space of $m \times n$ matrices is $\mathbb{C}^{m\times n}$.

The map $x \mapsto Ax$ is *linear*, which means that, for any $x, y \in \mathbb{C}^n$ and any $\alpha \in \mathbb{C}$,

$$\begin{aligned} A(x+y) &= Ax + Ay, \\ A(\alpha x) &= \alpha A x. \end{aligned}$$

Conversely, every linear map from $\mathbb{C}^n$ to $\mathbb{C}^m$ can be expressed as multiplication by an $m \times n$ matrix.

A Matrix Times a Vector

Let a_j denote the jth column of A, an m-vector. Then (1.1) can be rewritten

$$b = Ax = \sum_{j=1}^{n} x_j a_j. \tag{1.2}$$

This equation can be displayed schematically as follows:

$$\begin{bmatrix} \\ b \\ \\ \end{bmatrix} = \left[\begin{array}{c|c|c|c} \\ a_1 & a_2 & \cdots & a_n \\ \\ \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} \\ a_1 \\ \\ \end{bmatrix} + x_2 \begin{bmatrix} \\ a_2 \\ \\ \end{bmatrix} + \cdots + x_n \begin{bmatrix} \\ a_n \\ \\ \end{bmatrix}.$$

In (1.2), b is expressed as a linear combination of the columns a_j. Nothing but a slight change of notation has occurred in going from (1.1) to (1.2). Yet thinking of Ax in terms of the form (1.2) is essential for a proper understanding of the algorithms of numerical linear algebra.

We can summarize these different descriptions of matrix-vector products in the following way. As mathematicians, we are used to viewing the formula $Ax = b$ as a statement that A acts on x to produce b. The formula (1.2), by contrast, suggests the interpretation that x acts on A to produce b.

Example 1.1. Vandermonde Matrix. Fix a sequence of numbers $\{x_1, x_2, \ldots, x_m\}$. If p and q are polynomials of degree $< n$ and α is a scalar, then $p+q$ and αp are also polynomials of degree $< n$. Moreover, the values of these polynomials at the points x_i satisfy the following linearity properties:

$$\begin{aligned} (p+q)(x_i) &= p(x_i) + q(x_i), \\ (\alpha p)(x_i) &= \alpha(p(x_i)). \end{aligned}$$

Thus the map from vectors of coefficients of polynomials p of degree $< n$ to vectors $(p(x_1), p(x_2), \ldots, p(x_m))$ of sampled polynomial values is linear. Any linear map can be expressed as multiplication by a matrix; this is an example. In fact, it is expressed by an $m \times n$ *Vandermonde matrix*

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{bmatrix}.$$

If c is the column vector of coefficients of p,

$$c = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{bmatrix}, \qquad p(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{n-1} x^{n-1},$$

then the product Ac gives the sampled polynomial values. That is, for each i from 1 to m, we have

$$(Ac)_i \;=\; c_0 + c_1 x_i + c_2 x_i^2 + \cdots + c_{n-1} x_i^{n-1} \;=\; p(x_i). \tag{1.3}$$

In this example, it is clear that the matrix-vector product Ac need not be thought of as m distinct scalar summations, each giving a different linear combination of the entries of c, as (1.1) might suggest. Instead, A can be viewed as a matrix of columns, each giving sampled values of a monomial,

$$A = \left[\begin{array}{c|c|c|c|c} & & & & \\ 1 & x & x^2 & \cdots & x^{n-1} \\ & & & & \end{array}\right], \tag{1.4}$$

and the product Ac should be understood as a single vector summation in the form of (1.2) that at once gives a linear combination of these monomials,

$$Ac \;=\; c_0 + c_1 x + c_2 x^2 + \cdots + c_{n-1} x^{n-1} \;=\; p(x). \qquad \square$$

The remainder of this lecture will review some fundamental concepts in linear algebra from the point of view of (1.2).

A Matrix Times a Matrix

For the matrix-matrix product $B = AC$, *each column of B is a linear combination of the columns of A.* To derive this fact, we begin with the usual formula for matrix products. If A is $\ell \times m$ and C is $m \times n$, then B is $\ell \times n$, with entries defined by

$$b_{ij} = \sum_{k=1}^{m} a_{ik} c_{kj}. \tag{1.5}$$

Here b_{ij}, a_{ik}, and c_{kj} are entries of B, A, and C, respectively. Written in terms of columns, the product is

$$\left[\begin{array}{c|c|c|c} & & & \\ b_1 & b_2 & \cdots & b_n \\ & & & \end{array}\right] = \left[\begin{array}{c|c|c|c} & & & \\ a_1 & a_2 & \cdots & a_m \\ & & & \end{array}\right] \left[\begin{array}{c|c|c|c} c_1 & c_2 & \cdots & c_n \end{array}\right],$$