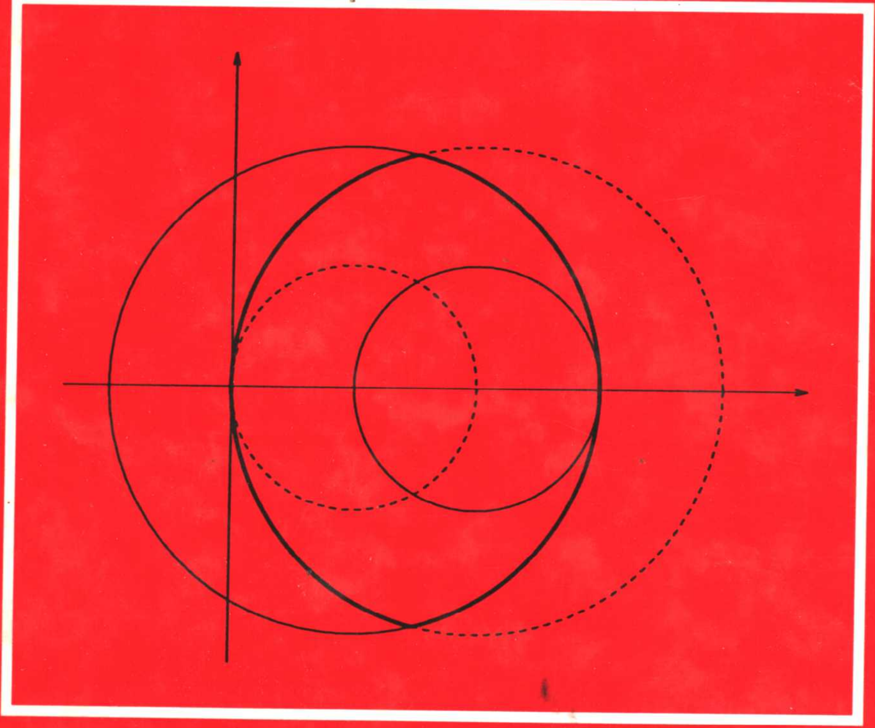
Owe Axelsson

Iterative Solution Methods



ITERATIVE SOLUTION METHODS

OWE AXELSSON

Faculty of Mathematics and Informatics
University of Nijmegen, The Netherlands

CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1994

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
Axelsson, O., 1934-
Iterative solution methods / O. Axelsson.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-44524-8
1. Iterative methods (Mathematics) I. Title.
QA297.8.A94 1993
411'.42–dc20 93-10658 CIP

A catalogue record for this book is available from the British Library.

ISBN 0-521-44524-8 hardback

Preface

Algorithms for the solution of linear systems of algebraic equations arise in one way or another in almost every scientific problem. This happens because such systems are of such a fundamental nature. For example, nonlinear problems are typically reduced to a sequence of linear problems, and differential equations are discretized to a finite dimensional system of equations.

The present book deals primarily with the numerical solution of linear systems. The solution algorithms considered are mainly iterative methods. Some results related to the estimate of eigenvalues (of importance for estimating the rate of convergence of iterative solution methods, for instance), are also presented. Both the algorithms and their theory are discussed. Many phenomena that can occur in the numerical solution of the above problems require a good understanding of the theoretical background of the methods. This background is also necessary for the further development of algorithms. It is assumed that the reader has a basic knowledge of linear algebra such as properties of sets of linearly independent vectors, elementary matrix algebra, and basic properties of determinants.

The first six or seven chapters and Appendix A can be (and have been) used as a textbook for an introductory course in numerical linear algebra, but this material demands students who are not afraid of theory. The theory is presented so that it can be followed even in selfstudy. Chapters 2, 3, and 4 contain much theoretical background for numerical linear algebra, the more difficult parts of which have been indicated by an asterisk. Some readers may wish to postpone study of some of this material until it is used in later chapters. To further help the reader, the definitions in each chapter are collected at the beginning of the chapter and appear in addition in the text at relevant places. Each of the above chapters and Appendix A contain a great number of exercises that either further illustrate the theory and the algorithms or, in many cases, give a fully programmed, step-by-step presentation of methods not treated in the text itself. These latter exercises can be given as homework assignments to students who want, for example, to take a supplementary course for extra credits. Topics discussed in these exercises include generalized inverses and singular values, the Aasens factorization algorithm, modified Gram-Schmidt orthogonalization, QR factorization methods, generalized eigenvalue problems, constrained optimization problems, and logarithmic norms.

