JOURNAL OF COMPUTATIONAL PHYSICS 24, 331–338 (1977)

On the Numerical Solution of Difficult Eigenvalue Problems

A. DAVEY

School of Mathematics, University of Newcastle upon Tyne, Newcastle upon Tyne, England

Received August 3, 1976; revised January 4, 1977

We present a generalization of the Riccati method for solving difficult linear differential eigenvalue problems, which allows the differential system to be of either even or odd order. As an example of an odd order problem we use the method to obtain the first four eigenvalues of the Blasius problem, making use of a complex contour of integration to avoid the singularities of the Riccati equation. We also give numerical results for an Orr–Sommerfeld problem which illustrate the efficacy of the Riccati method compared to orthonormalization.

1. INTRODUCTION

In an excellent paper Scott [6] has described the important Riccati method for solving linear differential eigenvalue problems for systems of ordinary differential equations. The basic solutions of a linear differential system are usually exponential in character and if the real parts of the characteristic values of the operator are widely separated then, when using an explicit integration scheme, parasitic growth problems occur and a special method, such as orthonormalization [1], will be needed to resolve the problem. The importance of the Riccati method lies in the fact that it transforms the linear problem into a nonlinear problem whose characteristic values all have negative real parts thus ensuring that the integration will be stable. The exponential character is essentially transformed by the nonlinearity to a tanh type behavior.

In Scott's paper he restricted his attention to differential systems of even order with an equal number of boundary conditions at each end of the range of integration. In Section 2 of this paper we formulate the Riccati method in a different way from Scott which will allow us to consider a differential system of either even or odd order and which can have more boundary conditions specified at one end of the range of integration than at the other end. We illustrate the use of the method by considering the Blasius eigenvalue problem as discussed by Wilks and Bramley [8] and we introduce the idea of using a complex contour of integration as an alternative to the "switching" procedure which is frequently necessary when the Riccati method is used. Then, in Section 3, we present a numerical comparison of the solution of the Orr–Sommerfeld equation for plane Poiseuille flow at high Reynolds numbers using the Riccati method and using orthonormalization.

ISSN 0021–9991

2. Generalization of the Riccati Method

Let us consider the homogeneous linear differential eigenvalue problem of order n

$$\mathbf{y}' = \mathbf{F}\mathbf{y}, \tag{1}$$

where $\mathbf{y} = (y_i)$ is a complex n-vector, and $\mathbf{F}$ is an $n \times n$ matrix of coefficients and a $'$ denotes differentiation with respect to the independent variable x. We suppose that x may be chosen so that the range of integration is $0 \leqslant x \leqslant 1$, and so that we know more boundary conditions, m, at $x = 0$ than at $x = 1$, unless n is even when m may be $\frac{1}{2}n$.

Let the boundary conditions at $x = 0$ be

$$\mathbf{G}\mathbf{y} = \mathbf{0}, \tag{2}$$

where $\mathbf{G}$ is an $m \times n$ matrix and the r.h.s. of (2) will be an m-vector. We suppose that the remaining $p = n - m$ boundary conditions can also be written in a form similar to (2) at $x = 1$. We shall write these down later.

Now consider all the solutions of (1) which satisfy (2), these will lie in a vector space of dimension p spanned by the *unknown* conditions at $x = 0$. Thus there will be a set of p of the elements of $\mathbf{y}$ which can be chosen *arbitrarily*, at any value of x, which will determine the other $m = n - p$ elements of $\mathbf{y}$ at the same value of x. Let the set of m such elements be denoted by the m-vector $\mathbf{u}$ and let the set of p elements be denoted by the p-vector $\mathbf{v}$. We choose $\mathbf{u}$ to be $\mathbf{G}\mathbf{y}$.

It follows that we can rewrite (1) and (2) in the form

$$\mathbf{u}' = \mathbf{A}\mathbf{u} + \mathbf{B}\mathbf{v}, \tag{3}$$

$$\mathbf{v}' = \mathbf{C}\mathbf{u} + \mathbf{D}\mathbf{v}, \tag{4}$$

and

$$\mathbf{u} = \mathbf{0}, \qquad \text{when} \quad x = 0. \tag{5}$$

The dimensions of the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, will be $m \times m$, $m \times p$, $p \times m$, $p \times p$, respectively. Moreover, at any value of x, for all the solutions of (1) which satisfy (2), $\mathbf{u}$ will be determined by $\mathbf{v}$ and there will exist a matrix $\mathbf{R}$ of dimension $m \times p$ such that

$$\mathbf{u} = \mathbf{R}\mathbf{v}. \tag{6}$$

The matrix $\mathbf{R}$ corresponds to Scott's matrix $\mathbf{R}_1$ when $m = p$, and we now obtain the Riccati equation for $\mathbf{R}$.

