QUARTERLY OF APPLIED MATHEMATICS

Vol. XVIII April, 1960 No. 1

NUMERICAL SOLUTION OF PARTIAL DIFFERENTIAL EQUATIONS BY BOUNDARY CONTRACTION*

BY

HAROLD W. MILNES AND RENFREY B. POTTS**

Research Laboratories, General Motors Corp., Detroit, Michigan

1. Introduction. The numerical solution of partial differential equations by finite differences can be accomplished either by explicit or by implicit difference schemes, depending in great measure upon the class of problems to which a given equation belongs and also upon the subsidiary conditions which are imposed upon the solution. For instance, if a hyperbolic differential equation with Cauchy boundary conditions is under consideration, it is generally possible to find an explicit difference scheme according to which the solution at any given instant is determined explicitly in terms of known values at a few preceding time intervals. On the other hand, it is inherent in the nature of an equation of elliptic type, with Dirichlet data, that the solution at any interior point of the region throughout which the equation is to be satisfied depends upon the data at every point of the boundary of the region. Consequently, in this case, any explicit technique must involve all the values prescribed at the boundary. In general, even if such an explicit difference relation can be found, it is not in a form which lends itself to practical computation. It is usually necessary, therefore, to have recourse in this case to some implicit difference scheme, that is to say, one in which an undetermined value is defined always in terms of relations among other values, equally undetermined. The solution of such implicit schemes in most instances is extremely difficult, as it reduces ultimately to the solution of simultaneous equations involving a large number of unknowns interrelated in a manner quite unsuitable for easy determination. If the system is linear, there are several classical methods by which it may be treated; most notably, by direct inversion of matrices of large size [1], by iterative procedures [2] and by relaxation [3]. It is the purpose of this paper to introduce a finite difference technique especially adapted to the solution of boundary value problems. This method depends upon the selection of an implicit difference scheme involving relations among only a limited number of the unknowns at a time, in a form that can be conveniently solved. It is especially well suited for use with digital computing machinery since the demand for memory capacity is small and the time required for computation is short compared with other existing methods, even when meshes of large size are involved.

In this preliminary report, the conditions for stability of the process are investigated. Criteria are derived for stars appropriate to various classes of partial differential equations. Applications of the theory to problems of both hyperbolic and elliptic type are discussed and the advantages of the method are illustrated.

*Received September 29, 1958; revised manuscript received April 10, 1959.

**Now at the University of Adelaide, South Australia.

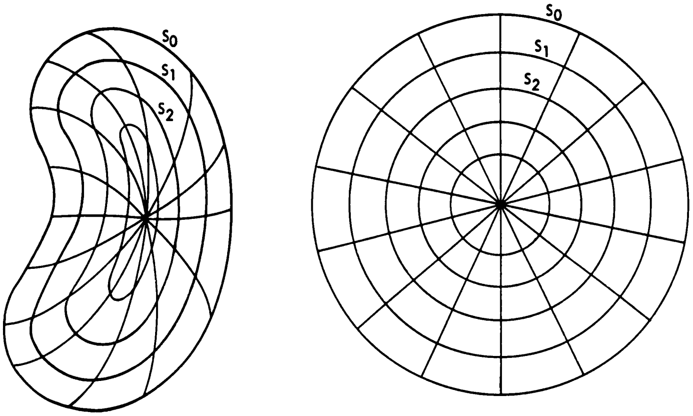

Fig. 1. Typical meshes for boundary contraction

2. Boundary contraction. Consider a region $\mathcal{R}$ (Fig. 1) bounded by a piecewise smooth Jordan curve S_0 and let $\mathcal{R}$ be referred to an (r, θ) coordinate system topologically equivalent to the polar system with the contour S_0 taken to be the unit circle. A sequence of concentric contours $S_1, S_2, \cdots$ of decreasing radius, which will be referred to as circles in the following, may be defined interior to S_0 and a mesh fitted onto $\mathcal{R}$ by considering, as nodal points of the mesh, the points of intersection of the circles of this sequence with a finite number $(N + 1)$ of radii from the origin. Suppose that the prescribed data are sufficient to determine initially the values of the solution at the nodal points on, say, $(j + 1)$ of the outer circles $S_0, S_1, \cdots S_j$. For instance, if Dirichlet data are given, S_0 is prescribed initially; if Cauchy type data are prescribed involving both the values of the solution as well as its normal derivative at the outer edge, then S_0 and S_1 are given initially. The contraction method for solving boundary value problems is a step-by-step process that is initiated at the outer boundary, where the originally prescribed data are given. At each stage, the solution is sought only at the nodal points corresponding to the outermost circle where the solution remains unknown. The newly computed data on this circle then replace those used in determining it and a new boundary value problem, similar to the original one, arises on a somewhat smaller region interior to the previous one. For example, if the data are known on $S_0, S_1 \cdots S_j$, this information is used to compute only the unknown data on S_{j+1}; then, the known information on $S_1, S_2, \cdots, S_{j+1}$ is used to determine S_{j+2}; and so forth. The new problem is treated in a similar fashion and the solution moves inwards progressively with the boundary being contracted onto the center. It should be observed that the data, as specified at the outset, are soon forgotten but that the information is propagated to the center in modified form as the new conditions on the inner boundaries. Thus, when using this method, the tremendous advantage is obtained that at any given stage it is necessary to work with only a small proportion of the total number of nodal points of the mesh.

