BULLETIN OF THE GREEK MATHEMATICAL SOCIETY 20(1979)

THE SOLUTION OF PARABOLIC PARTIAL DIFFERENTIAL EQUATIONS BY BOUNDARY VALUE METHODS

By
D.J. Evans*

1. Introduction

The solution of scientific problems defined by partial differential equations presents such analytical difficulties that numerical methods appear to be the only reasonable means of solution within the foreseeable future.

Usual numerical methods of solution for the initial boundary-value type problem posed by parabolic equations have been the explicit and implicit (Crank-Nicolson) methods in which the solution is obtained by a point or line step by step process on a chosen network of lines over the given open domain i.e. the semi-infinite strip R.

An alternative strategy which does not suffer from the row-to-row error accumulation of the initial value technique is to determine, the steady state solution u as $t\to\infty$. Then with the assumption that this is the solution on row t=T, we solve the resulting boundary value problem on the

*On leave of absence from Department of Computer Studies, University of Technology Loughborough Leicestershire, U.K.

truncated closed region $\tilde{R}$, (Greenspan, 1974).

2. Statement of Problem

Consider the solution of the parabolic partial differential equation

$$u_{xx} - u_t = f(x,t,u,u_x) \qquad (2.1)$$

in the semi-infinite strip $R\{(0 \leq x \leq d),\ t \geq 0\}$, given the initial condition,

$$u(x,0) = g(x), \quad 0 \leq x \leq d, \qquad (2.2a)$$

and boundary conditions,

$$\left.\begin{aligned} u(0,t) &= g_1(t), \quad t \geq 0, \\ u(d,t) &= g_2(t), \quad t \geq 0. \end{aligned}\right\} \qquad (2.2b)$$

Now we assume that equation (2.1) subject to the initial boundary conditions (2.2) has a solution at $t=\infty$, i.e. the steady state solution, which is determined by the boundary value problem,

$$d^2u/dx^2 = f(x,u,u_x), \quad 0 \quad x \leq d, \qquad (2.3)$$

with $u(0) = a, \quad u(d) = b,$

where $\lim_{t\to\infty} g_1(t) = a, \quad \lim_{t\to\infty} g_2(t) = b.$

The solution to (2.3) can be easily obtained by applying iteration techniques to the finite difference equations

$$(u_{i+1}-2u_i+u_{i-1})/h^2 = f(x_i,u_i,(u_{i+1}-u_{i-1})/2h), 1 \leq i \leq n-1, nh=d, (2.5)$$

where we have assumed that f has continuous derivatives and are bounded in the interval [0,d]. For a linear function f, equation (2.5) results in a tridiagonal system of equa -

tions which can be solved by a Gaussian elimination procedure for the solution values for t=T, whereas for a non-linear function, the set of (n-1) non-linear equations can be solved by Newton's method (Keller, 1968).

3. Formulation of the Boundary Value Problem

We consider now the reformulated boundary value problem,

$$u_{xx} - u_t = f(x,t,u,u_x), \tag{3.1}$$

over the rectangle $\bar{R}\equiv\{(x,t)\ 0 \le x \le d,\ 0 \le t \le T\}$ with boundary ∂R. For the numerical solution of this differential equation we impose a uniform (n×m) rectangular grid of mesh sizes h,k respectively on $R=\bar{R}\cup\partial R$ and by use of the following central difference operators:

$$u_{xx} = (u_{i-1,j}-2u_{i,j}+u_{i+1,j})/h^2+O(h^2), \tag{3.2}$$

and $$u_t = (u_{i,j+1}-u_{i,j-1})/2k+O(k^2),$$

we approximate the equation (3.1) at each internal node (i, j) by the 5 point finite difference equation,

$$u_{i,j+1}-2ru_{i-1,j}+4ru_{i,j}-2ru_{i+1,j}-u_{i,j-1}$$
$$= -2kf(x_i,t_j,u_{i,j},\{u_{i+1,j}-u_{i-1,j}\}/2h,)$$
$$\text{for } i=1(1)n-1,\ j=1(1)m-1, \tag{3.3}$$

where $r=k/h^2$.

