Computational Methods for Matrix Eigenproblems

Computational Methods for Matrix Eigenproblems

A. R. GOURLAY

Department of Mathematics
Loughborough University of Technology

G. A. WATSON

Department of Mathematics
University of Dundee

JOHN WILEY & SONS

London · New York · Sydney · Toronto

Library of Congress Catalog Card No. 73-2783

ISBN 0 471 31915 5

Printed in Great Britain by Adlard & Son Ltd.
Bartholomew Press, Dorking

Preface

Since 1947, numerical analysis has enjoyed dedicated effort from mathematicians (both pure and applied), computer scientists, engineers and the like, all striving to devise and analyse processes aimed at being of use and value in the solution of the problems of the outside world. One of the topic areas of numerical analysis in which this effort has led to rich rewards is that of computational linear algebra. Our understanding of matrix problems is much more complete than that of almost any other subset of numerical analysis. We have powerful algorithms for solving linear equations, and for tackling nearly all the commonly occurring eigenproblems. More important perhaps is the existence of reliable and robust computer routines for these algorithms. Much of this knowledge is contained in the monumental work by Wilkinson (*The Algebraic Eigenvalue Problem*, O.U.P., 1965). The rest is in the more recent literature.

The topics, numerical solution of linear systems, and numerical solution of matrix eigenproblems, are fundamental to courses in numerical analysis. Therefore, the availability of suitable textbooks for such courses is desirable. For the first topic, the book by Forsythe and Moler (*Computer Solution of Linear Systems*, Prentice Hall, 1967) is very suitable. Its style of short, single topic chapters is one which should appeal to lecturers and students alike. The present text is an attempt to provide a similar book for the second topic.

The content is based on lectures given by the authors to M.Sc. students at the University of Dundee. Some of the material has also been used in undergraduate courses at the same University. All these students would have attended a basic course in matrix algebra and be familiar with the fundamental concepts, for example the definitions of matrix multiplication, inversion and determinants. This book is written for students with such a background. The material is suitable for courses to students both in mathematical disciplines and in the more applied subjects such as engineering, and, because of the particular format in which the material is presented, it could form the basis for any level of course in its subject matter. Each chapter could be covered in one or two lectures, and thus the whole text could be suitable for a course of some fifteen to twenty lectures (including an allocation of time for consideration of some of the relevant chapters in the book by Forsythe and Moler).

The subjects have been chosen so as to present only the more commonly used and more reliable techniques for computing solutions to eigenproblems.

The aim is primarily to describe the techniques; therefore little will be said of the error analysis of each method, although the conclusions to be drawn from the relevant error analyses will of course be stressed. It is hoped that any student whose interest may be roused by the material covered here will turn to the book by Wilkinson for the 'complete story'.

Many colleagues and friends have contributed to the development of this book, particularly in the early stages of preparation. One of the authors (ARG) is especially grateful to Dr. J. Ll. Morris for his helpful criticism of first drafts of several of the chapters, to Professor D. S. Jones not only for the initial encouragement to undertake the preparation of this text but also for arranging the rescue of a foundering manuscript, and to his co-author (GAW) for effecting the rescue.

Both authors wish to express their thanks to Professor D. S. Jones for his advice and comments throughout the preparation of the manuscript, and to Mrs. Hilary Watson and Miss Frances Duncan for the expert typing of the manuscript.

A. R. GOURLAY
G. A. WATSON

Contents

1

Introduction

1.1 Introduction

The aim of this book is to provide an elementary text on the numerical techniques used in the solution of the algebraic eigenproblem

$$A\mathbf{x} = \lambda\mathbf{x}$$

where A is a known $n \times n$ matrix. The scalar λ is referred to variously as eigenvalue, latent root, characteristic value of the matrix A and the n column vector $\mathbf{x}$ as an eigenvector, latent vector, characteristic vector of A. The purpose of this introductory chapter is to provide some examples of the occurrence of such problems in a wide variety of areas of application. Certain standard definitions and conventions of notation are assumed, and the reader unfamiliar with these should refer to Chapter 2, which contains a collection of the standard notations and more useful results (to this text). Each of the following sections is intended to be self-contained and independent.

