

TR/01/83

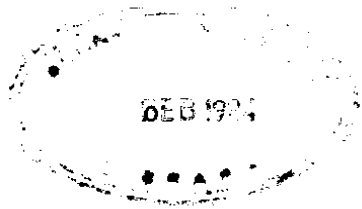
January 1983

Backward difference replacements of the
space derivative in first order
hyperbolic equations.

E. H. Twizell

and

A.Q.M. Khaliq



0. ABSTRACT

Two families of two-time level difference schemes are developed for the numerical solution of first order hyperbolic partial differential equations with one space variable. The space derivative is replaced by (i) a first order, (ii) a second order backward difference approximant and the resulting system of first order ordinary differential equations is solved using A_0 -stable and L_0 -stable methods.

The methods are tested on a number of problems from the literature involving wave-form solutions, increasing solutions with discontinuities in function values or first derivatives across a characteristic, and exponentially decaying solutions.

z147350x

1. INTRODUCTION

In recent years much attention has been devoted in the literature to the extrapolation in time of low-order methods for the numerical solution of first order hyperbolic partial differential equations and for second order parabolic equations.

Essentially the same procedure may be followed for such parabolic equations [11,4] and hyperbolic equations [8]: that is to say, the space derivatives in the differential equations are approximated by a suitable finite difference replacement, and the resulting system of first order ordinary differential equations solved using a stable numerical method. The accuracy in time can then be controlled by a suitable choice of method for solving an ordinary differential equation; improvement in the accuracy in space, on the other hand, requires a different replacement of the space derivative in the partial differential equation.

From the point where the replacement of the space derivative has been chosen, accuracy in time can be varied by a multistage method [4] which involves a spread over three or more time increments, or by a method involving a similar spread over more than three mesh points at a given time level [8,12,15]. The former type of method is, in effect, an application of linear multistep methods for systems of ordinary differential equations, while the latter is an application of multiderivative methods [15].

Both approaches have a weakness which is the other's strength: using a multistage method, seeking the solution at certain fixed times requires the time interval to be divided into two or more subintervals depending on the accuracy required, whereas the integration can be carried out without subdividing the time interval if an A-stable or L-stable multiderivative method is used. On the other hand, implicit multistage methods need only tridiagonal solvers to obtain the solution (five at each time level for

third order accuracy in time and nine for fourth order accuracy [4]) whereas those multiderivative methods based on central difference replacements of the space derivative [8] need only one quindagonal solver.

In the present paper attention will be given only to first order hyperbolic equations. The methods to be discussed are based on backward difference replacements of the space derivatives and can therefore be used explicitly so that here, too, they have an advantage over multistage formulations. The use of backward difference replacements has the advantage that the oscillations which are always present with central difference replacements, do not arise. Also, the difficulties which arise in parabolic equations because of stiffness are not present in solving hyperbolic equations by multiderivative techniques. The methods will use function values at only two time levels as in [8], unlike the methods developed by Olinger [13] where three time levels were used in the formulation.

The families of backward difference methods to be developed, like those of Olinger[13], depend on the theorems of Gustaffson [5] for the establishment of stability. The first methods developed are based on the usual first order replacements of the space derivative and the lower order Padé approximants to the matrix exponential function. Accuracy is then improved by approximating the space derivative at the mesh point adjacent to the boundary, at each time level, by the same low-order replacement, and by the usual second order replacement at all other mesh points. Finally, accuracy is improved further by using higher order Pade approximants to the matrix exponential function. The methods are divided into two classes, the first class using only the low order space replacement, the second using both space replacements. Extrapolation in time is also discussed for L_0 -stable methods.

The methods are tested on five problems involving wave-form solutions,

increasing solutions with discontinuities in function values or first derivatives across a characteristic, and exponentially decaying solution.

2. LOW ORDER APPROXIMATIONS IN SPACE AND TIME

Consider the first order hyperbolic partial differential equation

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = 0 ; \quad x > 0, \quad t > 0, \quad (1)$$

where $a > 0$ is a real constant, with initial conditions

$$u(x, 0) = g(x) ; \quad x \geq 0 \quad (2)$$

and boundary conditions

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

equations (1), (2), (3) form the initial-boundary value or outflow problem.