We now consider in detail the case when f is a linear function for here it is worth mentioning that equation (3.3) then represents the Richardson unstable formula for the diffusion equation when used in a marching type solution procedure (Forsythe & Wasow, 1960).

Consider now a columnwise ordering of the mesh points in the domain R such that for each point (i,j) the totality of the linear difference equations (3.3) (i.e. f=f(x,t)) produced in this way yields a (N×N) linear system of the form,

$$A\underline{u} = \underline{b}, \tag{3.4}$$

where

$$A = \begin{bmatrix} D_1 & F_1 & & & \\ E_1 & D_2 & F_2 & & \mathbf{0} \\ & \ddots & \ddots & \ddots & \\ \mathbf{0} & & & D_{n-2} & F_{n-2} \\ & & & E_{n-2} & D_{n-1} \end{bmatrix}_{(n-1)\times(n-1)} = D + E + F, \tag{3.5}$$

with

$$D_i = \begin{bmatrix} 4r & 1 & & \mathbf{0} \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ \mathbf{0} & & -1 & 4r \end{bmatrix}_{(m-1)\times(m-1)}, E_i = F_i = -2rI_{(m-1)\times(m-1)}, \tag{3.6}$$

where $m=T/k$, $n=d/h$ and $N=(n-1)\times(m-1)$, and $\underline{b}$ is obtained by inserting the known boundary values (when applicable) into the equations (3.5).

When f is a linear function of the form $f(x,t,u,u_x)$ then the difference equation assumes the form,

$$u_{i,j+1}-(2r+\frac{k}{h}\ f)u_{i-1,j}+(4r+2kf)u_{i,j}-(2r-\frac{k}{h}\ f)u_{i+1,j}-u_{i,j-1}$$
$$= -2kf(x_i,t_j), \tag{3.7}$$

and consequently we have

$$D_i = \begin{bmatrix} 4r+2kf & 1 & & & \mathbf{0} \\ -1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ \mathbf{0} & & & -1 & 4r+2kf \end{bmatrix}, \quad E_i = -(2r+kf/h)I, \text{ and } F_i = -(2r-kf/h)I, \tag{3.8}$$

for i=1(1)n-1.

It can be readily discerned from (3.5) and (3.6) that A is a block tridiagonal matrix where the D_i, E_i and F_i submatrices are of tridiagonal Toeplitz and diagonal forms respectively.

When f is a mildly non-linear function, the above analysis results in a similar system of non-linear equations of the form,

$$A\underline{u} = \underline{b}(u),$$

which can be solved by a generalised Newton method (Greenspan, 1974).

Similarly, we can consider a row-wise ordering of the mesh points (Carasso & Parter, 1970) in the domain R, from which we obtain the linear system

$$\tilde{A}\underline{u} = \tilde{\underline{b}}\ , \tag{3.9}$$

$$\tilde{A} = \begin{bmatrix} \tilde{D}_1 & \tilde{F}_1 & & & & \mathbf{0} \\ \tilde{E}_1 & \tilde{D}_2 & \tilde{F}_2 & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & & \tilde{D}_{m-2} & \tilde{F}_{m-2} \\ \mathbf{0} & & & & \tilde{E}_{m-2} & \tilde{D}_{m-1} \end{bmatrix}_{(m-1)\times(m-1)} = \tilde{D} + \tilde{E} + \tilde{F}\ , \tag{3.10}$$

where now,

$$\tilde{D}_j = \begin{bmatrix} 4r & -2r & & & \\ -2r & 4r & -2r & & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & & & 4r & -2r \\ & & & -2r & 4r \end{bmatrix}_{(n-1)\times(n-1)}, \tilde{E}_j = -\tilde{F}_j = -I_{(n-1)\times(n-1)}, \quad j=1,2,..,(m-1) \qquad (3.11)$$

with similar modifications to (3.7) and (3.8) for f non-linear.

