Lanczos Algorithms for Large Symmetric Eigenvalue Computations

Vol. I: Theory

Jane K. Cullum
Ralph A. Willoughby

C • L • A • S • S • I • C • S
In Applied Mathematics
siam
41

Lanczos Algorithms for Large Symmetric Eigenvalue Computations

Vol. I: Theory

Jane K. Cullum
Ralph A. Willoughby

siam
Society for Industrial and Applied Mathematics
Philadelphia

This SIAM edition is an unabridged republication of the work first published by Birkhäuser Boston, 1985.

10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Cullum, Jane K.,

Lanczos algorithms for large symmetric eigenvalue computations / Jane K. Cullum, Ralph A. Willoughby.

p. cm. -- (Classics in applied mathematics ; 41)

Originally published: Boston : Birkhäuser, 1985.

Includes bibliographical references and index.

Contents: vol. I. Theory

ISBN 0-89871-523-7 (v. 1 : pbk.)

1. Symmetric matrices--Data processing. 2. Eigenvalues--Data processing. I. Willoughby, Ralph A. II. Title. III. Series

QA193.C84 2002

512.9'434--dc21 2002075847

Classics in Applied Mathematics (continued)

Cornelius Lanczos, *Linear Differential Operators*

Richard Bellman, *Introduction to Matrix Analysis, Second Edition*

Beresford N. Parlett, *The Symmetric Eigenvalue Problem*

Richard Haberman, *Mathematical Models: Mechanical Vibrations, Population Dynamics, and Traffic Flow*

Peter W. M. John, *Statistical Design and Analysis of Experiments*

Tamer Başar and Geert Jan Olsder, *Dynamic Noncooperative Game Theory, Second Edition*

Emanuel Parzen, *Stochastic Processes*

Petar Kokotović, Hassan K. Khalil, and John O'Reilly, *Singular Perturbation Methods in Control: Analysis and Design*

Jean Dickinson Gibbons, Ingram Olkin, and Milton Sobel, *Selecting and Ordering Populations: A New Statistical Methodology*

James A. Murdock, *Perturbations: Theory and Methods*

Ivar Ekeland and Roger Témam, *Convex Analysis and Variational Problems*

Ivar Stakgold, *Boundary Value Problems of Mathematical Physics, Volumes I and II*

J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*

David Kinderlehrer and Guido Stampacchia, *An Introduction to Variational Inequalities and Their Applications*

F. Natterer, *The Mathematics of Computerized Tomography*

Avinash C. Kak and Malcolm Slaney, *Principles of Computerized Tomographic Imaging*

R. Wong, *Asymptotic Approximations of Integrals*

O. Axelsson and V. A. Barker, *Finite Element Solution of Boundary Value Problems: Theory and Computation*

David R. Brillinger, *Time Series: Data Analysis and Theory*

Joel N. Franklin, *Methods of Mathematical Economics: Linear and Nonlinear Programming, Fixed-Point Theorems*

Philip Hartman, *Ordinary Differential Equations, Second Edition*

Michael D. Intriligator, *Mathematical Optimization and Economic Theory*

Philippe G. Ciarlet, *The Finite Element Method for Elliptic Problems*

Jane K. Cullum and Ralph A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Vol. I: Theory*

PREFACE TO THE CLASSICS EDITION

Since 1985, when this book was first published, interest in practical Lanczos algorithms for computing eigenvalues of large scale problems has soared. The developments since 1985 could fill easily a new two-volume book.

The computers of today are many orders of magnitude more powerful than the computers that we used in the early 1980s when the algorithms which are described in this book were developed. In 1984 we labeled problems of size 200 or larger as *large scale* and speculated that our codes should be useful on problems up to size 10,000.

Today, eigenvalue problems that involve *millions* of degrees of freedom are of great interest for many different kinds of scientific and engineering studies, and, in fact, parallel extensions of the real symmetric Lanczos algorithm described in Chapter 4 of this monograph have been used on problems involving more than a million variables.

