Methods of Numerical Mathematics

G. I. Marchuk

G. I. Marchuk

Methods of Numerical Mathematics

translated by Jiri Ružička

Springer-Verlag
New York Heidelberg Berlin
1975

AMS Subject Classification
65-01

Library of Congress Cataloging in Publication Data

Marchuk, Guriĭ Ivanovich.
Methods of numerical mathematics.

(Applications of mathematics; v. 2)
Translation of Metody vychislitel'noi matematiki.
Bibliography: p. 295
Includes index.
1. Numerical analysis. I. Title.
QA297.M3413 519.4 75-15566

The original Russian edition METODY VYCHISLITEL'NOI MATEMATIKI was published in 1973 by NAUKA, Novosibirsk.

Printed in the United States of America.

ISBN 0-387-90156-6 Springer-Verlag New York Heidelberg Berlin

ISBN 3-540-90156-6 Springer-Verlag Berlin Heidelberg New York

Applied Probability
Control
Economics
Information and Communication
Modeling and Identification
Numerical Techniques
Optimization

Applications of Mathematics

2

Preface to the English Edition

The English translation of the present book fully corresponds to the Russian original version. The minor changes that have been made consist only of some remarks on the text and the correction of misprints that have been discovered. The author hopes that the English edition will help foreign readers obtain first hand knowledge of the methods in computational mathematics being developed in the Soviet Union and become aware of some scientific trends resulting from the necessity of solving complicated problems by reducing them to elementary ones.

The author expresses his gratitude to Professor I. Babuška who made a number of remarks on the book that have been considered in its English version. He would also like to acknowledge his deep gratitude to Dr. J. Ružička, the translator of the book, and to Professor A. V. Balakrishnan, head of the System Science Department at UCLA, who called attention to the need for an English language edition. The author appreciates very much the cooperation of Springer-Verlag who, with the present book, started a new series of monographs on mathematical systems and economics.

Preface

The present volume is an adaptation of a series of lectures on numerical mathematics which the author has been giving to students of mathematics at the Novosibirsk State University during the span of several years. In dealing with problems of applied and numerical mathematics the author sought to focus his attention on those complicated problems of mathematical physics which, in the course of their solution, can be reduced to simpler and theoretically better developed problems allowing effective algorithmic realization on modern computers.

It is usually these kinds of problems that a young practicing scientist runs into after finishing his university studies. Therefore this book is primarily intended for the benefit of those encountering truly complicated problems of mathematical physics for the first time, who may seek help regarding rational approaches to their solution.

In writing this book the author has also tried to take into account the needs of scientists and engineers who already have a solid background in practical problems but who lack a systematic knowledge in areas of numerical mathematics and its more general theoretical framework.

Consequently, the author has selected a form of exposition which in his opinion helps to attract the attention of a wide range of researchers to problems of numerical mathematics. This style has required certain concessions in the exposition, thus allowing concentration only on basic ideas and approaches. As for the details (sometimes important) and the possible generalizations (such as minimal smoothness requirements, constraints on the input data, etc.), they are obvious to the specialist and present useful exercises for a beginner.

Chapter 8 is an expanded version of the paper given by the author at the International Congress of Mathematicians in Nice (1970). This chapter gives some idea both of the material considered in the previous chapters, and of various methods and problems of numerical mathematics that are of fundamental importance but have not found their way into this volume.

In the process of preparation for publication this book has undergone considerable changes in response to advice and comments obtained by the author from his colleagues and associates. Those whose help is gratefully acknowledged include M. M. Lavrentiev, V. I. Lebedev, I. Marek, M. K. Fage, and N. N. Yanenko. They have made a number of constructive comments regarding the exposition of individual chapters, especially the first and fifth. The changes in the second chapter, which are due to Yu. A. Kuznetsov, are so profound that the nature of his contribution in this part is essentially that of coauthorship. The author has also enjoyed valuable advice and comments from V. T. Vasil'ev, V. P. Il'in, A. N. Konovalov,

V. P. Kochergin, V. V. Penenko, V. V. Smelov, U. M. Sultangazin, and others. G. S. Rivin did considerable work in editing the manuscript. To all these, as well as M. S. Yudin who took part in preparing the book for publication, the author expresses his deep gratitude.