The second half of the book presents recent results in the iterative solution of linear systems, mainly using preconditioned conjugate gradient methods. The latter have become well-established techniques since their early but rare use in the sixties and the beginning of the seventies. These chapters give research- or application-oriented students a thorough background that enables them not only to use these algorithms but also to derive and analyse algorithms for new types of problems and make them able, for instance, to read and apply new algorithms presented in numerical linear algebra journals.

This book has undergone many rewritings and has been developed over a long period of the author's teaching of numerical linear algebra. It reflects material that he has found most important or interesting, but this does not mean that topics not covered in the book cannot find use in practice. For instance, clearly many algorithms for the solution of eigenvalue problems and for linear least squares problems not discussed in the book are frequently used in practice. A thorough treatment of them would, however, require a second volume. Readers interested in these topics can consult the reference lists following each chapter.

Also in the interest of keeping the present volume at a reasonable size, certain solution methods that have attracted much recent interest, such as multilevel methods and domain decomposition methods, have not been presented.

Finally, any comments readers have on the contents of this book would be highly appreciated by the author.

Acknowledgments

During the writing of this book, the author benefited from the kind assistance of many colleagues, friends, and students. First of all I want to mention Allan Barker of the Technical University, Lyngby, Denmark, who made many valuable suggestions regarding presentation of the material and offered assistance in correct use of the English language. Radim Blaheta of the Mining Institute at Ostrava, Czech Republic; Ludwig Elsner of the University of Bielefeld, Germany; Igor Kaporin and Michail Makarov, both of the University of Moscow, Russia; Ivo Marek of the Charles University at Prague, Czech Republic; and Wim Lenferink, Hao Lu, Maya Neytcheva, and Ben Polman of the University of Nijmegen read parts of the manuscript and supplied the author with many helpful comments. In addition, Ben Polman and Maya Neytcheva helped with incorporating the figures into the text.

Of my students, Robert Hardin and Michael Bekker identified many opportunities for improvement of the manuscript.

It is a pleasure to express my gratitude to the secretarial staffs of the Department of Mathematics at the University of Nijmegen, whose expertise was indispensable in the writing of this book. In particular, Willy van de Sluis and Tarcise Bökkerink devoted countless hours to typing and retyping large parts of the manuscript.

Finally, I am indebted to the staff of Cambridge University Press and their associates for their fine work and cooperation throughout the publication process.

Owe Axelsson

Nijmegen
January 1993

Contents

Appendices:

1

Direct Solution Methods

The need to solve large linear systems of algebraic equations arises in almost any mathematical model, as illustrated in the instances below. In particular, we go into some detail regarding electrical networks. The most common method used to solve such linear systems is based on factorization of the matrix in triangular factors, which is discussed and shown to be equivalent to the Gaussian elimination method (learned in almost any elementary course in algebra).

We present this method in a form that can be the basis for a computer algorithm. When one wants to solve very large systems of equations, it is important to know how computational complexity grows with the size of the problem. This topic is also addressed, including the case where the matrix has a special structure in the form of a bandmatrix. The solution of tridiagonal systems is considered in particular detail. In addition, some basic dimension theory for matrices, such as the relation between rank and the dimension of the nullspace, is discussed and derived using the factored form of the matrix.

The following definitions or notations are introduced in this chapter:

Definition 1.1

(a) The *range* of the mapping $\widetilde{A} : \mathbb{R}^n \to \mathbb{R}^m$ is

$$\mathcal{R}(\widetilde{A}) = \{\tilde{\mathbf{y}} = \widetilde{A}\tilde{\mathbf{x}};\ \ \tilde{\mathbf{x}} \in \mathbb{R}^n\}.$$

(b) The *nullspace* of $\widetilde{A}$ is $\mathcal{N}(\widetilde{A}) = \{\tilde{\mathbf{x}} \in \mathbb{R}^n;\ \ \widetilde{A}\tilde{\mathbf{x}} = \mathbf{0}\}$.