The focus of this research monograph is on *symmetric* problems. Symmetric does not, however, imply Hermitian. The discussion not only covers real symmetric and Hermitian problems but also covers singular value problems and complex symmetric problems.

This monograph is unique in terms of the types of algorithms that are presented. Most of the numerical analysis community focuses on procedures that are firmly entrenched in the *orthogonal* world where spectral entities of orthogonal projections of operators are used to obtain corresponding approximations for the original problem. The use of orthogonal projections ensures direct relationships between these projections and the original problem. The popular package that is described in [1] utilizes this approach.

The algorithms in Chapters 4, 5, and 6 are based upon a different paradigm. The justification for our approach rests upon the nonintuitive discovery that the real symmetric Lanczos recursion is numerically stable in finite precision arithmetic [2, 3]. The analysis in [2, 3], when combined with our nonintuitive discovery that the eigenvalues of any Lanczos matrix can be *systematically* sorted into distinct subsets of *good* and *spurious* eigenvalues, enabled the development of practical Lanczos eigenvalue algorithms that do not require reorthogonalization. Details are in Chapter 4.

These algorithms function in two stages. Eigenvalues (singular values) are computed separately from corresponding eigenvector (singular vector) approximations. Separating these computations has two significant consequences. First, the amount of memory required for the eigenvalue computations is minimal, only some small multiple of the size of the original problem matrix. The size of this multiplier depends upon what the user wants to compute and upon the distribution of the spectrum of A. With minimal memory requirements, very large problems can be handled on not very large computers, and huge problems can be handled on large computers.

Second, the achievable accuracy of the eigenvalue approximations obtained by one of these algorithms can be greater than that achievable using a classical approach which incorporates reorthogonalization with respect to eigenvector approximations which are accurate to only a few digits. The procedures in Chapters 4, 5, and 6 do not invoke any type of deflation, and as long as the errors in the Lanczos recursions remain small, there is no degradation in the accuracy as more eigenvalue approximations are computed.

The eigenvalue procedures described in Chapters 4, 5, and 6 are simple to parallelize. There are no potential communication bottlenecks associated with a need to reorthogonalize every vector with respect to all vectors previously generated.

The penalty for not being *orthogonal* is that the matrix-vector multiplications required for the Lanczos recursions must be computed consistently at each stage of the Lanczos recursion. This limits the use of these methods in applications where the matrix-vector computations are inexact with varying accuracy. For example, these methods would not work well when these computations are accomplished via finite differences of nonlinear functions or as the numerical solution of a differential equation. In such a situation, the algorithms in Chapters 4, 5, and 6 could be used to obtain limited information about the spectrum but would eventually diverge. An approach based upon orthogonalization, such as in [1], may continue to function in such an environment, but the accuracy of the computed values may be difficult to determine.

Chapters 6 and 7 contain research that is not available in archival journals. In Chapter 6 we demonstrate that much of the analysis for the real symmetric Lanczos algorithm in Chapter 4 can be formally extended to the complex symmetric case.

The algorithms in Chapter 7 differ from those in Chapters 4, 5, and 6 in that they are block methods and belong to the *orthogonal* world. Reorthogonalization of vectors is invoked but in a very limited way. Of primary importance in Chapter 7 is our proposal for the *implicit deflation* of nearly dependent vectors from blocks. This form of deflation accelerates convergence, allows *converged* eigenvector approximation to continue to converge, and mollifies the negative effects which vector deflation can have upon approximations which have not yet converged. Details are in Chapter 7.

Algorithms that are based upon the work in this book, in spite of their relatively advanced age, are still being used. See, for example, [4], which contains numerical comparisons of the algorithm described in Chapter 4 with that in [1].

Volume 2 of this book, which is not being reproduced in the SIAM *Classics in Applied Mathematics* series, contains listings of all of the original implementations of the algorithms that are discussed in this monograph. The text for Volume 2, along with the source code for each of those listings, is available at the numerical analysis community Web repository, *www.netlib.org*, under the name "*lanczos*."