Suppose that the solution of (1) is sought in some region

$R = [0 < x < X] \times [t > 0]$ of the first quarter plane $x > 0, t > 0$. The interval $0 \leq x \leq X$ is divided into N equal parts each of width h , so that $Nh = X$, and the time variable t is discretized in steps of length ℓ . The open region R and its boundary ∂R , consisting of the axes $t = 0, x = 0$ and the line $x = X$, have thus been covered by a rectangular mesh of points having coordinates $(jh, n\ell)$, where $j = 0, 1, \dots, N$ and $n = 0, 1, 2, \dots$. The theoretical solution of a finite difference scheme approximating the differential equation (1) at the mesh point $(jh, n\ell)$ will be denoted by U_j^n , the theoretical solution of the differential equation at this point being $u_j^n \equiv u(jh, n\ell)$.

$$\underline{U}(t) = C^{-1} \underline{c}_t + \exp(-a t C) \{ \underline{g} - C^{-1} \underline{c}_t \}, \quad (8)$$

Where \underline{g} the vector of initial values. The solution given by (8) satisfies the recurrence relation

$$\underline{U}(t+\ell) = C^{-1} \underline{c}_{t+\ell} + \exp(-a \ell C) \{ \underline{U}(t) - C^{-1} \underline{c}_t \}. \quad (9)$$

Using the (m, k) Padé approximant to the exponential function defined by $R_{m,k}(\theta) = P_k(\theta)/Q_m(\theta) + O(\theta^{m+k+1})$ where $P_k(\theta)$, $Q_m(\theta)$ are polynomials of degrees k , m , respectively, to replace the matrix exponential function in (9) leads to a two-time level finite difference scheme which is unconditionally stable for $m \geq k$ and which may be used explicitly because of the nature of the initial and boundary conditions (2), (3). The principal part of the local truncation error of such a method has the form

$$\left(-\frac{1}{2} a \ell h \frac{\partial^2 \underline{u}}{\partial x^2} + C_q \ell^q \frac{\partial^q \underline{u}}{\partial t^q} \right)_j^n \quad (10)$$

where the constants C_q ($q = m+k+1$) are given in [15] and are reproduced in Table I.

The component of the local truncation error due to the chosen Padé approximant, namely $(C_q \ell^q \partial^q \underline{u} / \partial t^q)_j^n$, can be improved by at least one power of ℓ by extrapolating in time; the other component $\left(-\frac{1}{2} a \ell h \frac{\partial^2 \underline{u}}{\partial x^2} \right)_j^n$, which is related to the space discretization, will not change.

The extrapolating procedure determines $\underline{U}(t+2\ell)$ in terms of $\underline{U}(t)$: it first calculates $\underline{U}^{(1)} = \underline{U}^{(1)}(t+2\ell)$ by writing equation (9), in which the matrix exponential function has been replaced by an appropriate Padé approximant, over two single time steps, and then calculates $\underline{U}^{(2)} = \underline{U}^{(2)}(t+2\ell)$ by writing (9) over a double time step. The extrapolated value $\underline{U}^{(E)} = \underline{U}^{(E)}(t+2\ell)$ is then found from one of the formulas

$$\underline{U}^{(E)} = (2^{m+k} \underline{U}^{(1)} - \underline{U}^{(2)}) / (2^{m+k} - 1) + O(\ell^{m+k+2}) \quad (11)$$

for $m \neq k$, or

$$\underline{U}^{(E)} = (2^{2m} \underline{U}^{(1)} - \underline{U}^{(2)}) / (2^{2m} - 1) + O(\ell^{2m+3}) \quad (12)$$

for $m = k$.

The extrapolating formulas for $m = 1, 2$ and $k = O(1)m$ are contained in Table I. The term $[-\frac{1}{2} a \ell h \partial^2 u / \partial x^2]_j^n$ will still be present in the principal part of the local truncation error of the extrapolated form of each finite difference method. There will also be a term of the form $[E_s \ell^s \partial^s u / \partial t^s]_j^n$ ($s = m+k+2$ for $m \neq k$, $s = 2m+3$ for $m = k$). The constants E_s (Twizell and Khaliq [15]) are also contained in Table I.

Associated with each extrapolated method is the amplification symbol

$$s_{m,k}^{(0)} = A [P_k(\theta) / Q_m(\theta)]^2 - (A-1) P_k(2\theta) / Q_m(2\theta), \quad (13)$$

where $\theta = a \ell \lambda$, λ an eigenvalue of C (actually, the eigenvalues of the matrix C are all equal to $1/h$, but this will not be so in later sections of the paper). In (13), $A = 2^{m+k} / (2^{m+k} - 1)$.

The extrapolated form of a method is A_0 -stable, or stable in the conventional sense of perturbations in the initial conditions not being magnified as $t \rightarrow \infty$, if $|s_{m,k}(\theta)| \leq 1$. The extrapolated form of the five methods given in Table I are therefore unconditionally stable, except the extrapolated form of the method based on the (1,1) Padé approximant which is stable only for $0 < ar \leq 6 + 4\sqrt{3}$ where $r = \ell/h$.

If, in addition to the extrapolated form of a method being

A_0 -stable, its symbol satisfied $\lim_{\theta \rightarrow \infty} s_{m,k}(\theta) = 0$, the method is L_0 -stable.

Of the five methods listed in Table I only those based on the (1,0), (2,0), (2,1) Pade approximants have extrapolated forms which are L_0 -stable, the symbol $S_{2,2}(\theta)$ tending to +1 as $\theta \rightarrow \infty$. Lambert [10] notes that the method resulting from the (m,k) Pade approximant is

- (a) conditionally stable for $m < k$,
- (b) A_0 -stable for $m \geq k$,
- (c) L_0 -stable for $m > k$.