1.2 A geometrical example

The equation of an ellipsoid in n space dimensions is given, in cartesian form, by

$$\tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}z_iz_j + \sum_{j=1}^{n} b_jz_j + c' = 0,$$

where $\mathbf{z} = (z_1, z_2, \ldots, z_n)^T$ is a point in n dimensional space. Using matrix notation this may be compactly written as

$$\tfrac{1}{2}\mathbf{z}^TA\mathbf{z} + \mathbf{b}^T\mathbf{z} + c' = 0,$$

where c' is a known constant, $\mathbf{b}$ is a known vector and A a known symmetric positive definite matrix. By a suitable translation of axes

$$\mathbf{z} = \mathbf{x} - A^{-1}\mathbf{b},$$

the equation may be simplified to

$$\tfrac{1}{2}\mathbf{x}^TA\mathbf{x} + c = 0.$$

If the position vector of a point **x** on this *hyperellipsoid* is the same as the gradient vector at **x**, then **x** is a principal axis of the hyperellipsoid. Thus the set of principal axes are those directions which simultaneously correspond to a position vector and to a gradient vector. A two-dimensional example will help to clarify this. (Here the two components of **x** are denoted by x and y.)

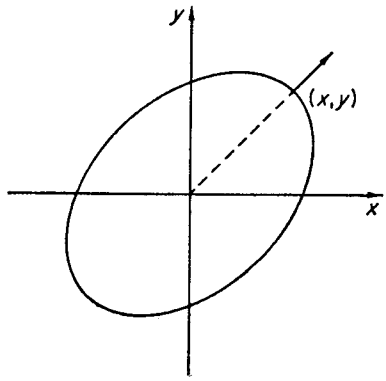

FIGURE 1

In Figure 1 the curve

$$ax^2+2bxy+cy^2 = d$$

is drawn. At a general point (x, y), the gradient is in the direction

$$(ax+by, bx+cy).$$

At the particular point (x, y) in Figure 1 the gradient vector is in the same direction as the position vector from the origin. It follows that at this point (x, y) (and at any similar points) there exists some scalar λ, such that

$$ax+by = \lambda x$$

$$bx+cy = \lambda y,$$

or in matrix notation

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}.$$

From this equation we deduce that the principal axes are given by the eigenvectors of the matrix

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

Returning to our n dimensional example we observe that the gradient vector is given by

$$\mathbf{g} = A\mathbf{x}.$$

Thus the principal axes are given by the n (non-trivial) vectors $\mathbf{x}$ satisfying

$$A\mathbf{x} = \lambda\mathbf{x},$$

that is by the n eigenvectors of A.

1.3 Small vibrations

An area which is a fruitful source of eigenproblems is the study of the vibrations of dynamical, and structural systems. The example given below considers the small vibrations of particles on a string under tension. Simplifying assumptions have been made to ensure that the analysis does not become too complicated. Thus we assume a uniform weightless string, no gravity and that the vibrations are small and in a direction perpendicular to the rest position of the string. We consider specifically the motion of four unequal, but equally spaced, particles on a string under tension F. The system is shown in Figure 2.

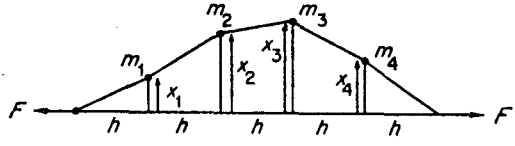

FIGURE 2

Making the standard assumptions, the equations for this system are given by

$$m_1 \frac{d^2x_1}{dt^2} = -F\frac{x_1}{h} + F\left(\frac{x_2 - x_1}{h}\right)$$

$$m_2 \frac{d^2x_2}{dt^2} = -F\left(\frac{x_2 - x_1}{h}\right) + F\left(\frac{x_3 - x_2}{h}\right)$$

$$m_3 \frac{d^2x_3}{dt^2} = -F\left(\frac{x_3 - x_2}{h}\right) - F\left(\frac{x_3 - x_4}{h}\right)$$

$$m_4 \frac{d^2x_4}{dt^2} = +F\left(\frac{x_3 - x_4}{h}\right) - F\frac{x_4}{h}.$$