3. Stability and propagation of error. The questions of stability and propagation of error are critical ones for this numerical method. This may be understood most easily in terms of the notations that will be used subsequently. Let the angular arguments of the $(N + 1)$ radii, to which reference was previously made, be denoted as θ_n, $(n = 0, 1, 2, \cdots, N)$ and $u(k, n) \equiv u(r_k, \theta_n)$ be the value of the solution at the nodal point defined as the intersection of the radial line determined by θ_n with the circle S_k, $(k = 0,$

1, 2, $\cdots$) which has radius r_k . Also, let $\mathbf{v}_k$ be the column vector with the components: $u(k, 0)$, $u(k, 1)$, $\cdots$, $u(k, N)$ which are the values of the solution on the circle S_k . The contraction process consists in the repeated application of a transformation $\mathbf{A}_k$ that relates the $(j + 1)$ vectors $\mathbf{v}_{k+j+1}$, $\mathbf{v}_{k+j}$, $\cdots$, $\mathbf{v}_{k+1}$ to the $(j + 1)$ vectors $\mathbf{v}_{k+j}$, $\mathbf{v}_{k+j-1}$, $\cdots$, $\mathbf{v}_k$. In the particular case of linear differential equations to be considered here, the transformation is expressible in matrix form. Using block representation, this can be written in the form:

$$\begin{bmatrix} \mathbf{v}_{k+1} \\ \mathbf{v}_{k+2} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_{k+j+1} \end{bmatrix} = \mathbf{A}_k \begin{bmatrix} \mathbf{v}_{k} \\ \mathbf{v}_{k+1} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_{k+j} \end{bmatrix}. \qquad (1)$$

Thus:

$$\{\mathbf{v}_{k+1}\mathbf{v}_{k+2} \cdots \mathbf{v}_{k+j+1}\} = \prod_{i=0}^{k} \mathbf{A}_i \{\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_j\}. \qquad (2)$$

For typographical reasons, here and subsequently, it is convenient to write column vectors row-wise, using braces, as:

$$\begin{bmatrix} u(0, 1) \\ u(0, 2) \\ \cdot \\ \cdot \\ \cdot \\ u(0, N) \end{bmatrix} \equiv \{u(0, 1)u(0, 2) \cdots u(0, N)\}; \qquad \begin{bmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_j \end{bmatrix} \equiv \{\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_j\}.$$

To illustrate the notation with an example, suppose that an explicit star relation is given by: $u(2 + k, n) = 1/2\,[u(1 + k, n - 1) + u(1 + k, n + 1)] + 1/4\,[u(k, n - 2) + 2u(k, n) + u(k, n + 2)]$ where the indices corresponding to the angular argument are assumed to be reduced modulo $N + 1$ and $k = 0, 1, 2, \cdots$. Then: $\mathbf{v}_{k+2} = \mathbf{P}\mathbf{v}_{k+1} + \mathbf{Q}\mathbf{v}_k$ where $\mathbf{P}$ and $\mathbf{Q}$ are the $(N + 1) \times (N + 1)$ circulant matrices

$$\mathbf{P} = \tfrac{1}{2}\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 1 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \qquad \mathbf{Q} = \tfrac{1}{4}\begin{bmatrix} 2 & 1 & 0 & \cdots & 0 & 1 \\ 1 & 2 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 1 & 0 & 0 & \cdots & 1 & 2 \end{bmatrix}.$$

Using block representation:

$$\begin{bmatrix} \mathbf{v}_{k+1} \\ \mathbf{v}_{k+2} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{P} & \mathbf{Q} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{k} \\ \mathbf{v}_{k+1} \end{bmatrix},$$

where $\mathbf{0}$ and $\mathbf{I}$ are respectively the $(N + 1) \times (N + 1)$ zero and identity matrices. In this simple case, the matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{P} & \mathbf{Q} \end{bmatrix}$$

is independent of k, so that

$$\begin{bmatrix} \mathbf{v}_{k+1} \\ \\ \mathbf{v}_{k+2} \end{bmatrix} = \mathbf{A}^{k+1} \begin{bmatrix} \mathbf{v}_0 \\ \\ \mathbf{v}_1 \end{bmatrix}.$$

As k increases and the solution moves inwards, an error made at some point in the calculation may be propagated forwards and may give rise to a first type of instability. This may be examined by assuming that for some integer m, an error $\boldsymbol{\varepsilon}_{m+j}$ is introduced in $\mathbf{v}_{m+j}$, and then determining the resulting error at a later stage. From (1) this is given by $\boldsymbol{\varepsilon}_{m+k+j+1}$, where:

$$\{\boldsymbol{\varepsilon}_{m+k+1} \cdots \boldsymbol{\varepsilon}_{m+k+j} \boldsymbol{\varepsilon}_{m+k+j+1}\} = \prod_{i=m}^{m+k} \mathbf{A}_i \{\mathbf{0} \cdots \mathbf{0}\ \boldsymbol{\varepsilon}_{m+j}\}. \tag{3}$$

The magnitude of the resulting error can be measured by comparing $|\boldsymbol{\varepsilon}_{m+k+j+1}|$ with $|\boldsymbol{\varepsilon}_{m+j}|$ where $|\boldsymbol{\varepsilon}|$ is the maximum modulus of the components of $\boldsymbol{\varepsilon}$. If $\mathbf{v}(r)$ is the vector consisting of the values of u on the mesh circle of radius r, and if an error occurs in $\mathbf{v}(r')$ for some $r' > r$, then the resulting error in $\mathbf{v}(r)$ will depend upon the choice of mesh used in going from r' to r, since, in general, the matrix $\mathbf{A}_i$ is determined by the mesh. For each mesh, let $m + j$ be such that $\mathbf{v}_{m+j} = \mathbf{v}(r')$ and k such that $\mathbf{v}_{m+k+j+1} = \mathbf{v}(r)$. Assuming that the solution is uniformly bounded over $\mathcal{R}$, for stability of the contraction process, it is required that $\boldsymbol{\varepsilon}_{m+k+j+1}$ be uniformly bounded as the mesh decreases, independently of r and r'. This in turn, requires that the matrix product $(\prod_{i=0}^{k} \mathbf{A}_i)$ should remain bounded for all k as the mesh decreases. Thus it is possible, in further consideration of the first type of instability, to limit discussions to this matrix product.