4. Iterative Methods of Solution

Point iterative methods of solution for the finite difference equations (3.3) can easily be given. Thus, assuming a consistent ordering of the points (i,j) the S.O.R. method is

$$u_{i,j}^{(\ell+1)} = u_{i,j}^{(\ell)} + \frac{\omega}{4r}[2ru_{i-1,j}^{(\ell+1)} + u_{i,j-1}^{(\ell+1)} + 2ru_{i+1,j}^{(\ell)} - u_{i,j+1}^{(\ell)} -4ru_{i,j}^{(\ell)} - 2kf], \quad i=1,2,...,n-1, \ j=1,2,....,m-1. \quad (4.1)$$

The eigenvalue spectrum of the point Jacobi method is given by

$$\mu_{p,q} = \{4r\cos(p\pi/n)+2i\cos(q\pi/m)\}/4r, \quad p=1(1)n-1, q=1(1)m-1,$$

and thus the spectral radius is

$$\rho = \max_{p,q} |\mu_{p,q}| = \{4r^2\cos^2(\pi/n)+\cos^2(\pi/m)\}^{\frac{1}{2}}/2r.$$

Similarly, from the block tridiagonal structure of A it is obvious that we can consider block iterative schemes for the solution of (3.4) in which each block of unknowns consist of all the points $u_{i,j}$ in a column of the grid, i.e.,

$(u_{i,1},u_{i,2},u_{i,3},\ldots,u_{i,m-1})$ for each i in the range $1 \le i \le n-1$.

Then, the block Jacobi method is given by

$$D\underline{u}^{(\ell+1)} = -(E+F)\underline{u}^{(\ell)} + \underline{b}, \qquad (4.2)$$

and the block successive overrelaxation method by

$$(D+\omega E)\underline{u}^{(\ell+1)} = -[\omega F+(\omega-1)D]\underline{u}^{(\ell)}+\omega b, \qquad (4.3)$$

or $\quad u^{(\ell+1)}=\mathscr{L}_\omega u^{(\ell)}+\tilde{b},\ \mathscr{L}_\omega\equiv(D+\omega E)^{-1}\{\omega F+(\omega-1)D\},\ \tilde{b}=\omega(D+\omega E)^{-1}b,$

where again the superscript ℓ denotes the iteration cycle and ω the block overrelaxation parameter.

Now the block Jacobi method is convergent if the spectral radius of the iteration matrix $\{-D^{-1}(E+F)\}$ satisfies

$$\rho\{-D^{-1}(E+F)\} < 1 .$$

The iterative method of the column block Jacobi method has an eigenvalue spectrum given by

$$\mu_{p,q} = 2r\cos(p\pi/n)/\{2r-i\cos(q\pi/m)\},\ p=1(1)n-1,q=1(1)m-1 \qquad (4.4)$$

and spectral radius

$$\rho = 2r\cos(\pi/n)/\{4r^2+\cos^2(q'\pi/m)\}^{\frac{1}{2}} < 1, \qquad (4.5)$$

where $q' = \left[\frac{m}{2}\right]$ is the integral part of m/2.

Then the column block Jacobi iteration method is convergent for all values of $r>0$.

The iterative matrix of the row block Jacobi method has a purely imaginary eigenvalue spectrum, namely

$$\mu_{p,q} = i\cos(q\pi/m)/[2r\{1-\cos(p\pi/n)\}]$$

so that

$$\rho = \cos(\pi/m)/[2r\{1+\cos(\pi/n)\}]. \qquad (4.6)$$

From the S.O.R. theory (Young, 1971) we have that since A has block property $A(A^{\pi})$ and is consistently ordered with $\rho\{D^{-1}(E+F)\}$ complex then the S.O.R. method is convergent if for some positive C, ω satisfies

$$0 < \omega < 2/(1+C) \tag{4.7}$$

and for each eigenvalue $(\gamma_k+i\delta_k)$ of $\{D^{-1}(E+F)\}$, the following relationship is satisfied, i.e.