Definition 1.2

(a) If $\mathbf{b} \in \mathcal{R}(A)$, then $A\mathbf{x} = \mathbf{b}$ is said to be a *consistent* system of linear equations. Otherwise, (if $\mathbf{b} \notin \mathcal{R}(A)$), it is said to be *inconsistent*.

(b) If A is square of order n and $\dim \mathcal{N}(A) = 0$, then A is said to be *nonsingular* or *regular*.

1.1 Introduction: Networks and Structures

A network consists of a set of nodes and a set of edges connecting certain pairs of nodes. Each node is connected to at least one other node. In a physical network, nodes are connected by some device, such as resistors in an electrical network, pipes in a gas pipeline network, and bars, beams, or similar devices in a frame structure.

A source, such as an electromotive voltage, gaswell, or outer pressure, is present to drive the currents, gasflow, or stresses (strains), respectively, through the network. A linear system of algebraic equations, usually with the same number of equations as unknowns, arises. The unknowns may be the potential—i.e., voltage or pressure at the nodes—or the rate of exchange of the potential along the edges (current, strains).

Example 1.1 (Electric network) To be specific, consider the case of an electric network consisting of a set V of nodes (vertices) and a set $L = \{(i, j)\}$ of edges, where (i, j) denotes the edge connecting nodes i and j. In general, not all nodes are connected. At a subset $V_0 \subset V$, the voltage is prescribed. The remaining set of nodes, $V \setminus V_0$, are called "free." Given resistances $r_{i,j}$ at the edges $(i, j) \in L$, we want to find the resulting voltage v_i at the free nodes.

A remarkable phenomenon in nature is the principle of minimal energy loss. Applied in the present context, it means that the distribution of electrical currents in the network will be such that total heat loss is minimized. As the heat loss along edge (i, j) is $(v_i - v_j)^2/r_{i,j}$, this means that

$$\sum_{(i,j)\in L} (v_i - v_j)^2/r_{i,j},$$

where v_i takes the given values for all $i \in V_0$, is minimized. This is a real valued function $f(v_1, v_2, \ldots, v_N)$ of the variables v_i at the free nodes, assuming N such nodes. Taking the partial derivatives with respect to these variables, we get the stationary equations

$$(1.1)\quad \frac{\partial f}{\partial v_i} = 2 \sum_{j,(i,j)\in L} \frac{1}{r_{i,j}}(v_i - v_j) = 0, \qquad i \in V \setminus V_0 = \{1, 2, \ldots, N\}.$$

This is, in fact, Kirchoff's law of electrical currents, implying that the sum of all currents entering and leaving a (free) node is zero. It gives us a system of N linear equations in the N unknowns $\{v_i\}_{i=1}^N$. Later in this book it shall be proved that the corresponding matrix is nonsingular. (It is a so-called irreducibly diagonally dominant matrix with diagonal entries d_i, where

$d_i = \sum_{j,(i,j)\in L} r_{i,j}^{-1}$, and with off-diagonal entries $-r_{i,j}^{-1}$, $i \neq j$, $i \in V \setminus V_0$, with strong inequality, $d_i > \sum_{k=1}^{N} r_{i,k}^{-1}$, for any free node i connected to a constrained node, $j_0 \in V_0$.)

Hence this system has a unique solution. Let the network consist of n nodes and m edges and let $1, 2, \ldots, n$ be a numbering (ordering) of the nodes and $1, 2, \ldots, m$ a numbering of the edges. In (1.1) we sum over the nodes. An alternative and interesting way to express (1.1) is by summing over the edges (cf. Strang [1986]). Let B be the set of edges (branches) and let $B_i^{(+)}$ and $B_i^{(-)}$ be the subsets of B of branches entering node i with $k > i$ and $k < i$, respectively, where k is the other node number of the branch.

Let $\tilde{x}_i$ be the current in branch i, directed from the lower node index to the higher. (Hence $\tilde{x}_i$ may be negative.) Similarly let $\tilde{v}_i$ be the potential difference at branch i, that is, $\tilde{v}_i = v_{j_i} - v_{k_i}$, where k_i, j_i, $k_i > j_i$ are the nodes of branch i. Hence

$$\tilde{\mathbf{v}} = E\mathbf{v},$$

where E is a matrix of order $m \times n$, which has zero entries everywhere except two entries per row, one entry $+1$ and the other -1.

