BOOK REVIEWS

Iterative Methods for Sparse Linear Systems

by Yousef Saad
PWS Publishing Company, Boston, 1996, XVI + 447 pages
A division of International Thomson Publishing Company
ISBN 0-534-94776-X

Two books dominate nowaday the numerical scientific community of computational linear algebra. The one is the book *of Wolfgang Hackbusch: Iterative Solution of Large Sparse Systems of Equations,* Springer- Verlag, 1994 and the other is the book *of Yousef Saad: Iterative Methods for Sparse Linear Systems*, PWS Publishing Company. Both of them deal with the same subject: iterative methods for solving large sparse systems of linear equations. Both of them are full of theory, methods, algorithms and numerical examples. Both of them are supported by computer programs.

The subject of iterative methods in computational linear algebra is not new. We can also mention here the books of *O. Axelsson: Iterative Solution Methods*, Cambridge University Press, 1994, and the excellent and older ones by *R.S. Varga: Matrix Iterative Analysis*, Prentice -Hall, Englewood Cliffs, 1960, and *D.M. Young: Iterative Solution of Large Linear Systems*, Academic Press, 1971.

Solving linear algebraic systems by iterative methods dates back in time. In a letter of 26th December 1823 dedicated to C.L. Gerling, Carl Friedrich Gauss had described an alternative method for solving systems of linear equations of type ATAx = ATb, namely an iterative method (called by Gauss an "indirect method"). This was the first iterative method for systems of linear equations. It was also described by Gauss in "Supplementum theoriae combinationis observationum erroribus minime obnoxiae" (1828). A very similar method was described by Carl Gustav Jacobi in his paper "Über eine neue Auflosungsart der bei der Methode der kleinsten Quadrate vorkommenden linearen Gleichungen", published in 1845. Later, in 1874 Philip Ludwig Seidel, an alumnus of Jacobi' s, designed a method for solving linear algebraic systems of equations. After 100 years of stagnation in this field, especially since the time that electronic computers became available for solving large systems of equations, Southwell and Young experimented with variants of the Gauss-Seidel method leading to important acceleration of the convergence.

A very important computational innovation of the early 1950 are the conjugate gradient and Lanczos algorithms for solving linear systems of equations and eigenproblems. These methods came into wide use only in the mid-1970s, and shortly thereafter vector computers and massive computer memories made it possible to use this iterative method to solve problems which could not be solved in any other way. The original development of this family of algorithms was made by Cornelius Lanczos and Magnus Hestens at the Institute for Numerical Analysis in the National Applied Mathematics Laboratories at the US National Bureau of Standards in Los Angeles, and by Eduard Steifel from the Technische Hochschule Zürich. Since then, the algorithms have been further refined and have become a basic tool for solving large systems of equations. A contribution to the understanding of these algorithms has been brought inter alios by John Reid who drew attention to the potential of the conjugate gradient method as an iterative one for sparse linear systems. Other developments were made by C. Paige and Michael Saunders who provided the first stable extension of the conjugate gradient algorithm to indefinite matrices. Paul Concus and Gene Golub considered a class of nonsymmetric matrices. Axelsson suggested preconditioning conjugate gradients by a scaled successive overrelaxation operator. Other preconditionings were discussed by D. Evans, Richard Bartels, James Daniel, R. Chandra, Stanley Eisenstat, etc. The dissertation of C. Paige (1971) provides, among other things, the first step to an understanding of the loss of orthogonality of the Lanczos vectors, thus giving the key to the development of stable algorithms that did not require complete reorthogonalization.

In the following we shall review the book of Professor Saad, from University of Minnesota.

From the very beginning the reader is intrigued with the title, as the sparse linear systems are always associated with direct methods. Here we have a new perspective of iterative methods for solving general, large sparse linear systems. Until recently, direct methods were preferred to iterative methods in real applications because of their robustness and predictable behaviour. However, the increased need for solving very large linear systems triggered the efforts to rapid developments of the iterative techniques. This volume tries to provide up-to-date coverage of iterative methods for solving large sparse linear systems. The author focussed the book on practical methods that work for general sparse matrices rather than for any specific class of problems. Two ideas are constantly followed along the 13 chapters of the book. The first is the consideration of sparsity for designing efficient iterative methods, and the second is the usage of preconditioners and accelerators to improve the behaviour of the iterative techniques.

The book is structured in four distinct parts. The first part (Chapters 1 through 4) has an introductory character, presenting the basic tools. The second part (Chapters 5 through 8) presents projections methods and the Krylov subspace techniques. The third part (Chapters 9 and 10) discusses preconditioning techniques. The last part (Chapters 11 through 13) discusses modern parallel implementations and parallel algorithms. The book is supported with numerous numerical examples, some of them with large scale dimensions, and documents computational results obtained with the SPARSKIT package: a basic tool kit for sparse matrix computations.

Chapter 1, "Background in Linear Algebra", gives an overview of relevant concepts in linear algebra which are used in later chapters. A review of basic matrix theory and the elementary notation used throughout the book is provided.