$$\gamma_k^2 + \delta_k^2/C^2 < 1. \tag{4.8}$$

We now seek a choice of the optimum overrelaxation parameter ω_b so that $\rho(\mathscr{L}_\omega)$ is minimised. Thus, we wish to calculate

$$\rho(\mathscr{L}_{\omega_b}) = \min_{\omega} \max_{\mu\varepsilon E} \left|\frac{\omega\mu}{2} \pm \sqrt{\frac{\omega^2\mu^2}{4} - (\omega-1)}\right|^2 \tag{4.9}$$

where E is an elliptical region containing all the eigenvalues $\mu=\gamma+i\delta$ of $(-D^{-1}(E+F))$. It has been shown (Young, 1971) that the optimal ω is given thus

$$\omega_b = 2/[1+\sqrt{\{1-(\gamma^2-\delta^2)\}}], \tag{4.10}$$

with
$$\rho(\mathscr{L}_{\omega_b}) = [(\gamma+\delta)/(1+\sqrt{1-(\gamma^2-\delta^2)})]^2 \tag{4.11}$$
$$= (\gamma+\delta)(\omega_b-1)/(\gamma-\delta) .$$

Finally, since $||(\mathscr{L}_{\omega_b})^n||$ where $||\ \ ||$ denotes the spectral norm, behaves as

$$\binom{n}{p-1}[\rho\{\mathscr{L}_{\omega_b}\}]^{n-p+1}$$

where $\binom{n}{p-1}$ is the binomial coefficient and p is the order of the largest diagonal block of the Jordon canonical form of $\{\mathscr{L}_{\omega_b}\}$.

In particular, if all the eigenvalues are distinct then p=1 and we obtain an approximate relation for the number of iterations required to satisfy the convergence criteria ε in the form

$$[\rho\{\mathscr{L}_{\omega_b}\}]^n \leq \varepsilon,$$

i.e. which gives for n, $n \simeq \log(\varepsilon)/\log[\rho\{\mathscr{L}_{\omega_b}\}]$. (4.12)

It is the purpose of the following numerical example to test the validity of this theory for the chosen reformulated boundary value problem.

5. Fast Algorithmic Solution of the Block Systems

Obviously the success of any block scheme depends solely on whether efficient algorithms for solving each block of unknowns exists and whether it is competitive with existing schemes.

The successive column line overrelaxation method (4.2) requires the direct solution of matrix equations of the form

$$D\underline{u}^{(\ell+1)} = \underline{b} \quad , \qquad (5.1a)$$

for each column of the rectangular domain $\tilde{R}$, where $\underline{b}$ is a sub-vector of known elements and each sub-matrix D is as given in (3.6). Similar systems will be required to be solved for the successive row line over-relaxation method (3.9).

Suitable efficient algorithms based on the Gaussian elimination and triangular factorisation methods can be derived where we note that because of the diagonal dominance property of the blocks no interchange of rows in the elimination process is necessary or required to maintain numerical stability in the solution process.

For the special symmetric form of $\tilde{D}_j$ in (3.9), a fast algorithmic solution procedure has already been given by Evans (1972). Here because of the special form of D_i, i.e. tridiagonal and Toeplitz it seems pertinent to investigate whether a similar fast algorithmic solution process can be derived. (Evans, 1980).