Ohm's law states that

$$R\tilde{\mathbf{x}} = \tilde{\mathbf{v}}, \quad \text{or } \tilde{\mathbf{x}} = R^{-1}\tilde{\mathbf{v}},$$

where R is a diagonal matrix with entries r_{k_i,j_i}, $i = 1, 2, \ldots, m$. (1.1) can also be written

$$\sum_{j\in B_i^{(-)}} \tilde{x}_j - \sum_{j\in B_i^{(+)}} \tilde{x}_j = 0, \quad i = 1, 2, \ldots, n,$$

which is readily seen to be equivalent with

$$E^T\tilde{\mathbf{x}} = \mathbf{0},$$

where E^T is the transpose of E.

The above relations can be described in a diagram as shown below:

$$\begin{tikzcd}
\boxed{\text{potential } \mathbf{v} \text{ at nodes}} \arrow[r, "E^T R^{-1} E\mathbf{v}"] \arrow[d, "E"'] & \boxed{\mathbf{0} = E^T\tilde{\mathbf{x}} = E^T R^{-1}\tilde{\mathbf{v}} = E^T R^{-1} E\mathbf{v}} \\
\boxed{\text{potential difference } \tilde{\mathbf{v}} \text{ at branches}} \arrow[r, "R^{-1}", "\text{Ohm's law}"'] & \boxed{\tilde{\mathbf{x}} = R^{-1}\tilde{\mathbf{v}}} \arrow[u, "E^T"']
\end{tikzcd}$$

By direct computation it can be shown that the matrix $A = E^T R^{-1} E$, which has order $n \times n$, has positive diagonal and nonpositive off-diagonal entries. Note also that $A\mathbf{e} = \mathbf{0}$, where $\mathbf{e} = (1, 1, \ldots, 1)^T$ because $E\mathbf{e} = \mathbf{0}$. In addition, $A = A^T$, so A is symmetric. On the other hand, given any symmetric matrix A of order $n \times n$ with positive diagonal and nonpositive offdiagonal entries, where $A\mathbf{e} = \mathbf{0}$, we can associate a network with A where the branches between two nodes i, j correspond to the nonzero entries of A. It can be shown that by letting R^{-1} hold all the values of the nonzero offdiagonal entries of A and letting E have entries 0 and one entry $+1$ and one -1 in positions corresponding to the nonzero elements of A, we can make $A = E^T R^{-1} E$. In fact, the so-called matrix graph of A and the network are topologically identical. (For a definition of matrix graphs, see a discussion later in this chapter and in Chapter 4.)

Remark Similar systems are found in networks of gas pipelines and frame structures, for example. In general, $r_{i,j}$ depends on the current (or the corresponding variable) through the edge (i, j), and (1.1) is then a nonlinear system of equations. When this dependence is neglected, however, (1.1) becomes a linear system.

Example 1.2 (Tomography) Consider a plate of *inhomogeneous* materials which is part of a three-dimensional body. It is not possible to observe the plate from above, but only from the boundary. This would be the case if we want to observe a plane section through a human body, for instance. Assume for simplicity that the plane is square and subdivided into n^2 small squares, say 9 (i.e., $n = 3$). (See Figure 1.1.) By sending X-rays with known intensities I_0 through the plate and measuring the intensity of the outgoing X-ray, we want to determine the damping factors x_i in the different squares.

This technique, called *tomography*, has been used since about 1973 in medicine (in cancer research, for instance). It is effective because each tissue (material within each square) has its own damping factor. If an X-ray is sent through the three top squares in Figure 1.1, we get

$$I_1 = I_0 e^{-x_1} e^{-x_2} e^{-x_3} = I_0 e^{-(x_1+x_2+x_3)},$$

assuming that the damping factor is exponential.

