

TR/17/84

November 1984

Extrapolation methods for
diffusion-convection equations.

E.H. Twizell

SUMMARY

A fully implicit method and a Crank-Nicolson type method are considered for the solution of the constant coefficient diffusion—convection equation.

The former method is seen to be first order accurate in time and to be L-stable or L₀-stable depending on the convection parameter, while the latter is seen to be second order accurate in time and to be A-stable A₀—stable with oscillations in the solution.

The accuracy in time of the fully implicit part of the local truncation error of the extrapolation method is shown to be slightly inferior to the Crank-Nicolson type method but the extrapolation method is shown to be L-stable or L₀-stable with only small oscillations evident in the solution. The computational requirements of the two second order methods are seen to be similar.



Z1485241

(Rec)
CH
1
R...

1. THE SPACE DISCRETIZATION AND A RECURRENCE RELATION

The linear form of the one dimensional diffusion-convection equation is given by

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} - \mu \frac{\partial u}{\partial x}, \quad 0 < x < X, t > 0 \quad (1)$$

with initial conditions

$$u(x,0) = g(x), \quad 0 \leq x \leq X \quad (2)$$

and boundary conditions

$$u(0,t) = v(t), \quad t > 0 \quad (3)$$

$$\frac{\partial u}{\partial x}(X,t) = 0, \quad t > 0. \quad (4)$$

In obtaining a numerical solution the interval $0 \leq x \leq X$ will be divided into N subintervals each of width $h > 0$, so that $Nh = X$, and the time variable t will be discretized in steps of length ℓ . The open rectangular region $R = [0 < x < X] \times [t > 0]$ in the (x,t) plane, together with its boundary ∂R consisting of the lines $t=0$, $x=0$, $x=X$ are thus covered by a rectangular mesh, the mesh points having coordinates $(mh, n\ell)$ with $m = 0, 1, \dots, N$ and $n = 0, 1, 2, \dots$.

The space derivatives in (1) and (4) are replaced by the central difference approximants

$$\frac{\partial^2 u}{\partial x^2} = h^{-2} \{u(x-h, t) - 2u(x, t) + u(x+h, t)\} + O(h^2) \quad (5)$$

and

$$\frac{\partial u}{\partial x} = \frac{1}{2} h^{-1} \{-u(x-h, t) + u(x+h, t)\} + O(h^2). \quad (6)$$

Equation (1) with (3),(4),(5),(6), is now applied to the mesh points $(mh, n\ell)$, $m=1, 2, \dots, N$ at time level $t = n\ell$ ($n = 1, 2, \dots$). This produces a system of first order ordinary differential equations of the form

$$\frac{d \tilde{U}}{dt} = A \tilde{U}(t) + \tilde{b}_t \quad (7)$$

(Smith⁷) and that $\underline{U}(t)$ satisfies the recurrence relation

$$\underline{U}(t + 1) = -A^{-1} \underline{b}_t + \exp(\ell A)(\underline{U}(t) + A^{-1} \underline{b}_t) . \quad (15)$$

To obtain an explicit solution to (1) with (2), (3), (4), the (0,1) Padé approximant may be used to replace the matrix exponential function in (15). This leads to

$$\underline{U}(t + \ell) = (I + \ell A) \underline{U}(t) + \ell \underline{b}_t \quad (16)$$

which, when applied to the mesh point $(mh, n\ell)$, gives

$$U_m^{n+1} = (P + \frac{1}{2} \mu r) U_{m-1}^n + (1 - 2p) U_m^n + (P - \frac{1}{2} \mu r) U_{m+1}^n \quad (17)$$

for $m \neq N$, with $p = \ell / h^2$ and $r = \ell / h$, For $m = N$, $U_{N+1}^n = U_{N-1}^n + O(h^3)$ from

(4) and (6), so that (17) takes the simplified form

$$U_N^{n+1} = 2P_{N-1} U_N^n + (1 - 2p) U_N^n. \quad (18)$$

This explicit method was developed and analyzed in Siemieniuch and Gladwell⁵ who used the matrix method in their stability analysis. Morton⁷ found that the stability interval given by Siemieniuch and Gladwell⁵ was too large and, using the Fourier method of analysis, showed that stability requires p to lie in the interval $0 \leq p \leq .\frac{1}{2}$

This restriction places a constraint on ℓ and with the ready availability of economic software for sparse solvers, the engineer can profitably turn to implicit methods to obtain a solution; these will be analyzed in the following sections.