To obtain this equation we use the fact that at this stage $\mathbf{v}$ is arbitrary and we substitute (6) in (3), (4) to obtain, respectively,

$$\mathbf{R}'\mathbf{v} + \mathbf{R}\mathbf{v}' = \mathbf{A}\mathbf{R}\mathbf{v} + \mathbf{B}\mathbf{v}, \tag{7}$$

and

$$\mathbf{v}' = (\mathbf{D} + \mathbf{C}\mathbf{R})\mathbf{v}. \tag{8}$$

We now eliminate $\mathbf{v}'$ between (7), (8) and since the resulting equation must be valid for all $\mathbf{v}$ we obtain the Riccati equation for $\mathbf{R}$, namely

$$\mathbf{R}' = \mathbf{B} + \mathbf{AR} - \mathbf{R}(\mathbf{D} + \mathbf{CR}). \tag{9}$$

We must now determine the boundary conditions for (9). When $x = 0$ the boundary condition is (5), which must be valid for all $\mathbf{v}$, so that

$$\mathbf{R} = \mathbf{0}, \qquad \text{when} \qquad x = 0. \tag{10}$$

When $x = 1$ we will be able to write the boundary condition in the form

$$\mathbf{Hu} = \mathbf{Jv},$$

where $\mathbf{H}$ is a $p \times m$ matrix anh $\mathbf{J}$ is a $p \times p$ matrix so that, using (6), then

$$(\mathbf{HR} - \mathbf{J})\mathbf{v} = \mathbf{0}, \qquad \text{when} \qquad x = 1. \tag{11}$$

Hence the differential equation to be integrated is (9) with initial condition (10) and the eigenvalues will be such as to make

$$\det(\mathbf{HR} - \mathbf{J}) = 0, \qquad \text{when} \qquad x = 1. \tag{12}$$

In general $\mathbf{H}$ will be nonzero and an iterative technique or a Dirac comb can be used to locate the eigenvalues. In the *special* case when $\mathbf{H} = \mathbf{0}$ then $\mathbf{R}$ is singular at $x = 1$ and (12) cannot be used to determine the eigenvalues. This difficulty can be overcome if we also integrate backward over a small range from $x = 1$ toward $x = 0$ and find the new matrix $\mathbf{S}$, corresponding to $\mathbf{R}$, which will be such that $\mathbf{v} = \mathbf{Su}$. The boundary condition (12) may then be replaced by

$$\begin{vmatrix} -\mathbf{I} & \mathbf{R} \\ \mathbf{S} & -\mathbf{I} \end{vmatrix} = 0,$$

at $x = 1 - \epsilon$, $\epsilon > 0$, say.

When the required eigenvalue has been found we may then obtain the eigenfunction. To do this we first solve (11) to find the associated eigenvector $\mathbf{v}$ at $x = 1$. We can then obtain $\mathbf{v}$ over the whole interval $0 \leqslant x \leqslant 1$ by integrating (8) backward from $x = 1$ to $x = 0$. When this integration is done it is important to use previously stored values of $\mathbf{R}$ rather than simultaneously integrating the Riccati equation backward as Sloan [7] has well explained. The point here is that the Riccati method ensures that all the real parts of the characteristic values of $(\mathbf{D} + \mathbf{CR})$ in (9) will be positive. Therefore the Riccati equation will be stable during a forward integration and Eq. (8) for $\mathbf{v}$ will be stable during a backward integration. (The Riccati equation will be unstable during a backward integration and the $\mathbf{v}$ equation will be unstable during a forward integration.) After $\mathbf{v}$ has been found over the whole range $0 \leqslant x \leqslant 1$ then $\mathbf{u}$ is given by (6) and so the complete eigenfunction is obtained. Gersting and

Jankowski [3] failed to obtain eigenfunctions because they tried to integrate the Riccati equation backward at the same time as they integrated their $\mathbf{v}$ equation.