Contents

Chapter 3

Methods for Solving Stationary Problems of Mathematical Physics

Chapter 4

Methods for Solving Non-Stationary Problems

Chapter 5

Numerical Methods for Some Inverse Problem

Chapter 6

The Simplest Problems of Mathematical Physics

Chapter 7

Numerical Methods in the Theory of Radiative Transfer

Chapter 8

A Review of the Methods of Numerical Mathematics

Chapter 1

Fundamentals of the Theory of Difference Schemes

This chapter briefly surveys those fundamentals of the theory of difference schemes that are used extensively in the following chapters. We restrict our theoretical considerations to the simplest and most easily interpreted cases, since our main purpose is to achieve familiarity with certain modern concepts in the construction of numerical algorithms in mathematical physics. For more refined and more complex theoretical developments we refer the reader to the specialized bibliography given at the end of the book.

1.1. Basic Equations and their Adjoints

Let us consider a region D in the n-dimensional Euclidean space E_n. Observe at the very outset that the regions usually encountered in applied mathematics are of such a structure that they will possess a measure: area in the two-dimensional case, volume in three dimensions and so on. Nevertheless the theory of Lebesgue measure is vital in the subsequent definitions and, thus, the reader is assumed to be familiar with measure theory and the Lebesgue integral. (Smirnov [2], Sobolev [1], Vladimirov [2], Natanson [1], and others.)

Next let us define the Hilbert space $L_2(D)$ of all real measurable square integrable functions:

$$\int_D f^2(x)\,dx < \infty,$$

with the inner product

$$(f, g) = \int_D f(x)g(x)\,dx. \tag{1.1}$$

As usual, the norm of the function $f \in L_2(D)$ is defined by

$$\|f\|^2 = (f, f). \tag{1.2}$$

Let us now choose a subspace (a linear manifold) Φ of the Hilbert space $L_2(D)$ by imposing certain additional conditions which every element $\phi \in \Phi$ must satisfy. For example, we may require some specified smoothness conditions, conditions on the limit behavior at the boundary D, etc. These conditions, however, must be sufficient to guarantee that an operator A, if given, maps the subspace Φ into $L_2(D)$.

A linear operator A, defined on the linear manifold Φ, is called positive semidefinite if

$$(A\phi, \phi) \geqq 0 \tag{1.3}$$

for all $\phi \in \Phi$, with the equality sign possibly holding for a nonzero element ϕ. It is customary to write $A \geqq 0$ in this case. If the equality sign above can not hold for nonzero elements, that is

$$(A\phi, \phi) > 0, \qquad \phi \neq 0, \tag{1.4}$$

then we say that the operator A is positive and write $A > 0$. Finally in the case of the stronger inequality

$$(A\phi, \phi) \geqq \gamma(\phi, \phi), \qquad \phi \in \Phi, \tag{1.5}$$

where $\gamma > 0$ is a positive constant independent of ϕ, the operator A is called positive definite.

Note that if A is a positive symmetric matrix, then it is positive definite (Faddeev, Faddeeva [8]).

The subspace Φ will be called the *domain* of the operator A and denoted by $\Phi(A)$.

Consider next the adjoint operator A^* defined by the Lagrange identity

$$(Ag, h) = (g, A^*h), \tag{1.6}$$

where $g \in \Phi(A)$, $h \in \Phi(A^*)$.