The principal part of the local truncation error of the finite difference scheme obtained by replacing $\exp(\ell A)$ in (15) by its (M,K) Padé approximant has the form

$$\left(\frac{1}{6} \mu h^2 \frac{\partial^3 u}{\partial x^3} + C_q \ell^{q-1} \frac{\partial^q u}{\partial t^q} \right)_m^n \quad (19)$$

at the mesh point $(mh, n\ell)$. The error constant C_q in (19) depends only on the Padé approximant being used and $q = M + K + 1$. For the explicit method (17), $q = 2$ and $C_2 = \frac{1}{2}$

2. A FULLY IMPLICIT SCHEME

In contrast to the (0,1) Padé approximant, the (1,0) Padé approximant may be used in (16). This gives

$$(I - \ell A) \underline{U}(t + \ell) - \ell \underline{b}_{t+\ell} = \underline{U}(t) \quad (20)$$

where $\underline{b}_{t+\ell} = h^{-2}[(1 + \frac{1}{2}\mu h)v_{t+\ell}, 0, \dots, 0]^T$ $v_{t+\ell}$ being the numerical value of $v(t + \ell)$.

Applying (20) to the mesh point $(mh, n\ell)$ gives the fully implicit scheme

$$-(p + \frac{1}{2}\mu r)U_{m-1}^{n+1} + (1 + 2p)U_m^{n+1} - (p - \frac{1}{2}\mu r)U_{m+1}^{n+1} = U_m^n \quad (21)$$

for $m \neq N$ and

$$-2pU_{N-1}^{n+1} + (1 + 2p)U_N^n$$

for $m = N$. The principal part of the local truncation error has $q = 2$ in

(19) with $C_2 = -\frac{1}{2}$ so that the method is first order accurate in time.

To properly investigate the stability of (20) the three cases $\mu h > 2$, $\mu h = 2$ and $0 < \mu h < 2$ must be examined separately. For $\mu h > 2$, the *amplification factor* $R_{1,0}(z)$, where $z = -\ell\lambda$ with $\lambda = -2h^{-2} + i\gamma$ some eigenvalue of A , is given by

$$R_{1,0}(z) = \frac{1}{1+z} = \frac{1 + 2p + i\ell\gamma}{(1 + 2p) + \ell^2\gamma^2}.$$

It is easy to see that

$$\left| R_{1,0} \right| = [(1 + 2p)^2 + \ell^2\gamma^2]^{-\frac{1}{2}} < 1$$

and that $|R_{1,1}| \rightarrow 0$ as $\text{Re}(z) = -\text{Re}(-\ell\lambda) = 2p \rightarrow \infty$, and (20) is L-stable for

$\mu h > 2$. (Note: the amplification factor is often referred to as the *amplification symbol* or *symbol*. This terminology was used by Gourlay and Morris¹ in their analyses of methods for the solution of the diffusion equation $\partial u / \partial t = \partial^2 u / \partial x^2$.)

Turning next to the case $\mu h = 2$, all eigenvalues of A have the value

$-2h^{-2}$ so that $z = 2p$ and

$$R_{1,0}(z) = (1+z)^{-1} = (1+2P)^{-1}.$$

It is easy to see that $|R_{1,0}| < 1$ and that $R_{1,0} \rightarrow 0$ as $z = \ell h = 2p \rightarrow \infty$.

The scheme is thus L_0 -stable for $\mu h = 2$.