For a real problem with real eigenvalues singularities of (9) may be encountered when the integration is along the real axis because a consideration of the general initial value problem shows that $\mathbf{R}$ is essentially the ratio of two matrices and the denominator will cause $\mathbf{R}$ to be singular if the problem with complementary boundary conditions at $x = 1$ has characteristic lengths on the range of integration. When $m = p$ this difficulty can be circumvented by switching to the inverse of $\mathbf{R}$ as has been explained by Scott [6]. When $m \neq p$, or even when $m = p$, this difficulty can however, also be easily overcome by deforming the contour of integration into the complex plane. For a problem defined on the real axis with complex eigenvalues singularities of $\mathbf{R}$ are not so likely to be encountered because the complementary problem will not be likely to have real characteristic lengths, but even if they are then again the contour can be deformed.

For instance if one uses the Riccati method to solve the adjoint Orr–Sommerfeld equation for plane Couette flow when the wave number $\kappa = 1$ and the Reynolds nomber Re $= 8000$, one cannot integrate along the real axis from $x = 0$ to $x = 1$ without "switching" due to the existence of a singularity near $x = 0.074$. If, however, one integrates along the contour $z = t + i\beta t(1 - t)$, $0 \leqslant t \leqslant 1$, with either $\beta = -0.05$ or $\beta = +0.05$ the eigenvalue is obtained without any difficulty, correct to six significant figures using 500 integration steps. We have also used the Riccati method together with complex contours to solve an astrophysical problem, Jones [4].

As an example of the application of the above generalization of the Riccati method to an *odd* order differential problem we have considered the eigenvalue problem of perturbations from the Blasius velocity profile for the boundary-layer flow of a viscous incompressible fluid past a semi-infinite flat plate as considered by Libby [5] and Wilks and Bramley [8].

The differential equation for the eigenfunction $y(x)$ and eigenvalue σ is

$$y''' + fy'' + \sigma f'y' + (1 - \sigma) f''y = 0, \tag{13}$$

and the boundary conditions are

$$y = y' = 0, \quad \text{when} \quad x = 0, \tag{14}$$

$$y' \to 0 \quad \text{exponentially as} \quad x \to \infty, \tag{15}$$

where f is the usual Blasius function.

We take

$$\mathbf{u} = \begin{pmatrix} y \\ y' \end{pmatrix}, \qquad \mathbf{v} = (y''), \qquad \mathbf{R} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}, \tag{16}$$

so that we set $y = r_1 y''$ and $y' = r_2 y''$ and the equations for r_1, r_2 are

$$r_1' = fr_1 + r_2 + (1 - \sigma) f'' r_1^2 + \sigma f' r_1 r_2, \tag{17}$$

and

$$r_2' = 1 + fr_2 + \sigma f' r_2^2 + (1 - \sigma) f'' r_1 r_2 . \tag{18}$$

Boundary conditions (14) and (15) yield

$$\begin{aligned} r_1 = r_2 = 0, &\qquad \text{when } x = 0, \\ r_2 \to (-1/f), &\qquad \text{as } x \to \infty. \end{aligned} \tag{19}$$

Usually with problems defined on a semi-infinite interval one starts the integration from x large, via a few terms of an asymptotic expansion, and integrates towards $x = 0$. We do not adopt such a course here because for this particular problem the asymptotic details make the Riccati method rather cumbersome when the integration is from x large to zero. A penalty which we have to pay because we integrate from $x = 0$ is to be careful to integrate far enough out to be able to select the required solution, but not too far out because all solutions decay as x increases whatever the value of σ.

A preliminary attempt to obtain the first eigenvalue by integration along the real axis met with failure, as expected, because a singularity was encountered. Since it is known that all the eigenvalues of this problem are real it is likely that all the characteristic lengths of the complementary problem, where the Riccati equation has its singularities, will also be real. The problem is therefore an ideal candidate for using a complex contour and we chose the contour $z = t - 0.02it(\lambda - t)$, $0 \leqslant t \leqslant \lambda$. The first four eigenvalues which we obtained were $\sigma = 2.0000$, 3.7737, 5.6287, 7.5132 using $\lambda = 6, 7, 8, 9$, respectively, and with $\lambda = 8$ we obtained 3.7736 for the second eigenvalue and hence complete agreement with the values found by Wilks and Bramley [8].