Chapter 2, "Discretization of PDES", considers the techniques for solving partial differential equations by discretization. The idea is to approximate them by linear equations that involve a finite (enough large) number of unknowns. The matrices that arise from these discretizations are generally large, sparse and often present special structures. Several different ways of discretizing a PDE are presented. The simplest method uses finite difference approximations for the partial differential operators. The Finite Element Method replaces the original function by a function which has some degree of smoothness over the global domain, but which is piecewise polynomial on simple cells, such as small triangles or rectangles. This method is the most general, the oldest and understood discretization technique available. In between these two methods there are a few conservative schemes called Finite Volume Methods, which attempt to emulate the continuous conservation laws of physics. The material of this Chapter is well organized and constitutes a good introduction into this subject matter.

Chapter 3, "Sparse Matrices", presents the fundamental aspects and the most important ideas concerning the sparse matrix technology. The sparse matrix techniques begin with the idea that the zero elements need not be stored. The key issue is to define data structures for these matrices that are well suited for efficient implementation of iterative or direct standard solution methods. The most important aspects considered here refer to the graph representation of sparse matrices, permutations and reordering, irreducibility, storage schemes, basic sparse matrix operations and sparse direct solution methods. The Cuthill - McKee and the reversed Cuthill-McKee orderings, which are important ingredients for solving large sparse systems of linear equations, are clearly presented and illustrated by means of numerical examples. No complexity aspects are considered here.

Chapter 4 is dedicated to presenting the **"Basic Iterative Methods"**. The first iterative methods used for solving linear systems were based on relaxation of the coordinates principle. Given an approximate solution, these methods modify the components of this approximation, one or a few of them at a time and in a certain order, until the convergence is obtained. The purpose of these modifications, called relaxation steps, is to annihilate one or some components of the residual vector. The chapter begins with the methods of Jacobi, Gauss-Seidel and the Successive Over Relaxation (SOR) method. These methods are then generalized to block relaxation schemes, and some classical results on general convergence of these schemes are provided. The iteration matrices and preconditioning are introduced. The word preconditioning was firstly used by Turing and since then seems to be standard terminology for problem transformation in order to make solution easier. This is a very important innovation in computational linear algebra and is reconsidered in Chapters 9 and 10 of this book in the context of the iterative methods for sparse linear systems.

Chapter 5, "Projection Methods", considers the so called Petrov-Galerkin framework for extracting an approximation to the solution of a linear system from a subspace. Firstly, these techniques are described into a very general frame, and later the one dimensional case is covered in detail providing a good preview of the more complex projection processes considered in later chapters. For the case where the matrix is symmetric and positive definite the steepest descent, the minimal residual and the residual norm steepest descent algorithms are presented. No numerical comparisons, and complexity analysis are provided.

Chapters 6 and 7 are dedicated to presenting the **"Krylov Subspace Methods".** Currently these methods are considered to be among the most important techniques available for solving large linear systems. These techniques are based on projection process onto Krylov subspaces, which are subspaces spanned by vectors of the form p(A)v, where p is a polynomial. The idea of these techniques is to approximate A-1b by p(A), where p is a good polynomial. These Chapters cover a number of methods and algorithms implementing this idea. Firstly the Krylov subspaces are introduced. Then the basic algorithm of Arnoldi is presented, as well as its variants: Arnoldi-Modified Gram-Schmidt, Householder-Arnoldi are discussed. A complexity analysis of these algorithms is presented. A very detailed description of Arnoldi's method for linear systems implemented as the Full Orthogonalization Method, the Incomplete Orthogonalization Method and the Direct Incomplete Orthogonalization Method are considered next. The Generalized Minimum Residual Method and its variant: the Householder version is then considered. To improve the behaviour of this method a number of variations like restarting and truncated techniques is also described. A simplification of Arnoldi's method for the particular case when the matrix is symmetric is the Symmetric Lanczos algorithm. On the other hand, for solving sparse symmetric positive definite linear systems the best known iterative method is the Conjugate Gradient Method. Both these methods are very well presented at the algorithmic and complexity analysis level. The Faber-Manteuffel theorem and the convergence analysis of different algorithms are the top subjects of Chapter 6. Some extensions of the above presented methods to nonsymmetric matrices are described in Chapter 7 where the Lanczos biorthogonalization, the Lanczos algorithm for linear systems, the Biconjugate Gradient algorithm, the Quasi-Minimal Residual algorithm, the Conjugate Gradient Squares algorithm, the Biconjugate Gradient Stabilized algorithm, the Transpose-Free Quasi-Minimal Residual algorithm are detailed. Numerical examples and comparisons are provided. The techniques and algorithms presented in these Chapters, which form the heart of the book, are under very intensive research activity, numerous variations and innovations being in process of elaboration.