Finally, for $0 < \mu h < 2$, all eigenvalues of A are real and negative with bounds given by (13), and $z = \ell \lambda$ satisfies

$$0 < 2p - 2p\left(1 - \frac{1}{4}\mu^2 h^2\right)^{\frac{1}{2}} \leq z \leq 2p + 2p\left(1 - \frac{1}{4}\mu^2 h^2\right)^{\frac{1}{2}}.$$

(22)

Clearly $z \rightarrow \infty$ only as $p \rightarrow \infty$ since $0 < \mu h < 2$. The symbol $R_{1,0}(z)$ is now bounded by

$$\frac{1}{1 + 2p\left[1 + \left(1 - \frac{1}{4}\mu^2 h^2\right)^{\frac{1}{2}}\right]} \leq \frac{1}{1 + 2p\left[1 - \left(1 - \frac{1}{4}\mu^2 h^2\right)^{\frac{1}{2}}\right]} \quad (23)$$

from which it follows that $|R_{1,0}| < 1$ and that $R_{1,0} \rightarrow 0$ as $p \rightarrow \infty$, and hence the scheme is L_0 -stable for $0 < \mu h < 2$.

Overall, the fully implicit scheme is L -stable or L_0 -stable whatever the value of μh and much larger time steps may be used than with the explicit method discussed in § 1.

3. A CRANK-NICOLSON TYPE METHOD

The use of the (1,1) Padé approximant in the recurrence relation (15) leads to

$$\left(I - \frac{1}{2}\ell A\right) \tilde{U}(t + \ell) - \frac{1}{2}\ell \tilde{b}_{t+\ell} = \left(I + \frac{1}{2}\ell A\right) \tilde{U}(t) + \frac{1}{2}\ell \tilde{b}_t \quad (24)$$

which, when applied to the mesh point $(mh, n\ell)$, gives

$$\begin{aligned} -\frac{1}{2}\left(p + \frac{1}{2}\mu r\right)U_{m-1}^{n+1} + (1+p)U_m^{n+1} - \frac{1}{2}\left(p - \frac{1}{2}\mu r\right)U_{m+1}^{n+1} \\ = \frac{1}{2}\left(p + \frac{1}{2}\mu r\right)U_{m-1}^n + (1-p)U_m^n + \frac{1}{2}\left(p - \frac{1}{2}\mu r\right)U_{m+1}^n \end{aligned} \quad (25)$$

for $m \neq N$ and

$$-pU_{N-1}^{n+1} + (1+p)U_N^{n+1} = pU_{N-1}^n + (1-p)U_N^n \quad (26)$$

when $m = N$.

The principal part of the local truncation error is given by (19)

with $q=3$ and $C_3 = -\frac{2}{12}$ so that the method is second order accurate in

time, an improvement on the fully implicit method of §2.

To investigate the stability of the scheme, the three cases $\mu h > 2$, $\mu h = 2$ and $0 < \mu h < 2$ must again be examined separately. For $\mu h > 2$ the amplification factor $R_{1,1}(z)$, where $z = -\ell\lambda$ with λ some eigenvalue of A , is given

by

$$R_{1,1} = \frac{1 - \frac{1}{2}z}{1 + \frac{1}{2}z} = \frac{1 - p^2 - \frac{1}{4}\ell^2\gamma^2 + i\ell\gamma}{(1 + p)^2 + \frac{1}{4}\ell^2\gamma^2}. \quad (27)$$

from which it follows that

$$|R_{1,1}| = \frac{[(1 - p^2 - \frac{1}{4}\ell^2\gamma^2)^2 + \ell^2\gamma^2]^{\frac{1}{2}}}{(1 + p)^2 + \frac{1}{4}\ell^2\gamma^2}.$$

An elementary calculation shows that $|R_{1,1}| \leq 1$ for $p \geq 0$ and that $|R_{1,1}| \rightarrow 1$ as $\text{Re}(z) = \text{Re}(-\ell\lambda) = 2p \rightarrow \infty$ so that (24) is A-stable for $\mu h > 2$.

For $\mu h = 2$, all eigenvalues of A have the value $-2h^2$ so that $z = 2p$ and

$$R_{1,1}(z) = \frac{1 - \frac{1}{2}z}{1 + \frac{1}{2}z} = \frac{1 - p}{1 + p} - \frac{p^{-1} - 1}{p^{-1} + 1}. \quad (28)$$

Clearly $|R_{1,1}| \leq 1$ for $p \geq 0$ and $R_{1,1} \rightarrow -1$ as $p \rightarrow \infty$ so that (24) is A_0 -stable for $\mu h = 2$.