3. HIGHER ORDER SPACE REPLACEMENT

Whereas extrapolation in time does, indeed, bring about some improvement in the principal parts of the local truncation errors of all finite difference schemes resulting from (9), the improvement of any one method may not be sufficient to justify its use for larger values of h . This is because the component of the local truncation error given by $[-\frac{1}{2}a\lambda h\partial^2 u/\partial x^2]_j^n$ is still present and tends to overshadow any improvement brought about by extrapolating in time.

Following Oliger [13], this difficulty is now partially overridden by introducing a second order backward difference approximant to $\partial u/\partial x$ at the mesh points $(jh, n\ell)$ for $j = 2, 3, \dots, N$ and $n = 0, 1, \dots$, whilst retaining the first order approximant (4) at the points $(h, n\ell)$ adjacent to the boundary $x = 0$. This mixture of approximants to $\partial u/\partial x$ is justified in the theorems of Gustafsson [5], so that, provided a Pad approximant is chosen which would lead to unconditional stability if the low order approximant (4) were used at every mesh point, the scheme resulting from the use of the mixture of approximants to $\partial u/\partial x$ will also be unconditionally stable and will have the convergence rate of the more accurate interior approximant (see also Oliger [13]).

The schemes resulting from the use of different backward difference

$$\mathbf{h} \underset{\sim}{\mathbf{d}}_t = [2\mathbf{v}_t, -\mathbf{v}_t, 0, \dots, 0]^T. \quad (18)$$

One eigenvalue of the matrix \mathbf{D} has the value $2/h$ and the other $N - 1$ eigenvalues have the value $3/h$.

The solution of (16) with (2) is

$$\underset{\sim}{\mathbf{U}}(t) = \mathbf{D}^{-1} \underset{\sim}{\mathbf{d}}_t + \exp\left(-\frac{1}{2}at\mathbf{D}\right) \left\{ \underset{\sim}{\mathbf{g}} - \mathbf{D}^{-1} \underset{\sim}{\mathbf{d}}_t \right\}, \quad (19)$$

and it is easy to show that (19) satisfies the recurrence relation

$$\underset{\sim}{\mathbf{U}}(t + \ell) = \mathbf{D}^{-1} \underset{\sim}{\mathbf{d}}_t + \exp\left(-\frac{1}{2}a\ell\mathbf{D}\right) \left\{ \underset{\sim}{\mathbf{U}}(t) - \mathbf{D}^{-1} \underset{\sim}{\mathbf{d}}_t \right\} \quad (20)$$

in which the matrix exponential function is replaced by an appropriate Padé approximant.

Only schemes based on Fade approximants for which $m \geq k$ will be considered. The amplification factors of the extrapolated forms of such schemes may be obtained from (13) with $\theta = \frac{1}{2}a\ell\lambda$, λ now an eigenvalue of \mathbf{D} . The schemes developed from (20) will be two-time level schemes which may be used explicitly because of (2), (3).

Using the (1,0) approximant in (20) gives the LQ-stable scheme

$$\left(\mathbf{I} + \frac{1}{2}a\ell\mathbf{D}\right) \underset{\sim}{\mathbf{U}}(t + \ell) - \frac{1}{2}a\ell \underset{\sim}{\mathbf{d}}_{t+\ell} = \underset{\sim}{\mathbf{U}}(t) \quad (21)$$

which, from Table I, is seen to be first order accurate in time. The principal part of the local truncation error at the mesh point $(h, n\ell)$ is, from (10),

$$\left(-\frac{1}{2}a\ell h \frac{\partial^2 \mathbf{u}}{\partial x^2} - \frac{1}{2}\ell^2 \frac{\partial^2 \mathbf{u}}{\partial t^2}\right)_1^n \quad (22)$$

and at the mesh point $(mh, n\ell)$ is

$$\left(-\frac{1}{3} a \ell h^2 \frac{\partial^3 \mathbf{u}}{\partial x^3} - \frac{1}{2} \ell^2 \frac{\partial^2 \mathbf{u}}{\partial t^2} \right)_j^n \quad (23)$$

for $j = 2, 3, \dots, N$ and $n = 1, 2, \dots$.

In view of its favourable stability properties, it is worthwhile to extrapolate (21) using (11). The extrapolated form can be used explicitly and is L_0 -stable; its local truncation error is

$$\left(-\frac{1}{2} a \ell h^2 \frac{\partial^2 \mathbf{u}}{\partial x^2} + \frac{4}{3} \ell^3 \frac{\partial^3 \mathbf{u}}{\partial t^3} \right)_1^n \quad (24)$$

at the mesh point $(h, n\ell)$ adjacent to the boundary, and

$$\left(-\frac{1}{3} a \ell h^2 \frac{\partial^3 \mathbf{u}}{\partial x^3} + \frac{4}{3} \ell^3 \frac{\partial^3 \mathbf{u}}{\partial t^3} \right)_1^n \quad (25)$$

at the interior mesh points $(jh, n\ell)$ where $j = 2, \dots, N$ and $n = 1, 2, \dots$.