Chapter 8, "Methods Related to the Normal Equations", is dedicated to present techniques and algorithms for solving the so- called normal equations: ATAx=ATb. This is a very important subject especially due to the impact of these methods on the interior point methods for large scale optimization. Firstly the normal equations are introduced, and then the Row Projection Methods (Gauss-Seidel on the normal equations and Cimmino's method), the Conjugate Gradient Methods and some variants, as well as the Saddle-Point problems (Uzawa's method and the Arrow-Hurwicz algorithm) are considered. These algorithms are illustrated on some numerical examples and comparisons are provided.

Despite their intrinsic appeal to solving very large linear systems, it is widely recognized that the lack of robustness of iterative methods, relative to direct solvers, is the main drawback of acceptance of these methods in industrial applications. Both the efficiency and robustness of iterative techniques can be improved by using preconditioning. **Chapters 9 and 10 "Preconditioned Iterations" and "Preconditioning Techniques"** respectively are for presenting this term which is the key ingredient for the accelerating of the Krylov subspace methods. In Chapter 9 the preconditioned versions of the iterative methods already considered are presented. Thus, the Preconditioned Conjugate Gradient algorithm, the Preconditioned Generalized Minimum Residual algorithm and the Preconditioned Conjugate Gradient for the Normal equations are presented. Chapter 10 covers the standard preconditioning techniques. The main purpose of this Chapter is to present techniques for constructing the preconditioning matrices. Thus, the Jacobi, SOR and Symmetric SOR preconditioners are firstly presented. Then the Incomplete LU factorization (ILU) preconditioners are deeply described, emphasizing a number of aspects concerning the filling, the numerical stability, preconditioners for matrices with regular structure, as well as the implementation details. Simple sample FORTRAN codes for computing preconditioners, as well as numerical experience and comparison are provided. For the indefinite matrices the standard incomplete LU factorization may face several difficulties. To cope with such a case some other techniques like approximate inverse preconditioners, factored approximate inverses are considered. The last part of Chapter 10 considers preconditioners for the normal equations. This is a very important numerical innovation with a great impact on the interior point algorithms for large scale linear programming. *The paper of Kim and Nazareth: A primal null-space affine scaling method*, ACM-TOMS, 1994, is a good starting research for implementing these iterative methods.

The last three **Chapters, 11 through 13, "Parallel Implementations", "Parallel Preconditioners", and "Domain Decomposition Methods"** consider the very modern parallel approach of linear computational algebra referring the iterative techniques for solving large sparse linear systems. Two approaches for developing parallel iterative algorithms are known. The first one extracts parallelism whenever possible from standard sequential algorithms. The second approach is to develop alternative algorithms which have enhanced parallelism. **Chapter 11** describes methods in the first category, giving an introduction into the field of parallel computing, and emphasizing some aspects on the types of parallel architectures and matrix by vector products. **Chapter 12** reinforces the preconditioning techniques into a parallel environment by presenting methods for finding preconditioners that have a high degree of parallelism, as well as good intrinsic qualities. Thus, the Block-Jacobi preconditioners, Polynomial preconditioners, Multicoloring, Multi-Elimination Incomplete LU factorization,

Distributed Incomplete LU factorization, as well as some other techniques are described. Some of the corresponding algorithms are illustrated by means of numerical examples of large scale dimensions. No complexity results are considered. The **last Chapter** refers to a collection of techniques which combine ideas from partial differential equations, linear algebra, mathematical analysis, graph theory, to solve large scale linear systems based on the principle of "divide-and-conquer". Different types of partitionings, the Schur complement elimination, the Schwarz alternating procedures, as well as graph partitioning are described at a high theoretical level.

The book is well- organized, providing an excellent text on the iterative methods in computational linear algebra, proving a lot of mathematical results, documenting numerous numerical experiments (with SPARSKIT), keeping an eye on the practical implementations of these techniques. At the end of each chapter numerous exercises are presented and 235 references are mentioned. Some other points like: how the sparsity could be more deeply speculated over in an iterative method, the sensitivity analysis of iterative methods, the complexity analysis of iterative methods, combinations between the direct and iterative methods, have not been considered. These subjects and some others could be considered into a new version of this book. All in all the reviewer considers this as a valuable work. He also recommends, to the extent possible, a parallel reading of this book and of Hackbush's book.

Neculai Andrei

Dr. Neculai Andrei graduated in Electrical Engineering, from the Polytechnical Institute of Bucharest and took his Ph.D degree for his contribution to digraph approach of large-scale sparse linear dynamic systems in 1973 and 1984 respectively. He is author of a number of published papers and technical reports in the area of mathematical modelling and optimization. He is author of the books: *Sparse Matrices and their Applications*, Technical Publishing House, 1983 Bucharest, Romania, and *Sparse Systems - Digraph approach of large-scale linear systems theory*, TUV Verlag - Rheinland, 1985 Köln, Germany. His main current scientific interests centre on languages for mathematical programming problems modelling, large-scale linear and non-linear optimization, interior point methods, penalty and barrier approaches, etc. Dr. Andrei is a founding member of The Operations Research Society of Romania, member of the Editorial Board of Computational Optimization and Applications - an International Journal, and of IFAC/WG2 - Control Applications of Optimization.