The subspaces $\Phi(A)$ and $\Phi(A^*)$ of the Hilbert space $L_2(D)$ do not coincide in general, despite the fact that their elements are defined on the same region D in E_n. In what follows we will assume that the adjoint operator exists and is *closed* in the following sense:

Consider a sequence $\phi_n^* \to \psi$ and let $A^*\phi_n^* \to \chi$. Then $\psi \in \Phi(A^*)$ and the limit relation $A^*\psi = \chi$ holds. The operator A is called *selfadjoint* if $A = A^*$ and $\Phi(A) = \Phi(A^*)$.

Let us note one important consequence regarding the properties of adjoint operators. Namely, if $\Phi(A) \equiv \Phi(A^*)$, then $A > 0$ implies $A^* > 0$.

A considerable role in analyzing algorithms is played by the Fourier expansions with respect to the eigenfunctions of operators and their adjoints.

Consider the following two spectral problems for $A \geqq 0$:

$$Au = \lambda(A)u, \qquad A^*u^* = \lambda(A^*)u^*. \tag{1.7}$$

Assume that each of the homogeneous equations (1.7) generates a complete set of eigenfunctions, $\{u_n\}$ and $\{u_n^*\}$, which are normalized as follows:

$$(u_n, u_m^*) = \begin{cases} 1, n = m, \\ 0, n \neq m, \end{cases} \tag{1.8}$$

and the corresponding eigenvalues $\lambda_n(A)$ belong to the interval

$$\alpha(A) \leq \lambda_n(A) \leq \beta(A).$$

This complete set of eigenfunctions will be called a *biorthogonal basis.* Thus, under the assumption of completeness, arbitrary functions $f \in \Phi$ and $f^* \in \Phi^*$ can be represented in the form of a Fourier series

$$f = \sum_n f_n u_n, \qquad f^* = \sum_n f_n^* u_n^* \tag{1.9}$$

where

$$f_n = (f, u_n^*), \qquad f_n^* = (f^*, u_n). \tag{1.10}$$

(In what follows, we use Φ, Φ^* instead of $\Phi(A)$, $\Phi(A^*)$ for the sake of simplicity).

Of great importance in the analysis of numerical algorithms are the norm estimates of operators. A *norm of an operator A* is defined as follows:

$$\|A\|^2 = \sup_{\substack{\phi \in \Phi \\ \phi \neq 0}} \frac{(A\phi, A\phi)}{(\phi, \phi)} \tag{1.11}$$

(in order to simplify notation the qualification $\phi \neq 0$ will not be explicitly mentioned again). Since

$$(A\phi, A\phi) = (\phi, A^*A\phi),$$

the square of the norm of A can also be expressed in the following way:

$$\|A\|^2 = \sup_{\phi \in \Phi} \frac{(\phi, A^*A\phi)}{(\phi, \phi)}. \tag{1.12}$$

The operator A^*A is symmetric and positive semidefinite. Consider the spectral problem

$$A^*A\Omega = \lambda(A^*A)\Omega \tag{1.13}$$

This problem defines a family of eigenfunctions $\{\Omega_n\}$ and eigenvalues $\lambda_n(A^*A) \geqq 0$. We will assume that $\{\Omega_n\}$ is a complete set. Then the function ϕ has the following Fourier expansion:

$$\phi = \sum_n \phi_n \Omega_n, \tag{1.14}$$

where

$$\phi_n = (\phi, \Omega_n). \tag{1.15}$$

Using the orthonormality of the functions Ω_n, the substitution of Series (1.14) into (1.12) yields

$$\|A\|^2 = \sup_{\{\phi_n\} \in Q} \frac{\sum_n \lambda_n(A^*A)\phi_n^2}{\sum_n \phi_n^2}, \tag{1.16}$$

where Q is the space of Fourier coefficients. It is easy to see that

$$\frac{1}{\|A^{-1}\|^2} = \lambda_{\min}(A^*A) = \alpha(A^*A),$$
$$\|A\|^2 = \lambda_{\max}(A^*A) = \beta(A^*A), \tag{1.17}$$

where $\lambda_{\min}$ is the smallest and $\lambda_{\max}$ is the greatest eigenvalue respectively in the set $\{\lambda_n(A^*A)\}$ for (spectral) Problem (1.13). The quantity $\beta(A^*A) = \lambda_{\max}(A^*A)$ is usually called the *spectral radius* of the operator A^*A. In general, the spectral radius is defined as $\beta(A) = \sup\{|\lambda(A)|\}$. Note that for $\lambda(A) > 0$ the spectral radius $\beta(A) = \sup\{\lambda(A)\}$.