We hope that republication of this book will be of benefit to the computational physics, chemistry, and engineering communities that have shown interest in these types of algorithms. We also hope that students in the numerical linear algebra community will find this book useful as a means for building appreciation for the potential gains that may be achievable by thinking outside of the *orthogonal* box.

Regrettably, my co-author, Ralph Willoughby, passed away in July 2001 prior to the republication of our book in the SIAM *Classics in Applied Mathematics* series.

Ralph and I worked together from 1977 to 1991. Most of our research on "symmetric" problems is contained in this book. From 1984 through 1991 the focus of our work on algorithms for eigenvalue problems moved to nonsymmetric problems and to the extension of many of the ideas in this book to those types of problems.

Our joint publication in 1991 provided strategies for selecting appropriate matrix shifts for use in either symmetric or nonsymmetric shift and invert algorithms.

I wish to express my personal gratitude to Ralph's wife, Nona Willoughby, for her support in the republication of this book.

Jane Cullum

March 2002

REFERENCES

[1] K. J. MASCHHOFF AND D. C. SORENSON, *P-ARPACK: An efficient portable large scale eigenvalue package for distributed memory parallel architectures,* in Applied Parallel Computing, Industrial Computation and Optimization, Third International Workshop, PARA '96 Proceedings (1996), Springer-Verlag, Berlin, 478–486.

[2] C. C. PAIGE, *Error analysis of the Lanczos algorithms for tridiagonalizing a symmetric matrix,* J. Inst. Math. Appl., 18 (1976), 341–349.

[3] C. C. PAIGE, *Accuracy and effectiveness of the Lanczos algorithm for the symmetric eigenproblem,* Linear Algebra Appl., 34 (1980), 235–258.

[4] M. SCHREIBER, F. MILDE, and R. A. ROMER, *Electronic states in the Anderson model of localization: Benchmarking eigenvalue algorithms,* Comput. Phys. Comm., 121-122 (1999), 517–523.

PREFACE

Energy levels, resonances, vibrations, feature extraction, factor analysis - the names vary from discipline to discipline; however, all involve eigenvalue/eigenvector computations. An engineer or physicist who is modeling a physical process, structure, or device is constrained to select a model for which the subsequently-required computations can be performed. This constraint often leads to reduced order or reduced size models which may or may not preserve all of the important characteristics of the system being modeled. Ideally, the modeler should not be forced to make such a priori reductions. It is our intention to provide here procedures which will allow the direct and successful solution of many large 'symmetric' eigenvalue problems, so that at least in problems where the computations are of this type there will be no need for model reduction.

Matrix eigenelement computations can be classified as small, medium, or large scale, in terms of their relative degrees of difficulty as measured by the amount of computer storage and time required to complete the desired computations. A matrix eigenvalue problem is said to be small scale if the given matrix has order smaller than 100. Well-documented and reliable FORTRAN programs exist for small scale eigenelement computations, see in particular EISPACK [1976,1977]. Typically those programs explicitly transform the given matrix into a simpler canonical form. The eigenelement computations are then performed on the canonical form. For the EISPACK programs the storage requirements grow as the square of the order of the matrix being processed and the operation counts grow cubically with the order.

A matrix eigenvalue problem is said to be medium scale if it is real symmetric, and if it is computationally feasible to compute the eigenelements by using Sturm sequencing and bisection in combination with inverse iteration directly on the original matrix. For example, a band matrix with a reasonable band width will be said to be of medium scale.

Matrix eigenvalue computations are said to be large scale if the size of the matrix and the pattern of the nonzeros in the matrix preclude the use of EISPACK-type procedures or of a Sturm sequencing/bisection/inverse iteration approach. For example, if the given matrix has order larger than 200 and is not banded, then we would classify the associated eigenvalue computation for that matrix as large scale. In most of our experiments, large scale has meant of order greater than 500.

We focus on large scale, 'symmetric' matrix eigenelement computations. Symmetric is in quotes because we include a procedure for computing singular values and vectors of real rectangular matrices. In addition we also include a procedure for computing eigenelements of nondefective complex symmetric matrices. Such matrices do not possess the desirable properties of real symmetric matrices, and the amount of computation required to process them can be significantly more than that required in the real symmetric case. We will not address the general nonsymmetric eigenvalue problem.

