H. P. Flatt

Chain Matrices and the Crank-Nicolson Equation

In Reference 1, a practical application of the special properties of a class of matrices called *chain matrices* is made to obtain numerical solutions of partial differential equations of the elliptic type. In this communication, we wish to present an application of these matrices to the study of the problem of obtaining numerical solutions of parabolic partial differential equations. This method is not proposed as a general method for obtaining a numerical solution to parabolic equations, for its limitations are obvious. Its use, however, does allow us to obtain some interesting theoretical results.

Under the proper conditions, solutions of parabolic partial differential equations satisfy a maximum principle. In a one-dimensional problem, for example, this principle states that the value of the solution at the interior point (x_0, t_0) does not exceed the largest and least value of the solution at any point on the boundary (x, t) such that $t \leq t_0$. This principle provides a powerful tool for the analysis of solutions of such equations.

In order to obtain a numerical solution of a parabolic partial differential equation, one generally replaces the equation by a finite difference equation. The numerical solution of this equation is then obtained as an approximation to the solution of the original equation. If the solution of the finite difference equation also satisfies a maximum principle, then one can use this to establish the convergence of the solution to the solution of the differential equation. This is used by Douglas, for example, in Reference 2. Also, from a computational point of view, one wishes the approximate solution to "behave" in the same manner as the exact solution.

For such difference equations, Douglas[3] has given an example of a stable difference equation which does not satisfy the maximum principle. By use of the properties of chain matrices, we are able to construct other such examples. A by-product of this construction is a relation which gives some insight into the problem of how to choose the mesh ratio so that the maximum principle will be satisfied.

Chain matrices

A chain matrix C_N of order N is a square matrix whose elements $c_{m,n} (m, n = 1, 2, \cdots, N)$ satisfy the recurrence relationship

$$c_{m,n-1} + c_{m,n+1} = c_{m-1,n} + c_{m+1,n}, \tag{2.1}$$

where zero is substituted for any element with one index 0 or $N + 1$.

Examples:

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \quad \begin{bmatrix} 1 & 0 & 3 \\ 0 & 4 & 0 \\ 3 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

An immediate consequence of this definition is that a chain matrix is uniquely determined by the elements of the first row, and, symbolically, we may write

$$C_N = \{c_{1,1}; c_{1,2}; \cdots ; c_{1,N}\}. \tag{2.2}$$

In Reference 1, the following properties of a chain matrix are established: If C_N is a chain matrix, and if

k is a scalar, then kC_N is a chain matrix. (2.3a)

B_N is a chain matrix, then $C_N + B_N$ is a chain matrix. Hence any finite linear combination of chain matrices is a chain matrix. (2.3b)

B_N is a chain matrix, then $B_N C_N$ is a chain matrix. (2.3c)

Further, if we say that the elements $c_{i,i}$ $(i = 1, \cdots, N)$ are on the main diagonal, and the elements $c_{i,N+1-i}$ $(i = 1, \cdots, N)$ are on the "second" diagonal, then a further useful property of the chain matrix is that it is symmetric about the main and second diagonals.

Consider now the chain matrices

$$E_{N,j} = \{\delta_{1,j}; \delta_{2,j}; \cdots ; \delta_{N,j}\}, \tag{2.4}$$

where $\delta_{i,j}$ is the Kronecker delta; i.e., $\delta_{i,j} = 1$ if $i = j$, and $\delta_{i,j} = 0$ if $i \neq j$. Note that $E_{N,1}$ is just the identity matrix. It is easily seen that if $C_N = \{c_{1,1}; c_{1,2}; \cdots ; c_{1,N}\}$, then

$$C_N = \sum_{j=1}^{N} c_{1,j} E_{N,j}. \tag{2.5}$$

Further, we may note that the matrix, all of whose elements are zero, is a chain matrix, and therefore, the set of all chain matrices of order N form a vector space of dimension N over the field of real or complex numbers, and the elements $E_{N,i}$ may be taken as a basis for the vector space.

Not all chain matrices have inverses. For example, the chain matrix $E_{3,2}$ is singular, and hence does not possess an inverse. However, as an almost immediate consequence of the Cayley-Hamilton theorem, it follows that if the chain matrix C_N is nonsingular, then its inverse C_N^{-1} is a chain matrix.