Turning lastly to the case $0 < \mu h < 2$, all the eigenvalues λ_j ($j = 1, 2, \dots, N$) are real and negative with bounds given by (13) and $z = -\ell\lambda$, where λ is any eigenvalue of A , satisfies (22); also $z \rightarrow \infty$ only as $p \rightarrow \infty$ since $0 < \mu h < 2$. The amplification symbol $R_{1,1}$ thus has the

bounds $R_{1,1}^{(1)}$ and $R_{1,1}^{(2)}$ given by

$$R_{1,1}^{(1)} = \frac{1 - p[1 - (1 - \frac{1}{4}\mu^2 h^2)^{\frac{1}{2}}]}{1 + p[1 - (1 - \frac{1}{4}\mu^2 h^2)^{\frac{1}{2}}]} = \frac{p^{-1} - [1 - (1 - \frac{1}{4}\mu^2 h^2)^{\frac{1}{2}}]}{p^{-1} + [1 - (1 - \frac{1}{4}\mu^2 h^2)^{\frac{1}{2}}]} \quad (29)$$

and

$$R_{1,1}^{(2)} = \frac{1 - p[1 + (1 - \frac{1}{4}\mu^2 h^2)^{\frac{1}{2}}]}{1 + p[1 - \frac{1}{4}\mu^2 h^2)^{\frac{1}{2}}]} = \frac{p^{-1} - [1 + (1 - \frac{1}{4}\mu^2 h^2)^{\frac{1}{2}}]}{p^{-1} + [1 + (1 - \frac{1}{4}\mu^2 h^2)^{\frac{1}{2}}]} \quad (30)$$

It follows from (29) and (30) that $|\mathcal{R}_{1,1}^{(1)}| \leq 1$ and $|\mathcal{R}_{1,1}^{(2)}| \leq 1$ for $p \geq 0$, and that $\mathcal{R}_{1,1}^{(1)} \rightarrow -1$ and $\mathcal{R}_{1,1}^{(2)} \rightarrow$ monotonically as $p \rightarrow \infty$. The method is therefore A_0 -Stable for $0 < \mu h < 2$.

Although more accurate than the fully implicit method (20), the Crank-Nicolson-type method (24) is known to give an oscillatory numerical solution for (1) - (4) (see Smith *et al*⁶). Similar oscillations also arise when the familiar Crank-Nicolson method is used to solve the diffusion equation

$\partial u / \partial t = \partial^2 u / \partial x^2$ when there are discontinuities between initial conditions and boundary conditions, Lawson and Morris³ found that such oscillatory components in the Crank-Nicolson solution of the diffusion equation are made to decay faster than the non-oscillatory components if some restriction is placed on r (a conventional stability analysis places no restriction on p), thus lessening their effect.

Following Lawson and Morris³, criteria must be found which ensure that the highest frequency component of the solution is damped to zero faster than the lowest frequency component. For $\mu h > 2$ this is equivalent to specifying that the magnitude of the amplification symbol relating to the maximum value of γ_j^2 (see (27) and (11)) does not exceed that relating to the minimum value.

This leads to

$$\left| \frac{1 - p^2 - p^2 \left(\frac{1}{4} \mu^2 h^2 - 1 \right)}{(1 + p)^2 + p^2 \left(\frac{1}{4} \mu^2 h^2 - 1 \right)} \right| < \left| \frac{1 - p^2}{(1 + p)^2} \right|$$

which is satisfied provided

$$-2p + p^3 + \frac{1}{4} \mu^2 p r^2 (1 + p) < 1. \quad (31)$$

The case $\mu h = 2$ is trivial but the case $0 < \mu h < 2$ requires

$$\left| \frac{1 - p \left[1 + \left(1 - \frac{1}{2} \mu^2 h^2 \right)^{\frac{1}{2}} \right]}{1 + p \left[1 + \left(1 - \frac{1}{4} \mu^2 h^2 \right)^{\frac{1}{2}} \right]} \right| < \left| \frac{1 - p \left[1 - \left(1 - \frac{1}{4} \mu^2 h^2 \right)^{\frac{1}{2}} \right]}{1 + p \left[1 - \left(1 - \frac{1}{4} \mu^2 h^2 \right)^{\frac{1}{2}} \right]} \right|$$

