

Note

Numerical Solution of Orr-Sommerfeld-Type Equations

1. INTRODUCTION

Equations of Orr-Sommerfeld type, homogeneous boundary value problems in which the differential equations have solutions with widely differing growth rates, are notoriously hard to solve numerically (for a general introduction to the earlier literature see the review by Gersting and Jankowski [4]). The methods used fall into two categories: first, those which seek to determine the entire solution at once, either by using matrix methods to solve a finite difference scheme or by determining the coefficients in some expansion of the solution; second, the shooting methods which attempt to solve the boundary value problem by initial value methods. In this note I only consider methods in the second category.

2. THE PROBLEM

Consider a system of linear ordinary differential equations depending analytically on some complex parameter p ,

$$\begin{aligned} \frac{d}{dx} y &= F(x, p) y, \\ y: [x_a, x_b] &\rightarrow V, \\ F: [x_a, x_b] \times \mathbb{C} &\rightarrow \mathcal{L}(V, V), \end{aligned} \tag{1}$$

where V is a complex vector space of dimension n and $[x_a, x_b]$ is a closed interval of the real line. Suppose it subject to k independent linear homogeneous boundary conditions at x_a and to $k' = n - k$ at x_b , that is, there exist linearly independent subsets

$$\{\alpha_i\}_{i=1 \dots k} \quad \{\beta_i\}_{i=1 \dots k'} \tag{2}$$

of the dual space V' and the boundary conditions are

$$\alpha_i(y(x_a)) = \beta_j(y(x_b)) = 0 \quad i = 1, \dots, k, j = 1, \dots, k'. \tag{3}$$

It is convenient to introduce bases for the annihilators of the subspaces spanned by these sets, i.e., linearly independent sets of vectors in V

$$\{a_i\}_{i=1 \dots k'} \quad \{b_i\}_{i=1 \dots k} \quad (4)$$

such that

$$\alpha_j(a_i) = \beta_i(b_j) = 0 \quad i = 1, \dots, k', j = 1, \dots, k. \quad (5)$$

Then any solution of the system satisfying both sets of boundary conditions will have for its value at x_a a linear combination of the a_i and at x_b a linear combination of the b_i . The problem is to find the values of p for which such solutions exist (the eigenvalues of the system).

3. SIMPLE SHOOTING

Integration of (1) defines a flow on V , i.e., a map

$$R(x_1, x_2): V \rightarrow V \quad (6)$$

such that

$$R(x_1, x_2) y(x_1) = y(x_2) \quad (7)$$

because the system is linear this map R is linear: if the value of the solution at x_a is a linear combination of the a_i , then at x_b it will be the same linear combination of the vectors

$$\{R(x_a, x_b)a_i\}_{i=1 \dots k'} \quad (8)$$

and this will be expressible as a linear combination of the b_i iff the exterior product (or determinant)

$$\Delta = R(x_a, x_b)a_1 \vee \dots \vee R(x_a, x_b)a_{k'} \vee b_1 \vee \dots \vee b_k \quad (9)$$

is zero. The simple shooting method consists of choosing k' sets of initial values satisfying the boundary conditions at x_a (the a_i), advancing them to x_b by some initial value method, and evaluating Δ . If the differential system is nonsingular in the range $[x_a, x_b]$, Δ is a holomorphic function of the parameter p and its zeros may be found by any of the standard methods.

This simple scheme breaks down when applied to problems of the Orr–Sommerfeld type. Although the initial vectors a_i are linearly independent, the final vectors may well be nearly dependent because the growth of some components of the solution is

so extreme as to swamp the less rapidly growing. A partial cure is to match at some point x_m in the middle of the range and set

$$\Delta = R(x_a, x_m)a_1 \wedge \cdots \wedge R(x_a, x_m)a_{k'} \wedge R(x_b, x_m)b_1 \wedge \cdots \wedge R(x_b, x_m)b_k. \quad (10)$$

4. THE METHOD OF ORTHONORMALIZATION

A much better solution is the method of orthonormalisation (see, e.g., Conte [1]). In this the range of integration is divided into a number of subranges. As in the simple shooting method, a set of k' vectors is advanced from x_a to x_b by integration across each subrange. However, at the end of each subrange the k' vectors are orthonormalized (say by the Gram-Schmidt process) with respect to some arbitrary inner product on V . For a suitable division into subrange this ensures that the vectors remain independent and that the growth of one solution does not swamp the others (or cause overflow problems!). As before, a determinant Δ can be formed, either at x_b or at some internal matching point x_m , the zeros of which identify nontrivial solutions of the boundary value problem. It may not, however, be a holomorphic function of the parameter; in particular, if the quadratic form used to define the inner product is a positive definite Hermitian one, it will not be holomorphic because of the occurrence of complex conjugate quantities in the expression for the inner product. For this reason the use of an orthogonal form (even though it is indefinite) has been recommended by some authors (starting with Gary and Helgason [5]). It is certainly true that the determinant found in this way will be locally holomorphic, but because of the different branches of the square root which may be taken when normalizing, the region of holomorphy will tend to be very small. Thus, it is impossible to use global zero finding methods based on the principle of the argument.