Some improvement, in accuracy may be achieved by using the (1,1) Padé approximant to the matrix exponential function in (20) to give

$$\left(I + \frac{1}{4} a \ell D \right) \underline{U}(t+\ell) - \frac{1}{4} a \ell \underline{d}_{t+\ell} = \left(I - \frac{1}{4} a \ell d \right) \underline{U}(t) + \frac{1}{4} a \ell \underline{d}_t \quad (26)$$

which is second order accurate in time and which is A_0 -stable, the amplification factor tending to -1 as $t \rightarrow \infty$.

The principal part of the local truncation error of (26) at the mesh point $(h, n\ell)$ is

$$\left(-\frac{1}{2} a \ell h \frac{\partial^2 \mathbf{u}}{\partial x^2} - \frac{1}{12} \ell^2 \frac{\partial^3 \mathbf{u}}{\partial t^3} \right)_1^n \quad (27)$$

and at the mesh points $(jh, n\ell)$ away from the boundary is

$$\left(-\frac{1}{3} a \ell h^2 \frac{\partial^3 \mathbf{u}}{\partial x^3} + \frac{1}{6} \ell^3 \frac{\partial^3 \mathbf{u}}{\partial t^3} \right)_j^n, j=4,5,\dots, N \quad (35)$$

which, on extrapolation, becomes

$$\left(-\frac{1}{3} a \ell h^2 \frac{\partial^3 \mathbf{u}}{\partial x^3} - \frac{1}{3} \ell^4 \frac{\partial^4 \mathbf{u}}{\partial t^4} \right)_j^n, j=4,5,\dots, N \quad (36)$$

Expressions (35), (36) show that the loss of accuracy at the mesh points $(h, n\ell)$, $n = 0, 1, \dots$, experienced by the methods based on the lower order Pade approximants, has spread to the mesh points $(2h, n\ell)$, $(3h, n\ell)$. This is not a grave problem, however, for a space discretization involving a large value of N . Furthermore, the constant $C_3 = \frac{1}{6}$ in (35) is greater in modulus than its counterpart in (28) which relates to the A_0 -stable method (26).

These observations indicate that the A_0 -stable method (26) is to be preferred to the L_0 -stable method (29). However, when a central difference replacement to the space derivative in (1) is made (see Khaliq and Twizell [8]), this is not so; neither is it so in the case of second order parabolic equations (Lawson and Morris [11]), for then the equivalent method based on the (1,1) Padé approximant (the Crank-Nicolson method), also requires a restriction on ℓ to ensure the decay of oscillations in U as $t \rightarrow \infty$. Numerical results to support all these observations are given in Khaliq [9].

Turning, next, to the (2,1) Pade approximant, (20) becomes

$$\begin{aligned} \left(I + \frac{1}{3} a \ell D + \frac{1}{24} a^2 \ell^2 D^2 \right) \underline{U}(t + \ell) - \left(\frac{1}{3} a \ell I + \frac{1}{24} a^2 \ell^2 D \right) \underline{d}_{t+\ell} \\ = \left(I - \frac{1}{6} a \ell D \right) \underline{U}(t) + \frac{1}{6} a \ell \underline{d}_t. \end{aligned} \quad (37)$$

Applying (37) to the mesh points $(jh, n\ell)$ requires the solution vector

$\underline{U}(t+\ell)$ to be determined implicitly from a linear system of the form (31).

The matrix E is still of the form (32) but its non-zero elements are now given by

$$\begin{aligned} e_1 &= 1 + \frac{2}{3} ar + \frac{1}{6} a^2 r^2, & e_2 &= -\frac{4}{3} ar - \frac{5}{6} a^2 r^2, & e_3 &= \frac{1}{3} ar + \frac{7}{8} a^2 r^2, \\ e_4 &= 1 + ar + \frac{3}{8} a^2 r^2, & e_5 &= -\frac{4}{3} ar - a^2 r^2, & e_6 &= \frac{1}{3} ar + \frac{11}{12} a^2 r^2, \\ e_7 &= -\frac{1}{3} a^2 r^2, & e_8 &= \frac{1}{24} a^2 r^2, \end{aligned} \quad (38)$$