which is satisfied provided

$$\mu r < 2. \quad (32)$$

4. EXTRAPOLATION OF THE FULLY IMPLICIT SCHEME

In view of the restrictions placed on the ratios $p = \ell/h^2$ and $r = \ell/h$ in (31) and (32), alternative methods of achieving higher accuracy in time must be investigated. The L-stable properties of the fully implicit method (20) indicate that this method is suitable for extrapolation.

The extrapolation is carried out by computing $\underline{U}(t + 2\ell)$ from $\underline{U}(t)$ using (a) two single time steps of length ℓ , and (b) a double time step of length 2ℓ . The values of $\underline{U}(t + 2\ell)$ so obtained are denoted by $\underline{U}^{(1)}(t + 2\ell)$ and $\underline{U}^{(2)}(t + 2\ell)$ and it may be shown that $\underline{U}^{(E)}(t + 2\ell)$ defined by

$$\underline{U}^{(E)}(t + 2\ell) = 2 \underline{U}^{(1)}(t + 2\ell) - \underline{U}^{(2)}(t + 2\ell) \quad (33)$$

is second order accurate in time. In (19), therefore, $q=3$ and it may be shown that $C_3 = \frac{4}{3}$

The extrapolated form of the fully implicit method is thus of the same order of accuracy as the Crank-Nicolson type method of §3 though its error constant is inferior. Its major benefit to the user, however, is that it is still L-stable. Using S instead of R to distinguish the amplification factor or symbol of the extrapolated method (33) from that of its original form (20), it may be shown that, for $\mu h > 2$

$$\begin{aligned} |s_{1,0}(z)|^2 &= \left| 2 \left(\frac{1}{1+z} \right)^2 - \left(\frac{1}{1+2z} \right) \right|^2 \\ &= \left[\frac{2\{1+2p\}^2 - \ell^2\gamma^2}{\{(1+2p)^2 + \ell^2\gamma^2\}^2} - \frac{1+4p}{(1+4p)^2 + 4\ell^2\gamma^2} \right]^2 \\ &\quad + 4 \left[\frac{2\ell\gamma(1+2p)}{\{(1+2p)^2 + \ell^2\gamma^2\}^2} - \frac{\ell\gamma}{(1+4p)^2 + 4\ell^2\gamma^2} \right]^2. \end{aligned} \quad (34)$$

It may be seen from (34) and (11) that $|S_{1,10}| \leq 1$ for $p \geq 0$; in addition it is easy to show from (34) that $|S_{1,10}| \rightarrow 0$ as $\text{Re}(z) = \text{Re}(-\ell) = 2p \rightarrow \infty$. The extrapolated form of the fully implicit method is therefore L-stable for $\mu h > 2$. It must be noted, however, that the real part of $S_{1,10}$ is negative

for values of $p < 1.3$ and $\ell^2 \lambda_2 < 0.7 + 0.2p + 0.8p^2$ (approximately) so that small oscillations will be present for these ranges.

Turning next to the case when $\mu h = 2$, the amplification symbol of the extrapolation method takes the form

$$S_{1,0} = 2 \left(\frac{1}{1 + 2p} \right)^2 - \frac{1}{1 + 4p}. \quad (35)$$

Clearly $S_{1,0} \leq 1$ for all $p \geq 0$ and $S_{1,0} \rightarrow 0$ as $z = 2p \rightarrow \infty$ so that the extrapolation method (33) is L° -stable for $\mu h = 2$. The amplification symbol given by (35) is negative for $p > \frac{1}{2}(1 + \sqrt{2})$ so that small oscillations are evident for $p > \frac{1}{2}(1 + \sqrt{2}) \simeq 1.21$.