It should be noted that Eq. (1) is an explicit vector relationship to determine $\mathbf{v}_{k+j+1}$; however, the components of this vector need not be explicitly determined and they are, indeed, usually related implicitly one with another. The solution of these implicit relations gives rise to a second type of instability which is described best, perhaps, by the aid of a simple example. In particular, for the case $k = 0, j = 0$, consider the approximating star relating the components of $\mathbf{v}_1$ to those of $\mathbf{v}_0$:

$$\begin{aligned} c_2^{(1)}u(1, n+2) + c_1^{(1)}u(1, n+1) + c_0^{(1)}u(1, n) + c_2^{(0)}u(0, n+2) \\ + c_1^{(0)}u(0, n+1) + c_0^{(0)}u(0, n) = 0 \qquad (n = 0, 1, 2, \cdots, N), \end{aligned} \tag{4}$$

where the indices corresponding to the angular argument are assumed to be reduced modulo $(N + 1)$. If $c_2^{(1)} \neq 0$, this may be solved as:

$$u(1, n+2) = -\left(\frac{1}{c_2^{(1)}}\right) \sum_{s=0}^{1} c_s^{(1)}u(1, n+s) - \left(\frac{1}{c_2^{(1)}}\right) \sum_{s=0}^{2} c_s^{(0)}u(0, n+s), \tag{5}$$

which implies that, for each $n = 0, 1, 2, \cdots, N$, $u(1, n + 2)$ is computed from $u(1, n + 1)$, $u(1, n)$ and the nodal values on S_0 . The system (5) can be expressed as a matrix equation:

$$\{u(1, n+1)\ u(1, n+2)\} = \mathbf{G}\{u(1, n)\ u(1, n+1)\} - \left(\frac{1}{c_2^{(1)}}\right)\mathbf{H}_0\mathbf{R}^n\mathbf{v}_0 , \tag{6}$$

where:

$$\mathbf{G} = \begin{bmatrix} 0 & 1 \\ -c_0^{(1)}/c_2^{(1)} & -c_1^{(1)}/c_2^{(1)} \end{bmatrix}, \tag{7}$$

$\mathbf{H}_0$ is the $2 \times (N + 1)$ matrix:

$$\mathbf{H}_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ c_0^{(0)} & c_1^{(0)} & c_2^{(0)} & 0 & \cdots & 0 \end{bmatrix} \tag{8}$$

and $\mathbf{R}$ is the $(N + 1) \times (N + 1)$ matrix:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \cdot & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdot & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & 1 & 0 \\ 1 & 0 & 0 & 0 & \cdot & 0 & 0 \end{bmatrix}. \tag{9}$$

The system is seen to be in a very convenient form for numerical solution since $\mathbf{v}_0$ is already known, so that once $u(1, 0)$ and $u(1, 1)$ are determined, then $u(1, 2)$ can be computed explicitly from (5) with $n = 0$; similarly, with $u(1, 1)$ and $u(1, 2)$ not known, $u(1, 3)$ follows with $n = 1$; and so on. To determine $u(1, 0)$ and $u(1, 1)$ it is merely necessary to start the process with a good initial guess for these quantities and perform the indicated calculations by making a complete circuit around S_1, coming back to $u(1, 0)$ and $u(1, 1)$. In virtue of the continuity of u in $\mathcal{R}$, the results obtained at the end of the circuit will agree with those at the beginning, whenever (4) is a linearly independent system of equations, if and only if they are the correct solutions. Thus, the implicit scheme (5) really reduces to two simultaneous equations in two unknowns. Analytically, the system (6) may be solved in the same way by carrying out the indicated substitutions and equating the end result to the initial value. Then:

$$\begin{aligned} \{u(1, N + 1)\ u(1, N + 2)\} &\equiv \{u(1, 0)\ u(1, 1)\} \\ &= \mathbf{G}^{N+1}\{u(1, 0)\ u(1, 1)\} - \left(\frac{1}{c_2^{(1)}}\right) \sum_{t=0}^{N} \mathbf{G}^{N-t}\mathbf{H}_0\mathbf{R}^t\mathbf{v}_0 . \end{aligned} \tag{10}$$

It is clear from (10) that the numerical calculations will become unstable as they proceed around S_1, if the iterated matrix $\mathbf{G}$ is unstable. Thus, a second stability condition, in addition to the one previously indicated, must be satisfied if a boundary value problem is to be solved satisfactorily by the contraction method.

4. Instability of the first type. In this section criteria will be developed for stability of the first kind discussed in Sec. 3 above. At the outset, a restriction is imposed upon the type of star to be considered and it will be assumed that this is limited to an approximating function of the general form:

$$\begin{aligned} &\sum_{s=0}^{N} c_s^{(j+1)}u(k + j + 1, n + s) + \sum_{s=0}^{N} c_s^{(j)}u(k + j, n + s) + \cdots \\ &\qquad + \sum_{s=0}^{N} c_s^{(0)}u(k, n + s) = 0, \qquad (n = 0, 1, 2, \cdots, N; k = 0, 1, 2, \cdots). \end{aligned} \tag{11}$$

It should be noted that the weights $c_s^{(j)}$ in this system of equations are independent of n as well as k, which implies that the basic form of this star is unchanged as contraction moves inwards and also as the computation moves around S_{k+j+1}. A star of this type is appropriate, for instance, to homogeneous linear partial differential equations which have constant coefficients when referred to the chosen mesh system. Certain generalizations upon this will be indicated later in Sec. 8.