This is a research monograph intended for engineers, scientists and mathematicians who are interested in computational procedures for large matrix eigenvalue problems. The discussion focuses on one particular subset of one particular family of procedures for large matrix eigenvalue computations. We are interested in Lanczos procedures.

Lanczos procedures derive their name from a famous (fifteen years ago some people may have been inclined to say infamous) 3-term recursion that was originally proposed as a means of transforming a general real symmetric matrix into a real symmetric tridiagonal matrix. Within the scientific community we find two basic approaches to Lanczos procedures. One approach maintains that 'global orthogonality is crucial' and the other approach maintains that 'local orthogonality is sufficient'. This terminology is explained in Chapter 2, Sections 2.4 and 2.5 where we briefly survey the literature on Lanczos procedures for eigenelement computations.

Our emphasis is on computation. We focus primarily on our research on single-vector Lanczos procedures with no reorthogonalization. These procedures belong to the class of 'local orthogonality is sufficient' procedures. The material is organized into two volumes. This volume contains the material necessary for understanding the proposed Lanczos procedures. The second volume contains the FORTRAN codes and documentation for each of the Lanczos procedures discussed in this volume.

Users with large problems are concerned about the amounts of computer storage and time required by the procedures which they have to use. Our single-vector Lanczos procedures are storage efficient. In most cases they are also time efficient if the matrix, whose eigenvalues (singular values) are to be computed, is such that matrix-vector multiplies can be computed rapidly and accurately. Typically if the given matrix is sparse, in the sense that there are only a few nonzero entries in each row and column, then this can be achieved.

Some of what is presented is new and has not yet been published elsewhere. Much of what is presented has appeared at least in preliminary form in papers and reports published in various places. It is hoped that by bringing all of this material together in one place, that these results will prove useful to a wide variety of users in the engineering and scientific community.

Jane K. Cullum
Ralph A. Willoughby
July 1984

INTRODUCTION

We consider the question of the computation of eigenvalues and eigenvectors of large, 'symmetric' matrices. While in a strictly mathematical sense, the scope of this book is very narrow, the potential applications for the material which is included are important and numerous. Perhaps the most familiar application of eigenvalue and eigenvector computations is to structural analysis, studies of the responses of aircraft, of bridges, or of buildings when they are subjected to different types of disturbances such as air turbulence, various types of loadings, or earthquakes. In each case, the physical system being analyzed varies continuously with time, and its true motion is described by one or more differential equations. Matrix eigenvalue problems and approximations to this motion are obtained by discretizing the system equations in some appropriate way.

For a given matrix A, the 'simple' eigenvalue-eigenvector problem is to determine a scalar λ and a vector $x \neq 0$ such that $Ax = \lambda x$. In structural problems, one typically encounters the generalized eigenvalue problem $Kx = \lambda Mx$, involving 2 different matrices, a mass matrix M and a stiffness matrix K. In fact the problems there can be nonlinear quadratic eigenvalue problems, see for example Abo-Hamd and Utku [1978]. However, the 'solution' of a linearization of one of these quadratic problems is often used as a basis for reducing the given nonlinear eigenvalue problem to a much smaller but dense generalized eigenvalue problem. In such a problem a few of the smallest eigenvalues and corresponding eigenvectors may be required or in some cases in order to determine the response of a given structure to external disturbances it may be necessary to compute eigenvalues and corresponding eigenvectors on some interior interval of the spectrum of the given matrix. In structures the matrices used are typically banded, that is all of the nonzero entries are clustered around the main diagonal of the system matrices. For many years simultaneous iteration techniques have been applied successfully to shifted and inverted matrices $(K - \mu M)$, using equation solving techniques designed for band matrices.