while the elements of ϕ^n are given by

$$\begin{aligned} \phi_1^n &= (1 - \frac{1}{3} ar)U_1^n + ar(\frac{2}{3} + \frac{1}{6} ar)v_{t+\ell} + \frac{1}{3} ar v_t, \\ \phi_2^n &= \frac{2}{3} arU_1^n + (1 - \frac{1}{2} ar)U_1^n + ar(\frac{1}{3} + \frac{11}{24} ar)v_{t+\ell} + \frac{1}{6} ar v_t, \\ \phi_3^n &= \frac{1}{6} arU_1^n + \frac{2}{3} arU_2^n + (1 - \frac{1}{2} ar)U_3^n + \frac{1}{4} a^2 r^2 v_{t+\ell}, \\ \phi_4^n &= -\frac{1}{6} arU_2^n + \frac{2}{3} arU_3^n + (1 - \frac{1}{2} ar)U_4^n - \frac{1}{24} a^2 r^2 v_{t+\ell}, \\ \phi_4^n &= -\frac{1}{6} arU_{j-2}^n + \frac{2}{3} arU_{j-1}^n + (1 - \frac{1}{2} ar)U_j^n, \quad j = 5, \dots, N. \end{aligned} \quad (39)$$

The vector $\underline{U}(t+\ell)$ is found from (31) using forward substitution.

The finite difference scheme based on the (2,1) Padé approximant is L_0 —stable; the principal part of its local truncation error is

$$\left(-\frac{1}{3} alh^2 \frac{\partial^3 u}{\partial x^3} - \frac{1}{72} \ell^4 \frac{\partial^4 u}{\partial t^4} \right)_j^n, \quad j = 4, \dots, N. \quad (40)$$

which, following extrapolation using (11), becomes

$$\left(-\frac{1}{3} alh^2 \frac{\partial^3 u}{\partial x^3} - \frac{8}{945} \ell^5 \frac{\partial^5 u}{\partial t^5} \right)_j^n, \quad j = 4, \dots, N. \quad (41)$$

Expressions (40), (41) do indicate an improvement on (28) and justify the use of (37) even though the three points near the boundary suffer

greater error at each time step than the remaining $N - 3$ points away from the boundary $x = 0$.

The final method of the family arising from (20) to be considered in this paper, is that obtained by replacing the exponential matrix function with its (2,2) Padé approximant. The recurrence relation becomes

$$\begin{aligned} & \left(I + \frac{1}{4} a \ell D + \frac{1}{48} a^2 \ell^2 D^2 \right) U(t+\ell) - \left(\frac{1}{4} a I + \frac{1}{48} a^2 \ell^2 D \right) d_{\sim t+\ell} \\ & = \left(I - \frac{1}{4} a \ell D + \frac{1}{48} a^2 \ell^2 D^2 \right) U(t) + \left(\frac{1}{4} a I + \frac{1}{48} a^2 \ell^2 D \right) d_{\sim t} \end{aligned} \quad (42)$$

which gives rise to an A -stable method with amplification factor tending to +1 as $t \rightarrow \infty$.

Applying (42) to each mesh point $(jh, n \ell)$, $j = 1, 2, \dots, N$, at time $t = n \ell$, $n = 0, 1, \dots$, leads to the solution vector $U(t+\ell)$ at the advanced time $t = (n+1)\ell$ being determined implicitly from a system of the form (31). The non-zero elements of E are arranged as in (32)

and have the values

$$\begin{aligned} e_1 &= 1 + \frac{1}{2} ar + \frac{1}{12} a^2 r^2, & e_2 &= -ar - \frac{5}{12} a^2 r^2, & e_3 &= \frac{1}{4} ar + \frac{7}{16} a^2 r^2, \\ e_4 &= 1 + \frac{3}{4} ar + \frac{3}{16} a^2 r^2, & e_5 &= -ar - \frac{1}{2} a^2 r^2, & e_6 &= \frac{1}{4} ar + \frac{11}{24} a^2 r^2, \\ e_7 &= -\frac{1}{6} a^2 r^2, & e_8 &= \frac{1}{48} a^2 r^2. \end{aligned} \quad (43)$$

The elements of the vector ϕ^n are

$$\begin{aligned} \phi_1^n &= \left(1 - \frac{1}{2} ar + \frac{1}{12} a^2 r^2 \right) U_1^n + ar \left(\frac{1}{2} + \frac{1}{12} ar \right) v_{t+\ell} + ar \left(\frac{1}{2} - \frac{1}{12} ar \right) v_t, \\ \phi_2^n &= ar \left(1 - \frac{5}{12} ar \right) U_1^n + \left(1 - \frac{3}{4} ar + \frac{3}{16} a^2 r^2 \right) U_2^n + ar \left(\frac{1}{4} + \frac{11}{48} ar \right) v_{t+\ell} + ar \left(\frac{1}{4} - \frac{11}{48} ar \right) v_t, \\ \phi_3^n &= ar \left(-\frac{1}{4} + \frac{7}{16} ar \right) U_1^n + ar \left(1 - \frac{1}{2} ar \right) U_2^n + \left(1 - \frac{3}{4} ar + \frac{3}{16} a^2 r^2 \right) U_3^n + \frac{1}{8} a^2 r^2 v_{t+\ell} - \frac{1}{8} a^2 r^2 v_t, \\ \phi_4^n &= -\frac{1}{6} a^2 r^2 U_1^n + ar \left(-\frac{1}{4} + \frac{11}{24} ar \right) U_2^n + ar \left(1 - \frac{1}{2} ar \right) U_3^n + \left(1 - \frac{3}{4} ar + \frac{3}{16} a^2 r^2 \right) U_4^n \\ & \quad - \frac{1}{40} a^2 r^2 v_{t+\ell} + \frac{1}{48} a^2 r^2 v_t, \end{aligned}$$