Defining tne vector $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$ and letting

$$d_i = \frac{m_i h}{F}, \quad i = 1, 2, 3, 4,$$

this system may be written in matrix notation as

$$D\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}t^2} = T\mathbf{x} \tag{1}$$

where D is the diagonal matrix

$$D = \begin{bmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & d_3 & 0 \\ 0 & 0 & 0 & d_4 \end{bmatrix}$$

and T is the tridiagonal matrix

$$T = \begin{bmatrix} -2 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix}.$$

When the system vibrates in a normal mode the equation

$$\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}t^2} = -w^2\mathbf{x} \tag{2}$$

holds. (In this situation the masses all vibrate in phase or in direct opposition.) Substituting (2) in (1), we obtain the eigenproblem

$$Dw^2\mathbf{x} = -T\mathbf{x} \tag{3}$$

for the normal frequencies $w_1, \ldots, w_4$ and the corresponding normal modes. Although this would appear at first sight to be a *generalized eigenproblem* of the form

$$(A - \lambda B)\,\mathbf{x} = 0,$$

it may easily be transformed into the standard symmetric tridiagonal problem

$$D^{-1/2}TD^{-1/2}\mathbf{y} = -w^2\mathbf{y}$$

where $\mathbf{y} = D^{1/2}\mathbf{x}$, since the elements of D are positive.

This model may easily be extended to the general case of n particles on a string, leading to an n dimensional analogue of (3). The matrix T is still tridiagonal. In fact, it is a special matrix which occurs frequently in numerical analysis, and whose eigenvalues are expressible in analytic terms.

1.4 An example in information system design

If we regard an information (both storage and retrieval) system as made up of component subsystems which operate together and perform a set of operations to accomplish the defined purpose of the system, then the aims in the design of such a system may be stated as:

(i) to define the purpose of the system
(ii) to select the component subsystems to achieve this purpose in an optimal way.

In the following analysis of this problem, we shall see that the eigenvectors of a particular matrix play an important role. We begin with a few definitions used in the model.

A *job* is defined to be the purpose of the system and it is composed of a set of *operations* $O_1, \ldots, O_m$ together with a *volume* (or work load) $V_1, \ldots, V_m$ for each operation. A *component* is a well-defined means of performing some parts of these operations. The *efficiency* with which a given component performs the execution of a given operation is measured as a function of cost, time and size of the operation. For example, a typical choice would be

$$e = \frac{ct}{n}$$

where c is the cost in pounds per unit time, t is the time taken and n is a measure of the size of the operation (e.g. bits processed). A *total system* constructed from a set of components $(S_1, \ldots, S_n)$ in order to achieve all the required operations is represented by an $n \times m$ efficiency matrix E with the (i, j) element e_{ij}, where for example,

$$e_{ij} = \frac{c_{ij}t_{ij}}{n_{ij}},$$

is the efficiency with which the ith component performs operation O_j.

Since the above system is designed to perform a specified job made up of the operations together with a volume for each operation, the cost of this system performing the job is

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

where

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$$

$$v = (v_1, v_2, \ldots, v_m)^T.$$

The system and its performance are thus given by E, $\mathbf{v}$ and $\mathbf{x}$. We require

some measure of the performance of the system on various tasks, defined by different volume vectors **v**. One measure is the *Rayleigh quotient*

$$a^2 = \frac{(E\mathbf{v})^T (E\mathbf{v})}{\mathbf{v}^T\mathbf{v}}.$$

The design of a general or total system involves the calculation of the maximum cost of the system, and the maximum of the above measure is given by the largest eigenvalue of E^TE. Further, the corresponding value of **v** giving the critical volume at which this maximum is achieved is given by the corresponding eigenvector of E^TE.

Further details of this area of application can be found, for example, in the book by Becker and Hayes (1967).