3. Comparison of the Riccati Method with Orthonormalization

To illustrate the usefulness of the method described in Section 2 we will use it to calculate the eigenvalue c which corresponds to the most unstable mode in the classical linear stability problem of plane Poiseuille flow, when the wavenumber $\kappa = 1$ and the Reynolds number Re is very large, so that the characteristic values of the differential operator will be of order ± 1 and $\pm \mathrm{Re}^{1/2}$. The differential equation for this problem is the Orr–Sommerfeld equation

$$L\phi \equiv \{D^2 - \kappa^2 - i\kappa \,\mathrm{Re}(1 - x^2 - c)\}\{D^2 - \kappa^2\}\phi - 2i\kappa \,\mathrm{Re}\, \phi = 0, \tag{20}$$

and the appropriate boundary conditions for the most unstable mode are

$$\phi' = \phi''' = 0, \qquad \text{when} \qquad x = 0, \tag{21}$$

and

$$\phi = \phi' = 0, \qquad \text{when} \qquad x = 1. \tag{22}$$

We shall be particularly concerned with solutions for very large values of Re in order to obtain a true comparison between using the Riccati method and using orthonormalization; this problem has been considered previously for smaller values of Re by several authors, see for example [2, 3]. See also Sloan [7] for details of the evaluation of the eigen*function*.

We integrate from $x = 0$ to $x = 1$ using a Runge–Kutta routine and first we use an orthonormalization method and a Newton–Raphson process to iterate to the eigenvalue and we determine, for $\log_{10} \mathrm{Re} = 5(1)10$, the *smallest* number of integration steps of equal length which we can use to obtain the eigenvalue c correct to four significant figures. When Re is very large this number is proportional to $\mathrm{Re}^{1/2}$ because the main restriction is that the Runge–Kutta routine must be convergent.

Second we repeat these calculations using the Riccati method of Section 2 and with

$$\mathbf{u} = \begin{pmatrix} \phi' \\ \phi''' - \kappa^2\phi' \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} \phi \\ \phi'' - \kappa^2\phi \end{pmatrix}, \qquad \mathbf{R} = \begin{pmatrix} r_1 & r_2 \\ r_3 & r_4 \end{pmatrix}. \tag{23}$$

We use this particular formulation because then it transpires that $r_3 \equiv 2i\kappa \,\mathrm{Re}\, r_2$ and so only r_1, r_2, r_4 need be found. (That we can do this is only made possible by the special form of the basic velocity profile for plane Poiseuille flow.) The differential equations for r_1, r_2, r_4 are

$$\begin{aligned} r_1' &= \kappa^2 - r_1^2 - 2i\kappa \,\mathrm{Re}\, r_2^2, \\ r_2' &= 1 - r_1 r_2 - r_4 r_2, \\ r_4' &= \kappa^2 + i\kappa \,\mathrm{Re}(1 - x^2 - c) - 2i\kappa \,\mathrm{Re}\, r_2^2 - r_4^2, \end{aligned} \tag{24}$$

and the boundary conditions are

$$\begin{aligned} r_1 = r_2 = r_4 &= 0, \qquad \text{when } x = 0, \\ r_2 &= 0, \qquad \text{when } x = 1. \end{aligned} \tag{25}$$

The comparison between the two methods is presented in Table I. The column headed ONIZ is approximately the smallest number of integration steps which may by used to calculate c correctly to four significant figures using orthonormalization. The column headed RICCATI is the corresponding number using the Riccati method via (24), (25). The absence of the r_3 equation does not affect the comparison.