It shall be important later to note the following result, which follows immediately from (2.5):

Theorem 2.1: *If the elements* $c_{1,j}$ $(j = 1, 2, \cdots, N)$ *of* C_N *are positive, then all of the elements of* C_N *are positive.*

Proof: It is sufficient to note that all elements of $E_{N,i}$ are nonnegative and that there is no zero element in the matrix $\sum_{i=1}^{N} E_{N,i}$.

Although further properties of chain matrices may be determined, it is of interest to consider an example wherein such matrices may arise.

Example

We shall consider the following problem: To obtain in R: $\{0 \leq x \leq 1, t \geq 0\}$ a numerical solution of the equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < 1, \qquad t > 0 \tag{3.1}$$

subject to the conditions

$$u(0, t) = u(1, t) = 0 \qquad t \geq 0 \tag{3.2a}$$

$$u(x, 0) = f(x) \qquad 0 < x < 1. \tag{3.2b}$$

We impose on R a lattice with grid points $x_m = m\Delta x$' $m = 0, 1, \cdots, N + 1$; $t_n = n\Delta t$, $n = 0, 1, 2, \cdots$, with $\Delta x = 1/(N + 1)$. The quantity N is an integer, and we shall always assume that N is an odd integer in order that the line $x = 0.5$ will have grid points on it. This is no essential restriction, for the modifications required in the work following are obvious. We choose $\Delta t = r(\Delta x)^2$, where r is a preassigned constant called the mesh ratio. We denote $g(m\Delta x, n\Delta t)$ by $g_{m,n}$.

We consider the following difference equation as an approximation to (3.1)–3.2b):

$$\begin{aligned} w_{m,0} &= u_{m,0} = f(x_m) \qquad 1 \leq m \leq N \\ w_{m,n+1} - w_{m,n} &= (r/2)(\Delta^2 w_{m,n+1} + \Delta^2 w_{m,n}), \\ &\qquad 1 \leq m \leq N, n \geq 0 \\ w_{0,n} &= w_{N+1,n} = 0 \qquad n \geq 0, \end{aligned} \tag{3.3}$$

where

$$\Delta^2 w_{m,k} = w_{m-1,k} - 2w_{m,k} + w_{m+1,k}. \tag{3.4}$$

These are difference equations arising from the use of the Crank-Nicolson procedure.[4] Equation (3.3) may be written in matrix notation as

$$A_N \mathbf{W}_{n+1} = -B_N \mathbf{W}_n, \tag{3.5}$$

where

$$\mathbf{W}_k = \begin{bmatrix} w_{1,k} \\ \vdots \\ w_{N,k} \end{bmatrix} \tag{3.6}$$

and A_N and B_N are the chain matrices

$$\begin{aligned} A_N &= \{a; 1; 0; \cdots; 0\} \\ B_N &= \{b; 1; 0; \cdots; 0\}. \end{aligned} \tag{3.7}$$

The value of a is $-2(1 + 1/r)$ and the value of b is $-2(1 - 1/r)$.

Since $|a| > 2$, we know by a theorem of Hadamard[5] that A_N is nonsingular, and hence A_N^{-1} exists. We may then write (3.5) as

$$\mathbf{W}_{n+1} = C_N \mathbf{W}_n, \tag{3.8}$$

where

$$C_N = -A_N^{-1} B_N. \tag{3.9}$$

Using properties previously noted, it is seen that C_N is also a chain matrix.

The elements of C_N

It is possible to write down explicitly the elements of the first row of C_N. Let $|A_N|$ denote the determinant of A_N and let $|A_0| = 1$ and $|A_1| = a$. Then

$$|A_k| = a\,|A_{k-1}| - |A_{k-2}|. \tag{4.1}$$

The solution of this difference equation is

$$|A_k| = (-1)^k[\sinh (k + 1)y/\sinh y], \tag{4.2}$$

where

$$y = \log[1 + (1/r) + \sqrt{(1 + 1/r)^2 - 1}]. \tag{4.3}$$

Let

$$\beta_k = \frac{|A_{N-k}|}{|A_N|}. \tag{4.4}$$

It is easily seen that $A_N^{-1} = \{\beta_1; +\beta_2; \beta_3; \cdots; \beta_N\}$.