By contrast, if the inner product is based on a Hermitian form, not only do we have the security of a positive definite inner product, but if the Gram-Schmidt process has been used the determinant formed is the product of a positive real function and that which would be obtained if the simple shooting method were used with exact arithmetic (the orthogonalisation of the vectors does not change Δ and the normalisation of each divides Δ by the modulus of the vector). Because of this we can use the principle of the argument to locate the zeros of Δ even though it is not a holomorphic function of p .

5. INVARIANT IMBEDDING

The method of orthonormalisation indicates that the particular set of vectors being advanced is less important than the linear subspace they span; this suggests formulating the problem and its solution directly in terms of subspaces of V .

The differential system induces a flow not only on V but also on most of the associated geometric objects; in particular, it induces a flow on the Grassmanians

$\mathcal{E}_k(V)$ (defined as the set of k -dimensional linear subspaces of V ; for example, $\mathcal{E}_2(\mathbb{R}^3)$ is the set of all planes through the origin of \mathbb{R}^3). The boundary conditions can be stated as

$$\begin{aligned} y(x_a) &\in \mathcal{A} = \text{Span}(a_i) \in \mathcal{E}_k(V), \\ y(x_b) &\in \mathcal{B} = \text{Span}(b_i) \in \mathcal{E}_k(V), \end{aligned} \quad (11)$$

so that if the induced flow on $\mathcal{E}_k(V)$ is denoted R_k , the condition for a nontrivial solution is

$$R_k(x_a, x_b)(\mathcal{A}) \cap \mathcal{B} \neq 0 \quad (12)$$

or if matching at some intermediate point,

$$R_k(x_a, x_m)(\mathcal{A}) \cap R_k(x_b, x_m)(\mathcal{B}) \neq 0. \quad (13)$$

$\mathcal{E}_k(V)$ is a manifold and has certain natural parametrizations. Let $U \in \mathcal{E}_k(V)$ and let \bar{U} be a complementary subspace,

$$\begin{aligned} V &= U \oplus \bar{U}, \\ \dim(U) &= k, \quad \dim(\bar{U}) = k'. \end{aligned} \quad (14)$$

Let $\alpha \in \mathcal{L}(U, \bar{U})$ and consider the map,

$$\alpha \mapsto \text{graph}(\alpha) = \{u \oplus \alpha u : u \in U\} \subset V. \quad (15)$$

This is a parametrization of an open neighbourhood of U in $\mathcal{E}_k(V)$ by elements of $\mathcal{L}(U, \bar{U})$ which can be identified (though noncanonically) with $\mathbb{C}^{kk'}$; the set of such parametrizations (or rather their inverse charts) constitutes an atlas for the manifold $\mathcal{E}_k(V)$.

Having parametrized $\mathcal{E}_k(V)$ (or at least an open subset of it) the flow R_k defines and is defined by a system of differential equation in the parametrization coordinates. The splitting of V induces a splitting of F into four linear maps A, B, C, D .

$$\begin{aligned} F(u \oplus \bar{u}) &= (Au + B\bar{u}) \oplus (cu + D\bar{u}) \\ \forall u \in U, \quad \forall \bar{u} \in \bar{U}. \end{aligned} \quad (16)$$

If

$$y = u \oplus \alpha u, \quad (17)$$

then

$$\begin{aligned} u' \oplus (\alpha' u + \alpha u') &= (Au + Bau) \oplus (Cu + Dau), \\ \Rightarrow u' &= Au + Bau, \\ \alpha' u &= Cu + Dau - \alpha Au - \alpha Bau. \end{aligned} \quad (18)$$

If the second equation is to hold for all initial values of u ,

$$a' = C + Da - aA - aBa, \quad (19)$$

which is the required equation determining the flow R_k . When V is of even dimension and $k = n/2$, this is the central equation of those methods generally known as Riccati or invariant imbedding methods (see, e.g., Scott [6]). An advantage of this derivation is that the extension to cases of arbitrary dimensionality (as recently achieved by several authors [3, 7]) is obvious, but equally as important is the conceptual gain from having a clear geometric formulation of the method. For instance, it is well known that singularities may be encountered in the integration of the Riccati equation and that the solution is to change to another set of variables which satisfy a related Riccati equation. From the geometric point of view it is clear that this behaviour is an essential consequence of the fact that the manifold $\mathcal{S}_k(V)$ is not homeomorphic to an open subset of $\mathbb{C}^{kk'}$ and so cannot be covered by any one parametrization; whenever in the integration we move out of the region covered by our chart we must switch to another. This raises the interesting question of how many charts are needed to form a complete atlas. If it is confined to natural charts (as defined above and used in the standard methods) it is easy to see that at least $k + 1$ are required (in contradiction to the impression often given that two will suffice). For given any k natural charts on $\mathcal{S}_k(V)$ defined by the decompositions of V ,