Very large matrix eigenvalue problems also arise in studies in quantum physics and chemistry, see for example Kirkpatrick [1972] and Gehring [1975]. The matrices generated are large and sparse. Typically significant numbers of the eigenvalues of these matrices are required. An entirely different kind of application is the use of eigenvectors in heuristic partitioning algorithms, see for example Barnes [1982]. For the particular application which Barnes considered, the placement of electrical circuits on silicon chips, the goal was to position a large number of circuits on a given number of chips in such a way that the resulting number of external connections between circuits on different chips was minimized.

Other applications for eigenvalue/eigenvector computations occur in quantum chemistry, see for example Nesbet [1981]; in power system analysis, see for example Van Ness [1980]; in oceanography, see for example Winant [1975] and Platzman [1978]; in magnetohydrodynamics, see for example Gerlakh [1978]; in nuclear reactor studies, see for example Geogakis [1977]; in helicopter stability studies, see for example Hodges [1979]; and in geophysics, see for example Kupchinov [1973].

As we said earlier we are considering the question of computing eigenvalues and eigenvectors of large 'symmetric' matrices which arise in various applications. The word symmetric is in quotes because we also present procedures for two types of matrix computations which are not symmetric in the ordinary sense. The basic ideas which we discuss are equally applicable to any matrix problem which is equivalent to a real symmetric eigenvalue/eigenvector problem. We consider several such equivalences. These include Hermitian matrices, certain real symmetric generalized eigenvalue problems, and singular value and singular vector computations for real, rectangular matrices. We also consider complex symmetric matrices which are not equivalent to real symmetric matrices.

The actual scope of this book is limited to a particular family of algorithms for large scale eigenvalue problems, the Lanczos procedures. Other types of eigenelement procedures suitable for large matrices exist, most of which are based upon either simultaneous iterations or upon Rayleigh quotient iterations, see Bathe and Wilson [1976] and Jennings [1977] for complete and very readable discussions of simultaneous iteration procedures. The research on Rayleigh quotient iteration procedures is scattered. Parlett [1980, Chapter 4] discusses the theoretical properties of such procedures and gives references for interested readers. We do not cover any of the non-Lanczos procedures in our discussions.

The research on Lanczos procedures for eigenelement computations (and for solving systems of equations) continues. Although many interesting results have been obtained, many of the theoretical questions concerning Lanczos procedures have not been satisfactorily resolved. Much of the existing literature on Lanczos procedures has not adequately incorporated the effects of roundoff errors due to the inexactness of the computer arithmetic. Numerical experiments with various Lanczos procedures have however clearly demonstrated their advantages and capabilities. Many different people have contributed to this research, and we apologize if we have neglected to mention one or more of these authors in our discussions or if for the authors we do mention we have not referenced all of their papers on this subject.

The demonstrated computational efficiences and excellent convergence properties which can be achieved by Lanczos procedures, have generated much interest in the scientific and engineering communities. Parlett [1980] is primarily devoted to discussions of small to medium size real symmetric eigenvalue problems where other type of eigenelement procedures are applicable. However, Chapter 13 of that book is devoted to Lanczos procedures for large matrices, but that discussion focuses on the 'global orthogonality is crucial' approach to Lanczos procedures and there is not much discussion of the 'local orthogonality is sufficient' approach which we use in our single-vector Lanczos procedures. Specific comments regarding the differences between these two approaches are given in Sections 2.4 and 2.5 of Chapter 2. We focus primarily on one subset of the Lanczos eigenelement procedures, the single-vector Lanczos procedures which do not use any reorthogonalization. Iterative block Lanczos procedures with limited reorthogonalization are also discussed but to a lesser extent.

This book is divided into two volumes. This volume provides the background material necessary for understanding the Lanczos procedures which we have developed and gives some perspective of the existing research on Lanczos procedures for eigenvalue or singular value computations. The second volume contains FORTRAN programs for each of the Lanczos procedures discussed in this volume. We have tried to make these volumes self-contained by including the material from matrix theory which is necessary for following the arguments given. Both volumes of this book should be accessible to engineers and scientists who have some knowledge of matrix eigenvalue problems. References are given to other books and papers where the interested reader can pursue various topics discussed.

Chapter 0 is intended as a reference chapter for the reader. Basic definitions and concepts from matrix theory which are used throughout the book are listed. Our notation is specified and special types of matrices are defined, along with special types of matrix transformations and projections.