The linear systems (5.1a) can be rewritten in the equivalent form

$$H\underline{u} = \underline{g} \tag{5.1b}$$

where

$$H = \begin{bmatrix} 1-\alpha^2 & \alpha & & & \mathbf{0} \\ -\alpha & 1-\alpha^2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \alpha \\ \mathbf{0} & & & -\alpha & 1-\alpha^2 \end{bmatrix}_{(m-1)\times(m-1)}$$

and we have used the simple transformations,

$$\alpha = 1/2(r+\sqrt{4r^2+1}), \tag{5.2}$$

and

$$g_q = (1-\alpha^2)b_q/4r \ , \ q=1,2,\ldots(m-1) .$$

Further, we can obtain a direct matrix factorisation of H such that H=PQ where P and Q are easily invertible matrices of the form,

$$P = \left[\begin{array}{ccccc|c} 1 & \alpha & & & & \\ & 1 & \alpha & & \mathbf{0} & \\ & & \ddots & \ddots & & \\ & \mathbf{0} & & \ddots & \alpha & \\ & & & & 1 & \alpha \end{array}\right]_{m\times(m-1)} \text{ and } Q = \left[\begin{array}{ccccc} 1 & & & & \\ -\alpha & 1 & & \mathbf{0} & \\ & \ddots & \ddots & & \\ \mathbf{0} & & \ddots & \ddots & \\ & & & -\alpha & 1 \\ \hline & & & & -\alpha \end{array}\right]_{(m-1)\times m} \tag{5.3}$$

Thus, the system (5.1) is solved by rewriting it as the alternative coupled system,

$$P\underline{v} = \underline{g} \quad , \tag{5.4}$$

and

$$Q\underline{u} = \underline{v} \quad , \tag{5.5}$$

where $\underline{v}$ (q=1,2,...,m) is now an intermediate (m×1) vector.

From (5.2) it can be verified immediately that $\alpha<1$ and therefore we can use α as a multiplier in an elimination process without incurring numerical instability from round-off error.

Hence, by applying elimination procedures similar to that used in Evans (1972) to the two coupled linear systems (5.4) and (5.5), it can be shown that the expression to evaluate u_{m-1} is given by

$$u_{m-1}[1-\alpha^2+\alpha^4-\ldots+(-1)^i\alpha^{2i}+\ldots+(-1)^{m-1}\alpha^{2(m-1)}]$$

$$=g_{m-1}[1-\alpha^2+\alpha^4-\ldots+(-1)^i\alpha^{2i}+\ldots+(-1)^{m-2}\alpha^{2(m-2)}]$$

$$+g_{m-2}[1-\alpha^2+\alpha^4-\ldots+(-1)^i\alpha^{2i}+\ldots+(-1)^{m-3}\alpha^{2(m-3)}]+\ldots+\alpha^i g_{m-i-i}$$

$$[1-\alpha^2+\alpha^4-\ldots+(-1)^{m-i-2}\alpha^{2(m-i-2)}]+\ldots\alpha^{m-3}g_2[1-\alpha^2]+\alpha^{m-2}g_1, \tag{5.6}$$

together with the connecting equation,

$$-\alpha u_{m-1} = v_m \; . \tag{5.7}$$

Then, equation (5.6) is simplified, for $|\alpha| < 1$, to give,

$$[1-(-\alpha^2)^m]u_{m-1} = g_{m-1}[1-(-\alpha^2)^{m-1}]+\alpha g_{m-2}[1-(-\alpha^2)^{m-2}]$$

$$+\ldots+\alpha^{m-3}g_2[1-(-\alpha^2)^2]+\alpha^{m-2}g_1[1-(-\alpha^2)] \; ,$$

i.e.,

$$[1-(-\alpha^2)^m]u_{m-1} = g_{m-1}+\alpha g_{m-2}+\ldots+\alpha^{m-3}g_2+\alpha^{m-2}g_1-\alpha^m[-g_1$$

$$+\alpha g_2-\alpha^2 g_3+\ldots+(-1)^i\alpha^{i-1}g_i+\ldots+(-1)^{m-1}\alpha^{m-2}g_{m-1}] \; , \tag{5.8}$$

from which, by using a nesting technique, u_{m-1} can be computed in 2m multiplications and 2m additions approximately.