Note that it follows from (4.1) and (4.4) that $\beta_{k+2} = a\beta_{k+1} - \beta_k$ $(k = 1, 2, \cdots, N - 2)$, and that $\beta_0 = 1$. Using this relationship, one finds easily by matrix multiplication that

$$c_{1,k} = (-1)^k(4/r)\beta_k - \delta_{k,1} \qquad (k = 1, 2, \cdots, N). \tag{4.5}$$

Then, from (4.1) and (3.9),

$$c_{1,k} = \frac{4}{r}\frac{\sinh (N - k + 1)y}{\sinh (N + 1)y} - \delta_{k,1} \qquad (k = 1, 2, \cdots, N). \tag{4.6}$$

Hence for all $r > 0$, and for $r > 1$, these elements are positive. However,

$$c_{1,1} = \frac{4}{r}\frac{\sinh Ny}{\sinh (N + 1)y} - 1. \tag{4.7}$$

Therefore, in order that $c_{1,1}$ be positive, it is necessary that the following condition hold:

$$r < 4\frac{\sinh Ny}{\sinh (N + 1)y}. \tag{4.8}$$

For N sufficiently large, the condition (4.8) is approximately satisfied if

$$r < 4e^{-y}, \tag{4.9}$$

i.e.,

$$r < 4/[1 + 1/r + \sqrt{(1 + 1/r)^2 - 1}]. \tag{4.10}$$

The relation (4.10) is equivalent to the requirement $r < 4 - 2\sqrt{2}$.

We may then state the following theorem:

Theorem 4.1: *For $r > 0$, the element $c_{1,1}$ is positive if and only if $r < \gamma$, where γ is the least upper bound of r such that*

$$r < 4\frac{\sinh Ny}{\sinh (N + 1)y}.$$

For N sufficiently large, γ is approximately $4 - 2\sqrt{2}$.

The importance of this result lies in Theorem 2.1, for, if $c_{1,1} > 0$, then all elements of C_N are positive. On the other hand, if $c_{1,1} < 0$, then the value of each of the elements on the main diagonal of C_N is decreased. In fact, by (2.5),

$$C_N = c_{1,1}E_{N,1} + \sum_{j=2}^{N} c_{1,j}E_{N,j}. \tag{4.11}$$

Since $E_{N,1}$ is just the identity matrix, if the value of $c_{1,1}$ is changed by any given amount, the value of each of the remaining elements of the main diagonal is changed by exactly the same amount.

Counterexample

Theorem 4.1 allows us immediately to construct an example to show that the system (3.3), although stable in the L_2 sense, has a solution which does not satisfy the maximum principle. Choose

$$f(x_m) = \delta_{m,1}(m = 1, 2, \cdots, N). \tag{5.1}$$

Then $w_{1,1} = c_{1,1}$, and by Theorem 4.1, $w_{1,1} < 0$ for $r > \gamma$. We may state this as a theorem:

Theorem 5.1: *If $f(x_m) = \delta_{m,1}$ ($m = 1, 2, \cdots, N$), the solution of (3.3) violates the maximum principle if $r > \gamma$.*

Two things should be noted. First, the value of γ depends in general upon the length of the interval, which has been assumed to be of unit length in this note. Secondly this example, while depending upon an unusual initial condition, actually is significant because it shows the smallest value of r for which the maximum principle is violated.