Chapter 1 contains brief summaries of fundamental results from matrix theory which are needed in later chapters. Properties of real symmetric matrices, of Hermitian matrices, and of real symmetric generalized eigenvalue problems are summarized. Sparse matrices are discussed along with sparse matrix factorizations.

Chapter 2 begins with a description of a basic single-vector Lanczos procedure for computing eigenelements of real symmetric matrices. Properties of this procedure are derived, assuming that the computations are being performed in exact arithmetic. However, we are interested in Lanczos procedures which do not use any reorthogonalization and must therefore be concerned with what happens in finite precision arithmetic. In Section 2.3 we summarize the results obtained by Paige [1971,1972,1976,1980], assuming finite precision arithmetic. These results are the basis for the arguments which are given in Chapter 4 to justify our Lanczos procedures with no reorthogonalization. In Section 2.4 we discuss the question of constructing practical Lanczos procedures, that is, procedures which are numerically-stable in finite precision arithmetic. Section 2.5 consists of a survey of the literature on Lanczos procedures.

Chapter 3 contains proofs of several basic properties of general tridiagonal matrices, including determinant recursions and formulas for computing eigenvectors from the determinants of the matrix, along with comments on inverse iteration computations. We need these properties in Chapters 4, 5, and 6.

Chapter 4 is the main chapter of this volume. Here we develop the single-vector Lanczos procedure with no reorthogonalization for real symmetric matrices. Included is a discussion of the relationships between Lanczos tridiagonalization and the conjugate gradient method for solving systems of equations. This relationship is used to construct a plausibility argument for the belief that the 'local orthogonality is sufficient' approach is legitimate. The key to the success of these types of eigenvalue procedures, an identification test which sorts the 'good' eigenvalues from the 'spurious' ones, is developed in Section 4.5. This test is justified heuristically using the connection of the Lanczos recursion with conjugate gradient iterations. Results of numerical experiments are used to demonstrate the performance of this procedure on different types of matrices. FORTRAN code for this procedure is given in Chapter 2 of Volume 2. Chapters 3, 4, and 5 of Volume 2 contain respectively, FORTRAN codes for corresponding Lanczos procedures for Hermitian matrices, for factored inverses of real symmetric matrices, and for certain real symmetric generalized eigenvalue problems.

Chapter 5 addresses the question of constructing a single-vector Lanczos procedure for computing singular values and singular vectors of real rectangular matrices. A general discussion of basic properties of singular values and singular vectors and of the relationships between singular values and eigenvalues is given. Section 5.3 contains a very brief discussion of several applications of singular values and vectors. Section 5.4 centers on our single-vector Lanczos procedure with no reorthogonalization. Results of numerical experiments are included to demonstrate the performance of this procedure. FORTRAN code for this procedure is given in Chapter 6 of Volume 2.

Chapter 6 addresses the question of constructing a single-vector Lanczos procedure for diagonalizable complex symmetric matrices. This class of matrices is genuinely nonsymmetric, possessing none of the desirable properties of real symmetric matrices. Relevant properties of complex symmetric matrices are included. The Lanczos procedure which is proposed maps the given complex symmetric matrix into a family of complex symmetric tridiagonal matrices. FORTRAN code for this procedure is given in Chapter 7 of Volume 2.

Chapter 7 addresses the question of iterative block Lanczos procedures. First a practical implementation of the block Lanczos procedure given in Cullum and Donath [1974] is discussed. We then describe a recently-developed hybrid procedure which combines ideas from the single-vector Lanczos procedures and from the iterative block procedure. FORTRAN code for this hybrid procedure is given in Chapter 8 of Volume 2.

Table of Contents

CHAPTER 2

LANCZOS PROCEDURES, REAL SYMMETRIC PROBLEMS

CHAPTER 3

TRIDIAGONAL MATRICES

CHAPTER 4

LANCZOS PROCEDURES WITH NO REORTHOGONALIZATION FOR REAL SYMMETRIC PROBLEMS