The stability criterion for the system (11) corresponds in a simple way to the system itself. To illustrate this, the final results obtained in this section will be stated here. By taking $k = n = 0$, a set of $(N + 1)$ algebraic equations can be derived from (11) by replacing the $u(0, s)$, $u(1, s)$, $\cdots$, $u(j + 1, s)$ with ω^s (where ω is any of the $(N + 1)$ roots of $\omega^{N+1} = 1$) and multiplying each sum starting from the left by λ^{j+1}, λ^j, $\cdots$, λ, 1 to obtain:

$$\lambda^{j+1} \sum_{s=0}^{N} c_s^{(j+1)} \omega^s + \lambda^j \sum_{s=0}^{N} c_s^{(j)} \omega^s + \cdots + \lambda \sum_{s=0}^{N} c_s^{(1)} \omega^s + \sum_{s=0}^{N} c_s^{(0)} \omega^s = 0. \tag{12}$$

The essential condition for the stability of (11) as contraction moves inwards relates to the roots of (12). *If the circulant matrix* $\mathbf{C}^{(j+1)}$ *defined by* (14) *is non-singular, necessary and sufficient conditions for the stability of the system* (11) *are that for each value of* ω, *the* $(j + 1)$ *roots of* (12) *should be distinct and have moduli less than or equal to unity.*

It may be remarked that in practice this stability criterion reduces to an exceedingly simple condition since a great number of the weight factors are usually zero. For instance, the stability equation associated with (4) is:

$$\lambda(c_2^{(1)}\omega^2 + c_1^{(1)}\omega + c_0^{(1)}) + (c_2^{(0)}\omega^2 + c_1^{(0)}\omega + c_0^{(0)}) = 0.$$

For the derivation of this result, Eqs. (11) can be combined into a matrix equation:

$$\mathbf{C}^{(j+1)}\mathbf{v}_{k+j+1} + \mathbf{C}^{(j)}\mathbf{v}_{k+j} + \cdots + \mathbf{C}^{(1)}\mathbf{v}_{k+1} + \mathbf{C}^{(0)}\mathbf{v}_k = \mathbf{0}, \qquad (k = 0, 1, 2, \cdots), \tag{13}$$

where $\mathbf{C}^{(i)}$ is the circulant matrix:

$$\mathbf{C}^{(i)} \equiv \begin{bmatrix} c_0^{(i)} & c_1^{(i)} & c_2^{(i)} & \cdot & c_N^{(i)} \\ c_N^{(i)} & c_0^{(i)} & c_1^{(i)} & \cdot & c_{N-1}^{(i)} \\ c_{N-1}^{(i)} & c_N^{(i)} & c_0^{(i)} & \cdot & c_{N-2}^{(i)} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_1^{(i)} & c_2^{(i)} & c_3^{(i)} & \cdot & c_0^{(i)} \end{bmatrix} \equiv \sum_{s=0}^{N} c_s^{(i)} \mathbf{R}^s. \tag{14}$$

If Eqs. (11) are sufficient to determine $u(k + j + 1, 0)$, $u(k + j + 1, 1)$, $\cdots$, $u(k + j + 1, N)$, the system must represent a linearly independent set of equations in these unknowns; this implies that $\mathbf{C}^{(j+1)}$ must possess an inverse $(\mathbf{C}^{(j+1)})^{-1} = \mathbf{D}$ and:

$$\mathbf{v}_{k+j+1} = -\mathbf{D}(\mathbf{C}^{(j)}\mathbf{v}_{k+j} + \mathbf{C}^{(j-1)}\mathbf{v}_{k+j-1} + \cdots + \mathbf{C}^{(0)}\mathbf{v}_k). \tag{15}$$

It is convenient to express (15) in a form similar to (1):

$$\{\mathbf{v}_{k+1}\mathbf{v}_{k+2} \cdots \mathbf{v}_{k+j+1}\} = \mathbf{A}\{\mathbf{v}_k\mathbf{v}_{k+1} \cdots \mathbf{v}_{k+j}\} \tag{16}$$

with

$$\mathbf{A} \equiv \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \cdot & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdot & \mathbf{I} \\ -\mathbf{DC}^{(0)} & -\mathbf{DC}^{(1)} & -\mathbf{DC}^{(2)} & \cdot & -\mathbf{DC}^{(j)} \end{bmatrix}, \tag{17}$$

where $\mathbf{I}$ is the $(N + 1) \times (N + 1)$ identity submatrix. It should be observed that by virtue of the assumption that the weights $c_s^{(i)}$ do not vary with r, the matrix **A** does not change as contraction proceeds. Equation (2), therefore, in this case becomes simply:

$$\{\mathbf{v}_{k+1}\mathbf{v}_{k+2} \cdots \mathbf{v}_{k+j+1}\} = \mathbf{A}^{k+1}\{\mathbf{v}_0\mathbf{v}_1 \cdots \mathbf{v}_j\}. \tag{18}$$

It is familiarly known [4] that the stability of an iterated matrix depends upon its eigenvalues and eigenvectors and that if the matrix has a full complement of eigenvectors then a necessary and sufficient condition for its powers to remain bounded is that the moduli of its eigenvalues are less than or equal to unity. The eigenvalues and eigenvectors of **A** can be found, by virtue of the fact that the matrices $\mathbf{C}^{(i)}$ are circulant, or rather, by virtue of the fact that the weights $c_s^{(i)}$ do not vary with θ. Let λ be an eigenvalue of **A** and $\{\mathbf{x}_0, \mathbf{x}_1, \cdots \mathbf{x}_j\}$ an eigenvector, where each $\mathbf{x}_i$ has the same dimension as $\mathbf{v}_i$. The relation

$$\begin{bmatrix} \mathbf{0} & \mathbf{I} & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdot & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \mathbf{0} & \cdot & \mathbf{I} \\ -\mathbf{DC}^{(0)} & -\mathbf{DC}^{(1)} & \cdot & -\mathbf{DC}^{(2)} \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \mathbf{x}_2 \\ \cdot \\ \mathbf{x}_{j-1} \\ \mathbf{x}_j \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \mathbf{x}_2 \\ \cdot \\ \mathbf{x}_{j-1} \\ \mathbf{x}_j \end{bmatrix} \tag{19}$$

gives $\mathbf{x}_{i+1} = \lambda\mathbf{x}_i$ $(i = 0, 1, \cdots, j - 1)$ so that the eigenvectors of **A** are of the form:

$$\{\mathbf{x}_0\ \lambda\mathbf{x}_0\ \lambda^2\mathbf{x}_0 \cdots \lambda^j\mathbf{x}_0\}. \tag{20}$$