Let

$$S_m = \sum_{k=1}^{N} c_{m,k} \qquad m = 1, 2, \cdots, N. \tag{5.2}$$

Theorem 5.2: $S_m < 1 \qquad m = 1, 2, \cdots, N.$

Proof: From Eq. (2.1), we see that

$$\sum_{k=1}^{N} c_{m+2,k} = \sum_{k=1}^{N} c_{m+1,k-1} + \sum_{k=1}^{N} c_{m+1,k+1} - \sum_{k=1}^{N} c_{m,k}. \qquad (m = 0, 1, \cdots, N - 2). \tag{5.3}$$

By a reordering of the above relation, we find that S_m satisfies the relation

$$S_{m+2} = 2S_{m+1} - S_m - [c_{1,m+1} + c_{1,n-m}]. \tag{5.4}$$

By definition, $S_0 = 0$, and by direct calculation

$$S_1 = \frac{2}{r}\frac{\sinh (N/2)y}{\sinh y/2 \cosh (N + 1/2)y} - 1. \tag{5.5}$$

We may use these initial conditions to solve the difference equation. We find that

$$S_m = 1[-\mu/(e^y - 1)^2][e^{(N-m+1)y} + e^{my}] + \delta_{m,0}, \tag{5.6}$$

where

$$\mu = \frac{2}{r}\frac{e^{-(N-1/2)y}}{\cosh (N + 1/2)y}. \tag{5.7}$$

We observe that

$$S_{m+1} - S_m = \frac{-\mu}{e^y - 1}[e^{my} - e^{(N-m)y}], \tag{5.8}$$

and so

$$S_1 < S_2 < \cdots < S_{(N+1)/2} > S_{(N+3)/2} > \cdots > S_{N-1} > S_N.$$

The maximum value for S_m occurs for $m = N + 1/2$, for which we have

$$S_{N+1/2} = 1 - \frac{2}{\cosh (N + 1/2)y} < 1. \tag{5.9}$$

In particular, if $r < \gamma$, then clearly $S_1 > 0$, and since then $0 < S_m < 1$ for $m = 1, 2, \cdots, N$, we see from Eq. (3.8) that the condition $r < \gamma$ is a *sufficient* condition in order that the maximum principle be satisfied. We have then, the following result:

Theorem 5.3: *A necessary and sufficient condition that the solution of the discrete system defined by (3.3) satisfy the maximum principle for arbitrary $f(x)$, $0 < x < 1$, is that $0 < r < \gamma$, where γ is defined in Theorem 4.1.*

It should be explicitly pointed out that for some choices of initial conditions larger mesh ratios may be used without violating the maximum principle. For example, if $f(x) \equiv 1$, $0 < x < 1$, then the maximum principle is satisfied if $r < 4$. The conditions which must be imposed upon $f(x)$ in order to use larger mesh ratios are not understood at the present time.

Remarks

The various examples given above point out again that the use of implicit difference methods to obtain numerical solutions of parabolic partial differential equations may contain some drawbacks. Theorem 4.1 shows that in order to be completely safe without any further analysis, r should be chosen less than $4 - 2\sqrt{2}(\cong 1.172)$ if one uses the Crank-Nicolson method. However, the use of this method involves about four times as much work per point as does the use of the forward difference method. Thus, it is clear that in certain circumstances, it is more efficient to use the forward difference method than the Crank-Nicolson method.

Note again, however, that the backward difference method, i.e., the use of the difference equation

$$W_{m,n+1} - W_{m,n} = r\Delta^2 W_{m,n+1} \tag{6.1}$$

instead of (3.3), does not have this particular drawback of the Crank-Nicolson method, for, as shown in Reference 2, solutions of this difference equation do satisfy the maximum principle.

References

1. T. Burgerhout, "On the Numerical Solution of Partial Differential Equations of the Elliptic Type-I," *App. Sci. Res., Series B*, **4**, 161–172 (1954).
2. J. Douglas, "On The Numerical Integration of Quasi-linear Parabolic Differential Equations," *Pacific J. Math.*, **6**, 35–42 (1956).
3. J. Douglas, "A Note on the Numerical Solution of Parabolic Differential Equations," Unpublished paper, Abstract in *Bull. Amer. Math. Soc.*, **62**, 588 (1956).
4. J. Crank and P. Nicolson, "A Practical Method for Numerical Evaluation of Solutions of Partial Differential Equations of the Heat-Conduction Type," *Proc. Camb. Phil. Soc.*, **43**, 50–67 (1947).
5. E. Bodewig, *Matrix Calculus*, 2nd ed., Interscience, New York, 1959, p. 79.

Received March 3, 1965