In the case $0 < \mu h < 2$, the amplification symbol $S_{1,0}$ is bounded by

$$\begin{aligned} & \frac{1}{2p^2 \left[1 + \left(1 - \frac{1}{4} \mu^2 h^2 \right) \frac{1}{2} \right]} - \frac{1}{4p \left[1 + \left(1 - \frac{1}{4} \mu^2 h^2 \right) \frac{1}{2} \right]} \\ & \leq S_{1,0} \leq \frac{1}{2p^2 \left(1 - \left(1 - \frac{1}{4} \mu^2 h^2 \right) \frac{1}{2} \right)} - \frac{1}{4p \left[1 - \left(1 - \frac{1}{4} \mu^2 h^2 \right) \frac{1}{2} \right]} \end{aligned} \quad (36)$$

so that $|S_{1,0}| < 1$ for $p > 0$ and $S_{1,0} \rightarrow 0$ as $z = 2p \rightarrow \infty$. The extrapolation of (20) is therefore L° -stable for $0 < \mu h < 2$, though for μh in this interval the solution oscillates for $p > 2$.

5. CONCLUSIONS AND COMPUTATIONAL ASPECTS

It has been seen that the third order accurate method (33), the extrapolated form of the fully implicit method (20), is L -stable. The method does give rise to small oscillations but these are damped to zero because of the L -stability property of the method.

This was not found to be so for the Crank-Nicolson type method which was seen to be A -stable. The order of accuracy of the Crank-Nicolson type method was seen to be the same as that of the extrapolation method. Indeed, the principal part of the local truncated error was seen to be slightly superior to that of the extrapolation method, but the L -stability property of the latter method makes it a better choice.

This preference for the extrapolation method is strengthened when the operations counts for the two methods are compared. In integrating from time t to time $t+2\ell$ using a stepsize of length ℓ , both methods need a total of $16N-12$ multiplications/divisions; the extrapolation method needs $13N-9$ additions/subtractions while the Crank-Nicolson method needs $12N-6$ additions/subtractions. The extrapolation method also requires the storage of one additional vector at length ℓ . Clearly, the additional costs are small.

Overall, the extrapolation method is to be preferred to the Crank-Nicolson type method in view of the L-stability properties of the former. The local truncation error of the fully implicit method, on which the extrapolation method is based, can be improved still further to $O(h^2 + \ell^3)$ with $q=4$ and

$$C_4 = -\frac{45}{8} \text{ in (19), by adapting the best of the combination methods of Gourlay}$$

and Morris¹. This third order method (in time) can also be shown to be L-stable for all values of $\mu h > 0$. The Crank-Nicolson type method (24) can also be extrapolated; this gives fourth order accuracy in time with $q=5$

$$\text{and } C_4 = \frac{1}{10} \text{ in (19), but this method is not even A-stable (see Khaliq}^2) \text{ and}$$

is not computationally economic.

REFERENCES

1. A.R. Gourlay and J.L.I. Morris, 'The extrapolation of first order methods for parabolic partial differential equations. I', *SINUM* 12, 641-655 (1980).
2. A.Q.M. Khaliq, *Numerical Methods for Ordinary Differential Equations with Applications to Partial Differential Equations*, Ph.D. thesis, Brunel University, 1983.
3. J.D. Lawson and J.L.I. Morris, 'The extrapolation of first order methods for parabolic partial differential equations, I', *SINUM* 15, 1212-1224 (1978).
4. K.W. Morton, 'Stability of finite difference approximations to a diffusion-convection equation', *Int. J. num. Meth. Engng.* 15, 677-683 (1980).
5. J.L. Siemieniuch and I. Gladwell, 'Analysis of explicit difference methods for a diffusion-convection equation', *Int. J. num. Meth. Engng.* 12, 899-916 (1978).
6. I.M. Smith, J.L. Siemieniuch and I. Gladwell, 'Evaluation of Nørsett methods for integrating differential equations in time', *Int. J. for Numerical and Analytical Methods in Geomechanics* 1, 57-74 (1977).
7. P. Smith, *Numerical Modelling of Human Thermoregulation*, Ph.D. thesis, CNA (Sunderland Polytechnic), 1981.