The vector $\mathbf{x}_0$ satisfies

$$(\lambda^{j+1}\mathbf{C}^{(j+1)} + \lambda^j\mathbf{C}^{(j)} + \lambda^{j-1}\mathbf{C}^{(j-1)} + \cdots + \lambda\mathbf{C}^{(1)} + \mathbf{C}^{(0)})\mathbf{x}_0 = \mathbf{0}, \tag{21}$$

which can be written as

$$\left(\lambda^{j+1} \sum_{s=0}^{N} c_s^{(j+1)}\mathbf{R}^s + \lambda^j \sum_{s=0}^{N} c_s^{(j)}\mathbf{R}^s + \cdots + \lambda \sum_{s=0}^{N} c_s^{(1)}\mathbf{R}^s + \sum_{s=0}^{N} c_s^{(0)}\mathbf{R}^s\right)\mathbf{x}_0 = \mathbf{0}, \tag{22}$$

where **R** is defined by Eq. (9). The eigenvectors of **R** are $\{1\ \omega\ \omega^2 \cdots \omega^N\}$, where ω ranges over the $(N + 1)$ roots of $\omega^{N+1} = 1$. If $\mathbf{x}_0 = \{1\ \omega\ \omega^2 \cdots \omega^N\}$, Eq. (22) gives:

$$\lambda^{j+1} \sum_{s=0}^{N} c_s^{(j+1)}\omega^s + \lambda^j \sum_{s=0}^{N} c_s^{(j)}\omega^s + \cdots + \lambda \sum_{s=0}^{N} c_s^{(1)}\omega^s + \sum_{s=0}^{N} c_s^{(0)}\omega^s = 0. \tag{23}$$

For each value of ω, $(j + 1)$ values of λ are determined and if these are all distinct, $(j + 1)$ distinct eigenvectors of **A** are found from (20). As ω ranges over the $(N + 1)$ roots of unity the full complement of eigenvectors of **A** are obtained and **A** is stable, provided that all the roots of Eq. (23) have moduli which are less than or equal to unity. This is the result stated at the beginning of the section.

It is interesting to comment upon the type of instability which occurs when the above conditions are relaxed. This is most easily discussed by consideration of the classical canonical form **B** to which **A** can be reduced by a similarity transformation [5]. The matrix **B** is composed of a number of primitive blocks (β_i) of the type

$$(\beta_i) = \begin{bmatrix} \lambda_i & 1 & 0 & \cdot & 0 \\ 0 & \lambda_i & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & 1 \\ 0 & 0 & 0 & \cdot & \lambda_i \end{bmatrix}. \tag{24}$$

Since **A** is similar to **B**, the powers of **A** are similar to the powers of **B** and the behavior of the latter is as the behavior of the powers of its primitive blocks (β_i). In case any of the roots λ of the system of Eqs. (23) has modulus greater than unity, then $\mathbf{B}^k$ increases exponentially like λ^k so that **B**, and consequently **A**, is unstable. In case all the roots of (23) have moduli less than or equal to unity but one or more has multiplicity two, then, at worst, $\mathbf{B}^k$ increases like

$$\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}^k = O(k). \tag{25}$$

More generally, if one of the roots has multiplicity m, then $\mathbf{B}^k = O(k^{m-1})$. In such instances, the matrix is subject to a controlled instablity, which may be suitable for many problems when the number of contracting steps inwards is not large.

5. Instability of the second type. In the example of Sec. 3, illustrating the second type of instability, Eq. (4) was solved for $u(1, n + 2)$. Provided $c_0^{(1)} \neq 0$, the same star could be used to compute $u(1, n)$ from $u(1, n + 2)$ and $u(1, n + 1)$; or alternatively, if $c_1^{(1)} \neq 0$, to compute $u(1, n + 1)$ in terms of $u(1, n + 2)$ and $u(1, n)$. In each case, however, a different stability condition is implied. The generalization to the more complex star (11) is quite simple but involves some notational inconveniences. Suppose that $c_p^{(j+1)} \neq 0$ and for each $n = 0, 1, 2, \cdots, N$ the value $u(k + j + 1, n + p)$ in (11) is computed from

$$\begin{aligned} u(k + j + 1, n + p) = {} & -(1/c_p^{(j+1)}) \sum_{s=0}^{N}{}' c_s^{(j+1)} u(k + j + 1, n + s) \\ & - (1/c_p^{(j+1)}) \sum_{i=0}^{j} \sum_{s=0}^{N} c^{(i)} u(k + i, n + s), \end{aligned} \tag{26}$$

where the prime on the sum means that $s \neq p$.

It is implied by (26) that an initial guess for $\mathbf{v}_{k+j+1}$ is made, $u(k + j + 1, n + p)$ is evaluated from Eq. (26) and then this newly computed value replaces the previous estimate. The same procedure is applied to each mesh point of S_{k+j+1} in turn and at the end of the first circuit the new values are compared with the old. If they do not agree, a second circuit is made and the process is continued until the solution of the system is obtained. With the indices corresponding to the angular argument reduced modulo $(N + 1)$, Eq. (26) can be written in matrix form as:

$$\begin{aligned} \begin{bmatrix} u(k + j + 1, n + 2 + p) \\ u(k + j + 1, n + 3 + p) \\ u(k + j + 1, n + 4 + p) \\ \cdot \\ u(k + j + 1, \quad n + p) \end{bmatrix} = {} & \mathbf{G} \begin{bmatrix} u(k + j + 1, n + 1 + p) \\ u(k + j + 1, n + 2 + p) \\ u(k + j + 1, n + 3 + p) \\ \cdot \\ u(k + j + 1, n - 1 + p) \end{bmatrix} \\ & - (1/c_p^{(j+1)}) \sum_{i=0}^{j} \mathbf{H}_i \mathbf{R}^n \mathbf{v}_{k+i}, \end{aligned} \tag{27}$$

where

$$\mathbf{G} \equiv \begin{bmatrix} 0 & 1 & \cdot & 0 \\ 0 & 0 & \cdot & 0 \\ 0 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 1 \\ -c_{p+1}^{(j+1)}/c_p^{(j+1)} & -c_{p+2}^{(j+1)}/c_p^{(j+1)} & \cdot & -c_{p-1}^{(j+1)}/c_p^{(j+1)} \end{bmatrix} \tag{28}$$

and $\mathbf{H}_i$ is the $N \times (N + 1)$ matrix:

$$\mathbf{H}_i \equiv \begin{bmatrix} 0 & 0 & 0 & \cdot & 0 \\ 0 & 0 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & 0 \\ c_0^{(i)} & c_1^{(i)} & c_2^{(i)} & \cdot & c_N^{(i)} \end{bmatrix} \tag{29}$$