1.5 An eigenproblem in non-linear optimization

A basic problem in non-linear optimization is the determination of the n dimensional vector **x** which minimizes the scalar function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$. Assuming that we are able to calculate the gradient of $f(\mathbf{x})$, denoted by $\mathbf{g}(\mathbf{x})$, then we may use a member of the class of *variable metric methods*. These algorithms assume an initial guess or estimate $\mathbf{x}_0$ of the solution and calculate a sequence of new points $\{\mathbf{x}_k\}$ by means of a relation of the form

$$\mathbf{x}_{k+1} = \mathbf{x}_k + t_k\mathbf{d}_k$$

where $\mathbf{d}_k$ is a direction vector and t_k is a positive scalar chosen to minimize

$$f(\mathbf{x}_k + t\mathbf{d}_k)$$

with respect to t—a univariate minimization problem. The variable metric methods are characterized by the use of a direction vector of the form

$$\mathbf{d}_k = -H_k\mathbf{g}_k$$

where $\mathbf{g}_k = \mathbf{g}(\mathbf{x}_k)$ and H_k is a symmetric positive definite matrix. It is beyond the scope of this section to explain in detail the theory of such algorithms.

At each step a new approximate matrix H_{k+1} is computed by a relation of the form

$$H_{k+1} = H_k + E_k, \qquad H_0 = I.$$

In practice, it is essential to ensure that the sequence of matrices $\{H_k\}$ remains positive definite. Whilst the correction E_k is usually chosen so that H_{k+1} will be positive definite if H_k is positive definite, the presence of rounding errors may cause H_{k+1} to become indefinite. J. Greenstadt has suggested a means of ensuring positive definiteness which involves a complete eigenanalysis of

H_{k+1}. If $\{\lambda_i^{k+1}\}$ and $\{\mathbf{u}_i^{k+1}\}$ are the eigenvalues and orthonormal eigenvectors of H_{k+1} then, we may write

$$H_{k+1} = \sum_{i=1}^{n} \lambda_i^{k+1} \mathbf{u}_i^{k+1} \{\mathbf{u}_i^{k+1}\}^T.$$

We now redefine H_{k+1} to be

$$H_{k+1} = \sum_{i=1}^{n} |\lambda_i^{k+1}| \, \mathbf{u}_i^{k+1} \{\mathbf{u}_i^{k+1}\}^T$$

which ensures that H_{k+1} is non-negative definite. If however any member of the set $\{\lambda_i^{k+1}\}$ were zero then the safest strategy would be to define

$$H_{k+1} = I.$$

This suggestion of Greenstadt, whilst ensuring positive definiteness, unfortunately involves a considerable increase in the computational requirements of the algorithms. For this reason it is only feasible for problems with small dimension n.

1.6 An example from mathematical economics

In the study of macroeconomics, one of the most useful tools available to the planner is input–output analysis introduced by Leontief. The input–output table or Leontief matrix links the individual industries to the overall working of the economy. To introduce the concepts we follow the book of Dernburg and Dernburg (1969).

Considering the sales and purchases of an industrial sector, we denote by b_{ij} the sales of industry i to industry j, and by b_{ii} the retention of goods produced by industry i. The sales of goods produced by industry i to outside users is denoted by y_i and the gross output by x_i. Thus

$$x_i = y_i + \sum_j b_{ij}. \tag{4}$$

The next step is to define the input coefficient. We assume that the sales of industry i to industry j are in constant proportion (a_{ij}) to the output of industry j, thus

$$b_{ij} = a_{ij} x_j.$$

The quantities a_{ij} are defined to be the input coefficients. From equation (4) we see that in a static situation

$$\mathbf{x} = \mathbf{y} + A\mathbf{x}, \tag{5}$$

where

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)^T,$$

$$\mathbf{y} = (y_1, y_2, \ldots, y_n)^T,$$

and A is the $n \times n$ matrix with the (i, j) element a_{ij}. The matrix $(I-A)$ is known as the Leontief matrix. Equation (5) can be used to determine the required gross outputs $\mathbf{x}$ of the industry sector to meet a preset final demand $\mathbf{y}$.