In the case of a selfadjoint operator A consider the spectral problem

$$Au = \lambda u. \tag{1.18}$$

We have

$$\|A\| = \beta(A). \tag{1.19}$$

It is not difficult to see that for a selfadjoint operator

$$\beta(A^2) = [\beta(A)]^2. \tag{1.20}$$

Consider a fixed closed positive operator C on the Hilbert space $L_2(D)$. We will call it the *energy operator*. Thus

$$(C\phi, \phi) > 0 \tag{1.21}$$

for all $\phi \in \Phi$, the domain of C dense in $L_2(D)$. In other words, for any element $f \in L_2(D)$ there is an element $g \in \Phi$ such that $\|f - g\| \leqq \varepsilon$, where ε is an arbitrarily small, positive constant. Denote by $\Phi^* = \Phi(C^*)$ the domain of definition of the adjoint operator C^*. Assume Φ^* coincides with Φ. Then $C^*\phi$ exists for all $\phi \in \Phi$ and $(C^*\phi, \phi) = (\phi, C\phi) = (C\phi, \phi)$. Consequently $(C\phi, \phi) = (\frac{1}{2}[C + C^*]\phi, \phi)$, where $\frac{1}{2}[C + C^*]$ is now a symmetric, positive operator. This allows one to introduce a new inner product in Φ, namely

$$(f, g)_C = (Cf, g),$$

and the norm

$$\|\phi\|_{\bar{C}}^2 = (C\phi, \phi) = (\bar{C}\phi, \phi),$$

where $\bar{C} = \frac{1}{2}[C + C^*]$. This norm will be called the *energy norm*. One can obtain the following significant estimate:

$$\|\phi\|_C^2 = \|\phi\|_{\bar{C}}^2 \leq \|\bar{C}\|\,\|\phi\|^2 = \beta(\bar{C})\|\phi\|^2, \tag{1.22}$$

where $\beta(\bar{C})$ is the largest eigenvalue of the operator $\bar{C}$.

In conclusion let us note that in dealing with problems of mathematical physics and their adjoints it is often convenient to use functions from the *Sobolev space* $W_2^l(D)$. This space is a Hilbert space of $L_2(D)$ functions whose generalized derivatives up to and including lth order are square integrable in

D. The inner product in such a space is defined by the formula (see Sobolev [1], Vladimirov [2])

$$(u, v)_{W_2^l} = \sum_{k=0}^{l} \sum_{(k)} \int_D \frac{\partial^k u}{\partial x^k} \frac{\partial^k v}{\partial x^k} dD. \tag{1.23}$$

Here we have used the following notation for the partial derivatives:

$$\frac{\partial^k \phi}{\partial x^k} = \frac{\partial^{\alpha_1 + \cdots + \alpha_n}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} \phi, \alpha_1 + \cdots + \alpha_n = k.$$

The norm in the space $W_2^l(D)$ is defined by the relation

$$\|\phi\|_{W_2^l}^2 = (\phi, \phi)_{W_2^l}. \tag{1,24}$$

1.1.1. Norm Estimates of Certain Matrices

Let us consider a positive semidefinite matrix $A \geqq 0$ on the Euclidean space. Then for any value of the parameter $\sigma \geqq 0$ we have the following relation:

$$\|(E + \sigma A)^{-1}\| \leqq 1. \tag{1.25}$$

For the proof of this important proposition we exploit the formula

$$\|(E + \sigma A)^{-1}\| = \max_{\phi} \frac{((E + \sigma A)^{-1}\phi, (E + \sigma A)^{-1}\phi)}{(\phi, \phi)}. \tag{1.26}$$