From Eq. (27), at the end of a circuit:

$$\begin{bmatrix} u(k+j+1, n+1+p) \\ u(k+j+1, n+2+p) \\ \cdot \\ u(k+j+1, n-1+p) \end{bmatrix} = \mathbf{G}^{N+1} \begin{bmatrix} u(k+j+1, n+1+p) \\ u(k+j+1, n+2+p) \\ \cdot \\ u(k+j+1, n-1+p) \end{bmatrix} - (1/c_p^{(j+1)}) \sum_{i=0}^{j} \sum_{t=0}^{N} \mathbf{G}^{N-t}\mathbf{H}_i\mathbf{R}^{n+t}\mathbf{v}_{k+i}\ . \tag{30}$$

It is clear from Eq. (30) that in making the circuit around S_{k+j+1} any error which may be introduced will be propagated in a manner which depends upon the properties of the iterated matrix **G**. To guarantee stability of **G**, it is essential that its non-zero eigenvalues should be distinct and have moduli less than or equal to unity. But **G** is seen to be in its Jordan canonical form [5] so that its characteristic equation is simply:

$$c_p^{(j+1)}\mu^N + c_{p-1}^{(j+1)}\mu^{N-1} + c_{p-2}^{(j+1)}\mu^{N-2} + \cdots + c_{p+2}^{(j+1)}\mu + c_{p+1}^{(j+1)} = 0. \tag{31}$$

Thus, the second stability criterion is that the non-zero roots of (31) *should be distinct and should have moduli less than or equal to unity.*

If the solution of Eq. (27) is to be found iteratively by using the results after one circuit as the initial guess for the next and if this process is to converge, conditions slightly stronger than those for stability must be satisfied. *It is essential for convergence that the moduli of the roots of* (31) *should be less than unity.*

It will be recalled that the first stability conditions required that the circulant matrix $\mathbf{C}^{(j+1)}$ should have an inverse. This is equivalent to the condition that it has no zero eigenvalues. If $\mathbf{C}^{(j+1)}$ is written:

$$\mathbf{C}^{(j+1)} = \sum_{s=0}^{N} c_s^{(j+1)}\mathbf{R}^s. \tag{32}$$

then its eigenvalues are given by:

$$\sum_{s=0}^{N} c_s^{(j+1)}\omega^s, \tag{33}$$

where $\omega^{N+1} = 1$. This is equivalent to:

$$\sum_{s=0}^{N} c_s^{(j+1)} \omega^s = \sum_{s=0}^{N} c_{p-s}^{(j+1)} \omega^{(p-s)} = \omega^{p-N} \sum_{s=0}^{N} c_{p-s}^{(j+1)} \omega^{N-s}, \tag{34}$$

where the subscript indices of the weight factors are reduced modulo $N + 1$. From this it follows that a sufficient condition for $\mathbf{C}^{(j+1)}$ to be regular is that the equation

$$\sum_{s=0}^{N} c_{p-s}^{(j+1)} \mu^{N-s} = 0 \tag{35}$$

should have no roots on the unit circle. Identifying this with (31) it is seen that the criterion for convergence is sufficient to ensure the regularity of $\mathbf{C}^{(j+1)}$.

6. Example of an equation of hyperbolic type. It is required to find the solution: $u = u(r, \theta)$, of the hyperbolic equation $r^2 (\partial^2 u/\partial r^2) - (\partial^2 u/\partial \theta^2) = 0$ in the region $0 < r \leqq 1$, satisfying $u(1, \theta) = f_1(\theta)$, $\partial u(1, \theta)/\partial r = f_2(\theta)$ on the boundary and $| u(r, \theta) | \leqq M$ in the interior of this region, where M is some constant. As mesh, let S_0 be the unit circle, S_k the circle of radius r_k and choose $r_{k+1}/r_k = \rho$ where ρ is a constant $0 < \rho < 1$; also, let $u(k, n) \equiv u(r_k, n\Delta\theta)$, where $\Delta\theta = 2\pi/(N + 1)$ for some integer N.

The first boundary conditions are satisfied by taking $u(0, n) = f_1(n\ \Delta\theta)$. The derivative appearing in the second boundary condition may be replaced by its divided difference approximation: $\partial u(1, \theta)/\partial r \approx [u(0, n) - u(1, n)]/(1 - \rho)$ so that: $u(1, n) = u(0, n) - (1 - \rho) f_2(n\Delta\theta)$.

It is possible to select a suitable star for this problem that involves eleven nodal points so that only ten multiplications are necessary to compute one of the nodal values from the rest. A typical stencil is given in Fig. 2.

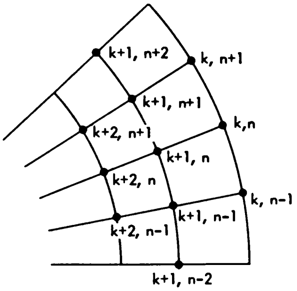

Fig. 2. Stencil for hyperbolic equation

In this case, the solution on S_{k+2} is determined always in terms of its known values on S_{k+1} and S_k so that $j + 1 = 2$ in (11).