$$\begin{aligned} \phi_j^n = & \frac{1}{48} a^2 r^2 U_{j-4}^n - \frac{1}{6} a^2 r^2 U_{j-3}^n + ar \left(-\frac{1}{4} + \frac{11}{48} ar \right) U_{j-2}^n + ar \left(1 - \frac{1}{2} ar \right) U_{j-1}^n \\ & + \left(1 - \frac{3}{4} ar + \frac{3}{16} a^2 r^2 \right) U_j^n ; \quad j = 5, \dots, N . \end{aligned} \quad (44)$$

The local truncation error of (42) for $j = 4, \dots, N$ and $n = 0, 1, \dots$, is

$$\left(-\frac{1}{3} a \ell h^2 \frac{\partial^3 \mathbf{u}}{\partial x^3} + \frac{1}{720} \ell^5 \frac{\partial^5 \mathbf{u}}{\partial t^5} \right)_j^n ,$$

the time component in which may be improved by extrapolating, using (12), to give

$$\left(-\frac{1}{3} a \ell h^2 \frac{\partial^3 \mathbf{u}}{\partial x^3} + \frac{1}{1890} \ell^7 \frac{\partial^7 \mathbf{u}}{\partial t^7} \right)_j^n . \quad (46)$$

In view of the dominant term $-\frac{1}{3} a \ell h^2 \frac{\partial^3 \mathbf{u}}{\partial x^3}$, however, the resulting

improvement in accuracy is unlikely to justify extrapolating in time unless h is very small.

In the event of an even higher order approximant to the space derivative being used in (1), instead of (14), the elegant methods of Gourlay and Morris [4] for improving the accuracy in time of numerical methods for parabolic equations, can be used with the relations (9), (20) of the present paper. Such an approach requires the matrix D to have increased band width. This band width would be increased still further on squaring D and more than three points near the boundary would suffer loss of accuracy when solving (31) using the higher order (2,0), (2,1), (2.2) Pade approximants, though stability would not be affected. It may, therefore, be advisable to use the technique of Gourlay and Morris [4] with a space replacement of order higher than (14), but the methods developed in the present paper can be implemented more quickly and are to be preferred for use with (14).

5. NUMERICAL EXPERIMENTS

To examine the behaviour of the methods discussed in sections 2,3,4, the methods based on the (1,1), (2,0), (2,1),(2,2) Pade approximants are tested on a number of problems from the literature. When these four Pade approximants are used in conjunction with the matrix C given by (6) they will be named C11,C20,C21,C22, respectively, and when used in conjunction with the matrix D they will be named D11,D20,D21,D22, respectively. All computations were carried out using single precision on a CDC 7600 computer.

The differential equation on which the methods are tested is

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0 ,$$

the initial and boundary conditions being different for each problem.

Problem 1 (Oliger [3])

Here, the initial conditions are taken to be

$$g(x) = \sin 2k\pi x ; \quad x \geq 0$$

and the boundary conditions to be

$$v(t) = -\sin 2k\pi t ; \quad t > 0$$

where k is a positive integer. The theoretical solution of this problem is

$$u(x,t) = \sin 2k\pi(x-t)$$

and the numerical solution will be calculated for $0 < x \leq 1$. The integer

k gives the number of complete waves in the interval $0 \leq x \leq 1$.

The boundedness of the solution and the build—up of error may be examined with reference to two norms, as in Oliger [13], Let

$z_j^n = u(jh, n\ell) - U_j^n$ with $j = 0, 1, \dots, N$ and $n = 0, 1, \dots$, so that \tilde{z}^n is the vector of such errors and has $N + 1$ elements, and let

$\tilde{V}^n = [U_0^n, U_1^n, \dots, U_N^n]^T$ be the vector (or order $N + 1$) of solutions,

including the boundary condition, at time $t = n\ell$. The norms are

defined by

$$\|\tilde{z}^n\|_{\infty} = \max_j |z_j^n|, \quad \|\tilde{z}^n\|_2^2 = h \sum_{j=0}^N |z_j^n|^2, \quad \|\tilde{V}^n\|_2^2 = h \sum_{j=0}^N |U_j^n|^2.$$

The solution was computed with $h = 1/640$, $\ell = 1/80$, $r = 8$ and

$k = 2$; the values of $\|\tilde{V}\|_2$, $\|\tilde{z}\|_2$, $\|\tilde{z}\|_{\infty}$ at time $t = 0.5, 1.0, 2.0$

and 4.0 are given in Table II. Choosing this small value of h has

the effect of lessening the emphasis of the component $-\frac{1}{2}a\ell h^2 u/\partial x^2$

in (22) *et seqq*, and the component $-\frac{1}{3}a\ell h^2 \partial^3 u/\partial x^3$ in (23) *et seqq*. The

increased number of mesh points at each time level can be appreciably offset by using a large value of ℓ , and consequently of r . In the

paper by Oliger [13], for example, r was given the value $\frac{1}{4}$ compared with the value 8 in the present experiment.