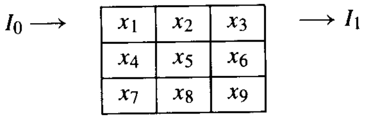


Figure 1.1. X-rays through a plate

Let $I_1 = I_0 e^{-b}$. Since I_1 is measured, we can determine b, and we have the linear relation $x_1 + x_2 + x_3 = b$. If we send three horizontal and three vertical X-rays we get the following linear algebraic system:

$$\begin{array}{ccccccccccccccccc}
x_1 & + & x_2 & + & x_3 & & & & & & & & & & & & = b_1 \\
 & & & & & & x_4 & + & x_5 & + & x_6 & & & & & & = b_2 \\
 & & & & & & & & & & & & x_7 & + & x_8 & + \ x_9 & = b_3 \\
x_1 & & & & & + & x_4 & & & & & + & x_7 & & & & = b_4 \\
 & & x_2 & & & & & + & x_5 & & & & & + & x_8 & & = b_5 \\
 & & & & x_3 & & & & & + & x_6 & & & & & + \ x_9 & = b_6.
\end{array}$$

Hence we have six equations but nine unknowns—i.e., the system is underdetermined. Accordingly, it has no unique solution. We may send five more X-rays, now through the diagonals. The system then becomes overdetermined. In practice one sends X-rays along directions incremented by a small angle. A unique solution may be determined, for instance, by a least squares approximation method.

Example 1.3 (Diffusion) Consider a tube with a liquid of concentration x_i, $i = 0, 1, \ldots, n+1$, in $n+2$ different cells. Initially the cells are assumed to be closed by impervious walls and to have concentration a_i, $i = 0, 1, \ldots, n+1$. At time $t = 0$ the walls become permeable and the concentration begins to diffuse between the cells. Assume that the left and right endcells have a fixed concentration $a_0 = 0$ and $a_{n+1} = 1$, respectively (see Figure 1.2).

To find the concentration in the other cells, we use the fact that the rate of diffusion $dx_i(t)/dt$ for a certain time is proportional to the difference of concentrations at time t, i.e. (for $n = 3$)

$$\frac{dx_1(t)}{dt} = c[(x_2 - x_1) - (x_1 - x_0)]$$

$$\frac{dx_2(t)}{dt} = c[(x_3 - x_2) - (x_2 - x_1)]$$

$$\frac{dx_3(t)}{dt} = c[(x_4 - x_3) - (x_3 - x_2)],$$

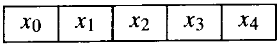


Figure 1.2. Diffusion in a tube ($n = 3$)

where c is a positive constant. This is a linear system of ordinary differential equations of first order. In matrix form we get

$$\frac{d\mathbf{x}(t)}{dt} = c[A\mathbf{x}(t) + \mathbf{f}], \quad t > 0, \quad x_i(0) = a_i, \quad i = 1, 2, \ldots, n,$$

where

$$A = \begin{bmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix}, \quad \mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} \quad \text{and} \quad \mathbf{f} = \begin{bmatrix} a_0 \\ 0 \\ a_4 \end{bmatrix}.$$

This is an initial value problem. It can readily be seen that the solution $\mathbf{x}(t)$ has a *steady state* solution $\mathbf{x}(\infty) = \mathbf{b}$, satisfying $A\mathbf{b} + \mathbf{f} = \mathbf{0}$ (because $d\mathbf{x}/dt = \mathbf{0}$ then). From this we easily find $b_i = i/(n+1)$, $i = 1, 2, 3$. Note in passing that the steady state, or state of equilibrium, is independent of the initial values a_1, a_2, and a_3.

Discrete diffusion systems are examples of more general compartment models, which are discussed next.

Example 1.4 (Compartment models) Compartment models are devices for describing the circulation of various elements in nature, in the human body, or in other systems. In a compartment system we have n compartments $X_1, X_2, \ldots, X_n$, which contain quantities (concentrations) $x_1, x_2, \ldots, x_n$ of some "matter." For natural reasons we must have $x_i \geq 0$ for all i. Then there is a "rule," a differential equation, which governs the circulation of matter among the compartments, as well as leakage from one or more compartments and also a possible external supply of matter to the system.

Let us consider three examples. First, diffusion and circulation of carbon dioxide in nature can be discussed within the framework of compartment models. A diffusive process can be described by compartment models; for a one-dimensional example, see Figure 1.2.

Second, in pharmacology the distribution of a drug in the human body is often discussed in terms of a compartment system. The various organs of the human body and the circulatory system are considered compartments. In medical compartment models, one often assumes that the transport between two compartments is governed by the so-called Fick's law, which leads to linear differential systems.

Third, there is an analogy between continuous-time Markov chains with a finite number of states and closed compartment systems (i.e., systems with no loss of mass). For a reference on compartmental analysis, see Jacquez (1972).