An approximation for the second derivatives appearing in the differential equation may be defined by taking linear combinations of the fundamental difference approximations at different points in the stencil. Thus for any selection of the real numbers a b, c such that $a + b + c = 1$, a difference approximation may be defined

$$\begin{aligned} \frac{\partial^2 u}{\partial \theta^2} \approx \frac{1}{\Delta\theta^2} \{ & a[u(k + 1, n + 2) - 2u(k + 1, n + 1) + u(k + 1, n)] \\ & + b[u(k + 1, n + 1) - 2u(k + 1, n) + u(k + 1, n - 1)] \\ & + c[u(k + 1, n) - 2u(k + 1, n - 1) + u(k + 1, n - 2)]\}. \end{aligned} \tag{36}$$

In a similar way, using divided differences:

$$r_k^2 \frac{\partial^2 u}{\partial r^2} \approx \frac{2}{\rho(1-\rho)^2(1+\rho)} \{d[u(k+2, n+1) - (1+\rho)u(k+1, n+1) \\ + \rho u(k, n+1)] + e[u(k+2, n) - (1+\rho)u(k+1, n) + \rho u(k, n)] \\ + f[u(k+2, n-1) - (1+\rho)u(k+1, n-1) + \rho u(k, n-1)]\}, \tag{37}$$

where $d + e + f = 1$. The stability criteria may be satisfied for a wide range of values of the parameters a, b, c, d, e, f and this fact may be used in actual practice to assist in the selection of an approximating star which is the most accurate possible, compatible with these conditions. However, we are not here concerned with the problem of accuracy and we therefore assign the parameters in such a way that the resulting algebra is somewhat simplified. Choosing $a = 1$, $b = -1/4$, $c = 1/4$; $d = 1$, $e = -1/4$, $f = 1/4$, the following difference approximation to $r^2\ \partial^2 u/\partial r^2 - \partial^2 u/\partial \theta^2 = 0$ results:

$$\frac{2}{\rho(1-\rho)^2(1+\rho)} [u(k+2, n+1) - \tfrac{1}{4}u(k+2, n) + \tfrac{1}{4}u(k+2, n-1)] \\ - \frac{1}{\Delta\theta^2} u(k+1, n+2) + \left(\frac{9}{4\Delta\theta^2} - \frac{2}{\rho(1-\rho)^2}\right) u(k+1, n+1) \\ + \left(\frac{1}{2\rho(1-\rho)^2} - \frac{7}{4\Delta\theta^2}\right) u(k+1, n) \\ + \left(\frac{3}{4\Delta\theta^2} - \frac{1}{2\rho(1-\rho)^2}\right) u(k+1, n-1) - \frac{1}{4\Delta\theta^2} u(k+1, n-2) \\ + \frac{2}{(1-\rho)^2(1+\rho)} [u(k, n+1) - \tfrac{1}{4}u(k, n) + \tfrac{1}{4}u(k, n-1)] = 0. \tag{38}$$

From Eq. (12) the first stability equation is

$$\frac{2(\omega - \frac{1}{4} + \frac{1}{4}\omega^{-1})}{\rho(1-\rho)^2(1+\rho)} \lambda^2 - \left[\frac{1}{\Delta\theta^2}\omega^2 + \left(\frac{2}{\rho(1-\rho)^2} - \frac{9}{4\Delta\theta^2}\right)\omega \right. \\ \left. + \left(\frac{7}{4\Delta\theta^2} - \frac{1}{2\rho(1-\rho)^2}\right) + \left(\frac{1}{2\rho(1-\rho)^2} - \frac{3}{4\Delta\theta^2}\right)\omega^{-1} + \frac{1}{4\Delta\theta^2}\omega^{-2}\right]\lambda \\ + \frac{2(\omega - \frac{1}{4} + \frac{1}{4}\omega^{-1})}{(1-\rho)^2(1+\rho)} \equiv \frac{2(\omega - \frac{1}{4} + \frac{1}{4}\omega^{-1})}{\rho(1-\rho)^2(1+\rho)} \left\{\lambda^2 \right. \\ \left. \frac{\rho(1-\rho)^2(1+\rho)}{2}\left[\frac{1}{\Delta\theta^2}\omega + \left(\frac{2}{\rho(1-\rho)^2} - \frac{2}{\Delta\theta^2}\right) + \frac{1}{\Delta\theta^2}\omega^{-1}\right]\lambda + \rho\right\} = 0, \tag{39}$$

where $\omega = \exp\{2n\pi i/(N+1)\}$, $(n = 0, 1, \cdots, N)$. It is advantageous geometrically to take $\Delta\theta = (1-\rho)$, so that Eq. (39) reduces to:

$$\lambda^2 - (1+\rho)\left[(1-\rho) + \rho\cos\left(\frac{2n\pi}{N+1}\right)\right]\lambda + \rho = 0. \tag{40}$$

A quadratic equation: $x^2 + px + q = 0$ with real coefficients has roots with moduli less than or equal to unity in case $1 \geqq q \geqq |\ p\ | - 1$. Application of this result easily demonstrates that the roots of Eq. (40) have moduli less than or equal to unity. The equation

$$\lambda^2 - (1 + \rho)[(1 - \rho) + \rho \cos \phi]\lambda + \rho = 0, \tag{41}$$

has equal roots only if

$$\cos \phi = \frac{\rho^2 \pm 2(\rho)^{\frac{1}{2}} - 1}{\rho(1 + \rho)}, \tag{42}$$

and Eq. (40) will have equal roots only if $\phi = 2n\pi/(N + 1)$ for some $n = 0, 1, 2, \cdots, N$. Since Eq. (42) can be satisfied, if at all, only for isolated values of n, it is not difficult to select values of N so that the roots of Eq. (40) are distinct.

The second stability equation, from (38), is seen to be

$$\mu^2 - \tfrac{1}{4}\mu + \tfrac{1}{4} = 0, \tag{43}$$

and it is readily verified that the roots of this equation are distinct with moduli less than unity.