If supply and demand are not in equilibrium then we must replace equation (5) by a dynamic model. The usual assumption is that the output in each industry changes at a rate which is proportional to the difference between the level of sales and the level of production. Thus our dynamic model takes the form

$$\frac{d\mathbf{x}(t)}{dt} = D[(A-I)\,\mathbf{x}(t)+\mathbf{y}(t)], \tag{6}$$

where D is a diagonal matrix of the reaction coefficients of the industries. Equation (6) thus is a simple model of the dynamic behaviour of the economic system we are considering. The question of the stability of the system being modelled can now be answered, by determining the eigensystem of the matrix $D(A-I)$ and thus considering the behaviour of the solutions to the system (6). In particular, for this model the existence of eigenvalues with positive real part would indicate an instability in the system because the required gross output would grow exponentially with time.

A similar use of the Leontief matrix and eigensystem analysis occurs in a discrete dynamic system of the form

$$\mathbf{x}(t+1)-\mathbf{x}(t) = D[(A-I)\,\mathbf{x}(t)+\mathbf{y}(t)].$$

Such models are of use in studying the stability of interindustry relations, multiple markets and intercountry trade. For fuller details the reader is referred to the text of Dernburg and Dernburg (1969).

1.7 A Sturm–Liouville problem

In the numerical analysis of ordinary and partial differential equations, a commonly occurring problem is the determination of an approximating eigensystem of the continuous problem. This may represent the vibration of bars, plates or structures, the oscillation of fluids, etc. Many of these problems are now tackled by variational means using a technique frequently referred to as a finite element or Rayleigh–Ritz method. Our example in this section is a straightforward Rayleigh–Ritz attack on a Sturm–Liouville problem. Our aim is to demonstrate the technique and the resulting eigenproblem in as simple a manner as possible.

We therefore consider the problem of determining those values of λ for which there exists a non-trivial differentiable function $\phi(x)$ on $[a, b]$ which, under suitable assumptions, satisfies the differential equation

$$(p(x)\phi'(x))'-q(x)\phi(x)+\lambda r(x)\phi(x) = 0, \tag{7}$$

the boundary conditions

$$\phi(a) = \phi(b) = 0, \tag{8}$$

and the normalization condition

$$\int_a^b r(x)\phi^2(x)\,dx = 1. \tag{9}$$

The functions $p(x), q(x), r(x)$ satisfy $p(x) > 0, q(x) \geq 0$ and $r(x) > 0$. The next step in the analysis is to place a mesh on the interval $[a, b]$ consisting of the points

$$a = x_0 < x_1 < x_2 < \ldots < x_{N+1} = b.$$

Let M denote a subspace of functions defined on the mesh on the interval $[a, b]$. For example, M might be chosen to be a *spline* subspace such that each $\psi_j \in M$ is a cubic polynomial on each interval $[x_j, x_{j+1}], j = 0, \ldots, N$ and such that ψ_j has continuous second derivative at the points x_j. Practical considerations regarding the choice of M and the basis functions ψ_j are beyond the scope or intention of this example.

Returning to our Sturm–Liouville problem we cast our eigenproblem in the form of a Rayleigh–Ritz minimization. Thus the solution to (7), (8) and (9) is equivalent to finding the stationary values, and corresponding functions ϕ of the Rayleigh quotient

$$R[\phi] = \left\{\int_a^b [p(\phi')^2 + q\phi^2]\,dx\right\} \Big/ \left\{\int_a^b r\phi^2\,dx\right\}. \tag{10}$$

In general, we cannot deal with (10) unless we make some simplifying assumptions. If we restrict our approximate solution ϕ to lie in the (usually finite) subspace M then we may carry the analysis further. Thus letting

$$\phi = \sum_{j=1}^{J} c_j \psi_j(x)$$

where c_j are constants to be determined, the problem reduces to that of determining the values of c_j corresponding to the stationary values of (10). For ease of writing we use the notation

$$\phi = \mathbf{c}^T \boldsymbol{\psi} \tag{11}$$

where

$$\mathbf{c} = (c_1, c_2, \ldots, c_J),$$

$$\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_J).$$

If we substitute into (10) the assumption (11), then

$$R[\mathbf{c}] = N(\mathbf{c})/D(\mathbf{c}),$$