It is clear from Table I that when $\mathrm{Re}^{1/2} > 1000$, so that the characteristic values of the differential operator, ± 1, $\pm \mathrm{Re}^{1/2}$ say, are *very* widely separated then the Riccati method requires approximately twice as many integration steps as orthonormalization. This is only to be expected in view of the fact that the basic solutions of the original linear equation are of exponential character whereas the solutions of the Riccati equation are of a tanh character and so for very large characteristic values the integration routine will only be convergent if the step length is halved.[1] Both methods

[1] For if say, $\phi \sim ae^{\lambda x} + be^{x} + ce^{-x} + de^{-\lambda x}$, and the step length is h the Runge–Kutta routine will require $\lambda h < 2$, say, for convergence, and $r_i \sim (a'e^{\lambda x} + b'e^{x} + c'e^{-x} + d'e^{-\lambda x})/(ae^{\lambda x} + be^{x} + ce^{-x} + de^{-\lambda x})$ so that $r_i \sim a'/a + \cdots + d''e^{-2\lambda x}$, as x increases; and now we shall need $2\lambda h < 2$ for convergence, where $\lambda \sim \mathrm{Re}^{1/2}$.

TABLE I

The smallest number of integration steps of equal length which may be used to calculate the eigenvalue c correct to four significant figures by the two methods

$\log_{10}\mathrm{Re}$	c	ONIZ	RICCATI
5[a]	$0.14592 - 0.01504i$	325	300
6	$0.06659 - 0.01398i$	600	800
7	$0.03064 - 0.00726i$	1200	2600
8	$0.01417 - 0.00351i$	3700	8200
9	$0.006566 - 0.001660i$	12000	26000
10	$0.003045 - 0.000777i$	37000	82000

[a] For this entry c was calculated correct to five decimal places instead of four significant figures.

needed roughly the same number of iterations, usually three or four, and when $\mathrm{Re}^{1/2} > 1000$ we found that the computing times required by the two methods differed by less than 7%. This is mainly because the factor of two as regards the number of steps required by the two methods is balanced by the fact that the orthonormalization method has to integrate twice as many differential equations.

When $\mathrm{Re}^{1/2} \leqslant 1000$ so that the characteristic values are not too widely separated then the step length for the Riccati method is not so severely restricted and, as Table I indicates, as Re is lowered the Riccati method requires *comparatively* fewer integration steps and hence requires less computing time than orthonormalization. This advantage of the Riccati method should be enhanced if a sophisticated variable-step integrator is used to accommodate the tanh character of the r_i relative to similar accommodation of the structure of ϕ using orthonormalization.

All the comments above as regards the relative computing times required did *not* take advantage of the relationship $r_3 \equiv 2i\kappa \, \mathrm{Re} \, r_2$, all four equations for r_1, r_2, r_3, r_4 were integrated when timing.

4. Conclusions

For a general difficult problem we expect the Riccati method to require approximately the same number of integration steps as the orthonormalization method, when using a Runge–Kutta type of integration procedure. Also, in general, the Riccati method will only need to integrate about half as many differential equations as the orthonormalization method. In consequence the Riccati method will be faster than orthonormalization by a factor of order two. Even if the problem is *very* difficult so that the characteristic values are *very* widely separated the Riccati method will still be as fast as orthonormalization.

Another important advantage of the Riccati method is that it is so simple to formulate and program compared with the intricacies of orthonormalization. This is particularly highlighted by the fact that it is much easier to evaluate the eigen*function* using the Riccati method, and this advantage, together with the others mentioned above would seem to indicate that the Riccati method may often be preferable to orthonormalization.

If the posed problem is not self-adjoint it is often useful to solve the adjoint problem during initial tests since it may require less computing time. Moreover, a comparison of the solutions of the posed problem and its adjoint is invariably a good guide to the accuracy of these solutions.

It is interesting that whereas the *analytical* solution of a nonlinear differential problem may be facilitated by transforming it into a linear system, the *numerical* solution of a linear differential system may be facilitated by transforming it into a nonlinear problem!

References

1. S. D. Conte, *SIAM Rev.* **8** (1966), 309.
2. M. L. Curl and W. P. Graebel, *SIAM J. Appl. Math.* **23** (1972), 380.
3. J. M. Gersting and D. F. Jankowski, *J. Appl. Mech.* **39** (1972), 280.
4. C. A. Jones, *Mon. Notic. Roy. Astron. Soc.* **176** (1976), 145.
5. P. A. Libby, *AIAA J.* **3** (1965), 2164.
6. M. R. Scott, *J. Computational Phys.* **12** (1973), 334.
7. D. M. Sloan, *J. Computational Phys.* **24** (1977), 303–319.
8. G. Wilks and J. S. Bramley, *J. Computational. Phys.* **24** (1977), 320–330.

Printed in Belgium