With the boundary conditions taken to be $f_1 = \cos \theta$, $f_2 = 1/2 \cos \theta$, the analytic solution: $u = r^{1/2} \cos [(3)^{1/2} (\log r)/2] \cos \theta$ for the differential equation can be found. A comparison between the computed and the actual values indicates that with $(N + 1)$ as great as 1000, stability was indicated for as many as 250 steps inwards so that a mesh of 250,000 nodal points was involved.

7. Example of an equation of elliptic type. The most important application of the boundary contraction method is to partial differential equations of elliptic type. No essential difficulty exists as a consequence of the well-known distinction between elliptic and other types of partial differential equations. In practice, however, it is not easy to find difference schemes for this class of equations that fulfill the stability requirements. In another paper [6], the authors have presented a detailed analysis of Laplace's equation in the circle. It was found that for this equation a modified version of the method presented here was preferable. However, the stability criteria related back to a simple form of those which have been given here. Solutions have been computed over a mesh of 2,500 points and accuracy to six significant figures has been obtained for a variety of boundary conditions. Stability was verified up to 200,000 nodal points but similar accuracy cannot be claimed for meshes of this size.

8. More general forms of stars. By arguments analogous to those used in the foregoing paragraphs, the contraction method may be shown to be valid equally well for stars of the form

$$\begin{aligned}\sum_{s=0}^{N} c_s^{(j+1)} u(k + j + 1, n + s) + \sum_{s=0}^{N} c_s^{(j)} u(k + j, n + s) + \cdots \\ + \sum_{s=0}^{N} c_s^{(0)} u(k, n + s) = F(r, \theta)\end{aligned} \tag{44}$$

which correspond to a non-homogeneous linear partial differential equation. Equation (44) can be expressed as:

$$\{\mathbf{v}_{k+1}\mathbf{v}_{k+2} \cdots \mathbf{v}_{k+j+1}\} = \mathbf{A}\{\mathbf{v}_k\mathbf{v}_{k+1} \cdots \mathbf{v}_{k+j}\} + \mathbf{B}_k, \tag{45}$$

where $\mathbf{B}_k$ is a matrix independent of the $\mathbf{v}_i$. It follows readily that the stability properties of Eq. (45) depend only on those of the matrix $\mathbf{A}$, which has been fully discussed.

A further generalization is possible to a system in which the weights $c_s^{(i)}$ may vary with the radius, as the contraction moves inwards. This implies from Eq. (1) that the matrices $\mathbf{A}_k$ depend upon r. It is found that sufficient conditions for stability of the first type for a star such as:

$$\sum_{s=0}^{N} c_s^{(j+1)}(r_{k+j+1})u(k+j+1, n+s) + \sum_{s=0}^{N} c_s^{(j)}(r_{k+j})u(k+j, n+s) + \cdots + \sum_{s=0}^{N} c_s^{(0)}(r_k)u(k, n+s) = F(r, \theta) \tag{46}$$

are: (a) that the mesh be taken sufficiently fine, and (b) for each ω, such that $\omega^{N+1} = 1$, the roots of the equations

$$\lambda^{j+1} \sum_{s=0}^{N} c_s^{(j+1)}(r_{k+j+1})\omega^s + \lambda^j \sum_{s=0}^{N} c_s^{(j)}(r_{k+j})\omega^s + \cdots + \sum_{s=0}^{N} c_s^{(0)}(r_k)\omega^s = 0 \tag{47}$$

should be distinct and should have moduli less than unity, uniformly in k. This result follows from the proposition [7] that if a matrix $\mathbf{A} = \mathbf{A}(r)$ is continuously dependent upon a parameter r which takes on a sequence of values: $\{r_i \mid i = 0, 1, 2, \cdots\}$ in a closed bounded region, then the matrix product $\prod_i \mathbf{A}(r_i)$ is bounded, provided $\mathbf{A}(r)$ has a full completement of eigenvectors, the moduli of the eigenvalues of $\mathbf{A}(r)$ are uniformly less than unity and $\delta = \max_i [|r_{i+1} - r_i|]$ is sufficiently small. The second stability criterion for Eq. (46) becomes simply that the non-zero roots of

$$\sum_{n=0}^{N} c_{p-n}^{(j+1)}(r_{k+j+1})\mu^{N-n} = 0 \tag{48}$$

should be distinct and have moduli less than unity.

The generalization to stars in which the weights $c_s^{(i)}$ are dependent upon the angular argument has not yet been made.

9. Further investigations. Further investigations of the contraction method are being made by the authors. The method is being generalized to include other classes of partial differential equations and to stars which depend on the angular as well as the radial arguments. The possible applications of the method to irregularly shaped boundaries, to multiply connected regions, to mixed boundary value problems and to non-linear equations will also be considered.

Acknowledgements. The authors are much indebted to Dr. R. Herman, Dr. D. C. Gazis and Dr. R. Bartels, University of Michigan, for their helpful suggestions and criticisms. They are also grateful to Mrs. E. McManus and Miss P. Bakey for their assistance in programming the calculations for the IBM 704 computer at the Research Laboratories.

References

1. E. Bodewig, *Matrix calculus*, Interscience Publishers, New York, 1956, p. 182 ff
2. L. Collatz, *Numerische Behandlung von Differentialgleichungen*, Springer Verlag, 2nd ed., 1955, p. 320 ff
3. R. V. Southwell, *Relaxation methods in engineering science*, Oxford University Press, 1940
4. E. Bodewig, *Matrix calculus*, Interscience Publishers, 1956, p. 83
5. Van der Waerden, *Modern algebra*, vol. II, Frederic Unger Publishing Co., New York, 1950, p. 120-121
6. H. W. Milnes and R. B. Potts, *Boundary contraction solution of Laplace's differential equation*, J. Assoc. Comp. Mach. 6 (1959) 226
7. H. W. Milnes, *Bounded continuous matrix products*, Michigan Math. J. 6 (1959)