Visual analysis of Table II, and comparison with Table 3.1 in [13], shows that the errors for all eight formulations involving the matrices C and D show very little increase in magnitude after time $t = 1.0$.

That is to say, the errors reach their maximum values very quickly, there being very little accumulation of errors after time $t = 1.0$.

This observation contrasts with the results of Table 3.1 in [13] where the errors, generally, show a gradual growth as time increases. The stagnation of errors experienced using the two-time level methods of

the present paper make them suitable for use with large values of t . The maximum error of each method was seen to be in keeping with the truncation errors given in sections 2,3,4. The methods based on the (2,1) and (2,2) Padé approximants showed the greatest improvement when used with the matrix D (for any value of t), the corresponding improvements in the performance of the methods based on the (1,1) and (2,0) Padé approximants being less pronounced.

Problem 2 (Abarbanel et al. [1])

The boundary conditions and the initial conditions for this problem are the same as for Problem 1. The parameter k is given the value 4 and the solution computed with $h = 1/640$, $\ell = 1/80$, $r = 8$; the numerical results at time $t = 10.0$ are given in Table III. The corresponding results for $k = 4$ are given in Table 4 of Abarbanel et al. [1] where the ratio r was given the value 0.9. In their Table 4 Abarbanel et al. [1] compare their results with earlier work by a number of authors including Boris and Book [2], Kreiss and Oliger [9], Oliger [13], and Richtmyer [14]. The results of the present paper show that the methods developed are very competitive with all methods tested in [1] for $k = 4$. The growth of errors as a result of increasing the wave frequency was not as pronounced as any of the methods tested in [1]. Allowing a factor of 3 for the faster CDC 7600 over the CDC 6600 used by Abarbanel et al. [1], the CPU times quoted in Table III are generally superior to the figures quoted in [1]. This observation is strengthened when it is further noted that the CPU times in Table III include the time taken to compute $\left\| z \right\|_{\infty}$ by 640 comparison statements in the computer program.

It is confirmed again that the use of a small value of h in the methods which have higher accuracy in time, produces accuracy as high as do

those methods, tested in [1,13] with a larger value of h , which have $O(h^4)$ error in space.

Problem 3 (Khaliq and Twizell [8])

The boundary condition for this problem is

$$u(0,t) = t ; \quad t > 0$$

and the initial condition is

$$u(x,0) = 1 + x ; \quad x \geq 0 .$$

The theoretical solution of the problem is

$$u(x,t) = 1 + x - t , \quad x \geq t$$

$$u(x,t) = t - x , \quad x < t$$

so that there exists a discontinuity in the solution across the line $t = x$ in the (x,t) plane.

Problem 4 (Khaliq and Twizell [8])

Here, the initial condition is

$$u(x,0) = e^x , \quad x \geq 0$$

and the boundary condition is

$$u(0,t) = e^t , \quad t > 0 .$$

The theoretical solution of the problem is

$$u(x, t) = e^{x-t} \quad , \quad x \geq t$$

$$u(x, t) = e^{t-x} \quad , \quad x < t$$

so that there exist discontinuities in the first derivatives across the line $t = x$ in the (x, t) plane.

Problems 3 and 4 were tested with $h = 1/80$, $\ell = 1/20$, $r = 4$ and the results are given at time $t = 1.0$ in Tables IV, V respectively. It is noted again that the methods based on the (2,1) and (2,2) Pade approximants showed greater improvements than the improvements shown by the methods based on the (1,1) and (2,0) Pade approximants. Using the higher order space approximation the highest accuracy was achieved by method D22 followed, in succession, by D21, D11, D20. This is in keeping with the local truncation errors of these methods and with the numerical results obtained for Problems 1 and 2. It was also found, as the computation proceeded, that, away from the boundary, the greatest errors were at those mesh points close to the line $t = x$ across which there were discontinuities.

Problem 5

The boundary condition for this problem is taken to be

$$u(0, t) = e^{-t} \quad , \quad t > 0$$

and the initial condition to be

$$u(x, 0) = e^x \quad , \quad 0 \leq x \leq 1 .$$

The theoretical solution of the problem is

$$U(x, t) = e^{x-t}$$

which decays as time increases. The problem was run with $h = 1/80$, $\ell = 1/20$ and $r = 4$; the numerical results at time $t = 10.0$ are given in Table VI.