$$V = U_1 \oplus \bar{U}_1 = \dots = U_k \oplus \bar{U}_k, \quad (20)$$

choose one vector from each of the complementary subspaces $\bar{U}_1 \dots \bar{U}_k$ and if necessary adjoin to this set such additional arbitrary vectors that the subspace of V it generates is of dimension k . Then this subspace is an element of $\mathcal{S}_k(V)$ that does not belong to any of the coordinate domains of the given charts.

6. OTHER METHODS

This geometric point of view can be used to describe other integration techniques and suggests several new ones. By integrating in the dual space V' we get the method of adjoints, the simple shooting method and the method of orthonormalisation can be regarded as an integration in $\oplus^k V$, and the method proposed by Davey [2] in a recent paper involves integrating in $\wedge^k V$. A novel method which is not very efficient, but easy to program, consists of integrating in the space of sets of k orthogonal vectors

This is obtained from the simple shooting method by a slight modification of the derivative subroutine; instead of integrating

$$y'_\alpha = F y_\alpha, \quad \alpha = 1, \dots, k', \quad (21)$$

we take an initial set of orthonormal vectors and integrate

$$y'_\alpha = \prod_{\beta=1}^{k'} (1 - y_\beta y_\beta^\dagger) F y \alpha_\alpha, \quad \alpha = 1, \dots, k'; \quad (22)$$

in other words we only keep that component of each derivative which is perpendicular to the subspace spanned by the vectors. This means that the vectors remain orthonormal and only rotate as much as is necessary for them to stay in the correct subspace.

7. COMPARISON OF METHODS

The invariant imbedding method and that of orthonormalisation require the integration of differential systems of order kk' and $\max[nk, nk']$, respectively; however, the invariant imbedding method needs a steplength about half that used in the other (unless a stiff equation integrator is used) so that the computational overheads are comparable. The chief difference between them is that while both determine the subspace $R_{k'}(x_a, x_b)(a)$, the method of orthonormalisation also determines a specific basis for it. This additional information allows one to use the principle of the argument (the argument of Δ is essentially the orientation of the basis) in a robust and efficient scheme for determining the number and approximate location of the eigenvalues. I have written a FORTRAN package which, given a rectangular region in the complex p plane, scans the boundary recording where (normalising the argument to $(+\pi, -\pi]$) $\arg(\Delta(p))$ jumps between $+\pi$ and $-\pi$ and the direction of the jump. From this information the number of eigenvalues within the rectangle is calculated and if it is nonzero the rectangle is bisected, the new is line scanned, and the process is repeated as many times as desired. Once approximate locations have been found for the eigenvalues they can be refined by faster iterative processes; although it would be possible to devise a process for finding zeros of products of a holomorphic function and a positive real function or to restore the analyticity of Δ by retaining the scale factors used at each orthonormalisation, I have found the ordinary secant method quite satisfactory even though Δ is not strictly analytic.

Location of the eigenvalues by the invariant imbedding method is equivalent to finding the zeros of a meromorphic function with equal densities of poles and zeros. While this may facilitate the precise determination of individual eigenvalues it makes any global search very difficult. However, the subspace must vary less rapidly than the vectors so that for a given amount of computing the method may be more accurate (if there is a large difference between the rates of variation the Riccati equations will be stiff and substantial computational savings might be made by using a stiff equation integrator). The method I have proposed, advancing an orthonormal frame, combines some of the features of both methods; while allowing the use of the principle of the argument it should be as stiff as the Riccati method.

ACKNOWLEDGMENT

This work was carried out on a research studentship from Trinity College Cambridge and on an Isaac Newton studentship.

REFERENCES

1. S. D. CONTE, *SIAM Rev.* **8**(1966), 309.
2. A. DAVEY, *J. Comput. Phys.* **30** (1979), 137.
3. A. DAVEY, *J. Comput. Phys.* **24** (1977), 331.
4. J. M. GERSTING AND D. F. JANKOWSKI, *Internat. J. Num. Meth. Eng.* **4** (1972), 195.
5. J. GARY AND R. HELGASON, *J. Comput. Phys.* **5** (1970), 169.
6. M. R. SCOTT, *J. Comput. Phys.* **12** (1973), 334.
7. D. M. SLOAN, *J. Comput. Phys.* **24** (1977), 320.

RECEIVED: February 9, 1979; REVISED: September 13, 1979

L. O'C. DRURY*
Institute of Astronomy
Madingley Road
Cambridge, England

* Present address: Max Planck Institute für Kernphysik, Postfach 103980, D-6900 Heidelberg, West-Germany.