Let us introduce new elements

$$\psi = (E + \sigma A)^{-1}\phi.$$

Then

$$\|(E + \sigma A)^{-1}\| = \max_{\psi} \frac{(\psi, \psi)}{((E + \sigma A)\psi, (E + \sigma A)\psi)}$$

$$= \frac{1}{\min_{\psi}\left[1 + 2\sigma \frac{(A\psi, \psi)}{(\psi, \psi)} + \sigma^2 \frac{(A\psi, A\psi)}{(\psi, \psi)}\right]}.$$

Since $A \geqq 0$ on the elements ϕ, ψ, the last relation implies (1.25). If $A > 0$, then for $\sigma > 0$ we have immediately

$$\|(E + \sigma A)^{-1}\| < 1. \tag{1.27}$$

Kellogg's lemma [15]. For any matrix $A \geqq 0$ and for any $\sigma \geqq 0$ one has

$$\|(E - \sigma A)(E + \sigma A)^{-1}\| \leq 1. \tag{1.28}$$

For the proof let us define T by

$$T = (E - \sigma A)(E + \sigma A)^{-1},$$

and consider the expression for $\|T\|^2$:

$$\|T\|^2 = \max_{\phi} \frac{((E - \sigma A)(E + \sigma A)^{-1}\phi, (E - \sigma A)(E + \sigma A)^{-1}\phi)}{(\phi, \phi)}$$

$$= \max_{\psi} \frac{((E - \sigma A)\psi, (E - \sigma A)\psi)}{((E + \sigma A)\psi, (E + \sigma A)\psi)}$$

$$= \max_{\psi} \frac{(\psi, \psi) - 2\sigma(A\psi, \psi) + \sigma^2(A\psi, A\psi)}{(\psi, \psi) + 2\sigma(A\psi, \psi) + \sigma^2(A\psi, A\psi)} \leq 1.$$

Here the crucial role has been played by the positive semidefinitness of the matrix A. The lemma is proved.

In the case when the matrix A is positive and $\sigma > 0$, the expression (1.28) is replaced by

$$\|(E - \sigma A)(E + \sigma A)^{-1}\| < 1. \tag{1.29}$$

1.1.2. Computing the Spectral Bounds of a Positive Matrix

Consider the problem of finding the largest and smallest eigenvalues of a matrix $A > 0$ with a positive spectrum. The approach below is due to Lyusternik [4].

Assume that the spectral problem

$$Au = \lambda u \tag{1.30}$$

defines a complete set of eigenfunctions $u_k \in \Phi$, and a set of eigenvalues $\lambda_k(A)$. (A fairly complete treatment of spectral problems can be found in the papers by Marek [8].) Consider the iterative process

$$\phi^{(n+1)} = (1/c_n)A\phi^{(n)},$$
$$\phi^{(0)} = g,$$

where g is an arbitrary nonzero vector, and c_n is a normalizing factor which can be conveniently chosen in the form

$$c_n = \|\phi^{(n)}\| = \sqrt{\sum_p |\phi_p^{(n)}|^2}.$$

Here $\phi_p^{(n)}$ is the pth component of the vector $\phi^{(n)}$. Thus

$$\phi^{(n+1)} = A \frac{\phi^{(n)}}{\|\phi^{(n)}\|}. \tag{1.31}$$

Let $0 < \alpha(A) = \lambda_1 \leqq \cdots \leqq \lambda_{m-1} < \lambda_m = \beta(A)$. Clearly, the following relation holds:

$$\beta(A) = \lim_{n \to \infty} \|\phi^{(n)}\|. \tag{1.32}$$