A back-substitution process, using (5.4) yields the components of the auxiliary vector as,

$$v_i = g_i - \alpha v_{i+1}, \quad i=m-1,m-2,\ldots,1 \qquad (5.9)$$

where

$$v_m = -\alpha u_{m-1} \quad .$$

Finally, the solution vector $\underline{u}$ is given by a forward substitution process obtained from (5.5) to give,

$$u_1 = v_1 \ , \qquad (5.10)$$

$$u_i = v_i + \alpha u_{i-1}, \quad i=2,3,\ldots,m-2.$$

The algorithmic solution (5.8)-(5.10) together with the connecting equation (5.7) requires 5m multiplications, and m additions provided that (5.9) is evaluated efficiently by a nesting technique together with the calculation of predetermined quantities.

A similar rounding error analysis to investigate the cumulative effect of the rounding errors incurred during this algorithmic process can be completed similar to that given in Evans (1972) to verify stability.

6. Numerical Example

In order to test the validity of using the unstable Richardson formula within the specified block methods, we consider the initial boundary problem posed by the linear parabolic partial differential equation

$$u_{xx} - u_t = xu_x \ , \qquad (6.1)$$

in the semi-infinite strip $0 \le x \le 1$, $t \ge 0$, under the initial and boundary conditions

$$u(x,0) = x \quad , \quad 0 \le x \le 1$$

$$u(0,t) = 0 \quad , \quad t \ge 0 \tag{6.2}$$

and $$u(1,t) = e^{-t} \quad , \quad t \ge 0.$$

The steady state solution to (6.1) is given by the differential equation

$$u'' - xu' = 0, \tag{6.3}$$

whilst the end points of the range (0,1) given by (6.2)for $t=\infty$ (T) are

$$u(0) = u(1) = 0, \tag{6.4}$$

from which the analytical solution at $t=T$, $0 \le x \le 1$ can be verified to be $u=0$.

Now choosing the grid size $h=0.1$; $k=0.1$, which gives the value of the parameter $r=10$ for the chosen mesh we assume that the steady state solution is achieved at $T = 10$ then the difference equations become,

$$u_{i,j+1}-r(2+hx_i)u_{i-1,j}+4ru_{i,j}-r(2-hx_i)u_{i+1,j}-u_{i,j-1}=0, \tag{6.5}$$
$$i=1,2,\ldots,9;\ j=1,2,\ldots,99.$$

We can test the convergence of the point and block successive overrelaxation methods and the theory for the optimal parameter and convergence rates (number of itera - tions) as discussed briefly in Section 4 when the iteration matrix has complex eigenvalues.

The above-mentioned investigation results in a large system of linear equations of order $N=891$ which were sol - ved on the computer at Loughborough University for values of ω in the range (0,2). The numerical solution agreed with the exact solution $u=xe^{-t}$ to within a prescribed degree of accuracy, $\epsilon=5\times10^{-6}$and the results are given in the accompanying table.

Method	Theoretical Results		Experimental Results	
	optimal ω_b (4.10)	no. of iterations (4.12)	optimal over-re-laxation parame - ter ω_b	minimum no. of iterations
Point S.O.R.	1.523	23	1.5	23
Col.block S.O.R.	1.515	24	1.5	24
Row block S.O.R.	0.823	7	0.78	7

References

1 GREENSPAN, D., (1974), Discrete Numerical Methods in Physics and Engineering, Academic Press.

2 KELLER, H.B., (1968), Numerical Methods for Two Point Boundary Value Problems, Blaisdell Pub. Co.

3 FORSYTHE, G.E. & WASOW, W.R., (1960), Finite Diffe - rence Methods for Partial Differential Equations J. Wiley.

4 YOUNG, D.M., (1971), Iterative Solution of Large Li-near Systems, Academic Press.

5 CARASSO, A., & S.V. PARTER, (1970), An Analysis of Boundary Value Techniques for Parabolic Problems, Math. Comp. 23, 315-340.

6 EVANS, D.J., (1972), An Algorithm for the Solution of Certain Tridiagonal Systems of Linear Equations, Comp. J., 15, 356-359