The errors were found to behave in much the same way as in the other problems; that is, using the higher order space approximant produced a more noticeable improvement in the methods based on the (2,1),(2,2) Pade approximants than in the other two methods. The two formulations based on the (1,1) Pade approximant are seen to give very good results at time $t = 10.0$ when for $0 \leq x \leq 1$, the solution lies in the approximate interval $4.540 \times 10^{-5} < u < 1.234 \times 10^{-4}$. This is due to these formulations using fewer mesh points and thus experiencing smaller round off errors.

6. CONCLUSIONS

Two families of two—time level finite difference schemes, based on Pade approximants to the matrix exponential function, have been developed in this paper for the numerical solution of first order hyperbolic partial differential equations with initial and boundary conditions specified.

First of all, the space derivative was replaced by the usual first order backward difference approximant at each mesh point at a given time level and the resulting system of first order ordinary differential equations was solved using the (1,1),(2,0),(2,1),(2,2) Padé pproximants. Next, the space derivative at the mesh point adjacent to the boundary, at a given time level, was replaced by the same low order approximant, and by the usual second order backward difference approximant at all other

mesh points. The resulting system of ordinary differential equations was solved using the same four Pade approximants.

All four numerical methods of each family were implicit in nature; those based on the (1,1) and (2,2) Pade approximants were seen to be A_0 -stable and those based on the (2,0) and (2,1) Padé approximants were seen to be L_0 -stable. The form of the given boundary conditions, however, meant that the methods were all used explicitly, obviating the need to solve a linear algebraic system. The CPU times for all eight methods were found to be fast.

The methods were tested on five problems from the literature; the results obtained were better than other results in the literature, even though the orders of the methods in the present paper are, in many cases, lower. It was found that the lower order (1,1) and (2,0) Padé approximants gave good results when the low order replacement of the space derivative was used at each mesh point at a given time level, and that the higher order (2,1) and (2,2) Padé approximants gave their best results when the higher order replacement of the space derivative was used at interior mesh points. This implies that low order replacements in both space and time, or higher order replacements in both space and time, are most effective; this observation was also made by Abarbanel et al.[1,p.351]. For problems with decaying solutions, the two formulations based on the (1,1) Pade approximant give very good results due to the smaller number of mesh points used, thus reducing round-off errors.

Table I: Error constants and the extrapolated forms of the five methods based on the (m,k) Pade approximants $m = 1,2$, $k = 0(1)m$.

Method (Padé)	Error constant C_q	Extrapolated From	Error constant E_s
(1,0)	$c_{q=-1/2}$	$2 \underline{U}^{(1)} - \underline{U}^{(2)}$	$E_3 = 4/3$
(1,1)	$c_3 = -1/2$	$(4 \underline{U}^{(1)} - \underline{U}^{(2)})/3$	$E_5 = 1/10$
(2,0)	$c_3 = 1/6$	$(4 \underline{U}^{(1)} - \underline{U}^{(2)})/3$	$E_4 = -1/3$
(2,1)	$c_4 = 1/72$	$(8 \underline{U}^{(1)} - \underline{U}^{(2)})/7$	$E_5 = -8/945$
(2,2)	$c_5 = 1/720$	$(16 \underline{U}^{(1)} - \underline{U}^{(2)})/15$	$E_7 = -1/1890$

Table II: Numerical results for Problem 1 at time $t = 0, 0.5, 1.0, 2.0, 4.0$

Method	$\ \underline{V}_{\sim}\ _2$	$\ \underline{Z}_{\sim}\ _2$	$\ \underline{Z}_{\sim}\ _{\infty}$	CPU(sec)	$\ \underline{V}_{\sim}\ _2$	$\ \underline{Z}_{\sim}\ _2$	$\ \underline{Z}_{\sim}\ _{\infty}$	CPU(sec)
	$t = 0.5$				$t = 1.0$			
C11	6.75(-1)	3.55(-2)	6.08(-2)	0.062	6.65(-1)	5.00(-2)	1.07(-1)	0.115
C20	6.70(-1)	5.76(-2)	1.01(-2)	0.070	6.58(-1)	8.55(-2)	1.56(-1)	0.123
C21	6.74(-1)	1.79(-1)	6.01(-2)	0.074	6.64(-1)	2.07(-1)	1.07(-1)	0.137
C22	6.75(-1)	1.71(-1)	5.98(-2)	0.078	6.64(-1)	1.97(-1)	1.04(-1)	0.145
D11	7.07(-1)	7.03(-3)	1.21(-2)	0.084	7.06(-1)	1.00(-2)	2.40(-2)	0.158
D20	7.03(-1)	4.67(-2)	9.11(-2)	0.088	7.00(-1)	7.20(-2)	1.17(-1)	0.169
D21	7.06(-1)	1.31(-2)	2.66(-3)	0.095	7.05(-1)	1.89(-2)	2.70(-2)	0.179
D22	7.06(-1)	1.23(-3)	2.42(-3)	0.119	7.06(-1)	1.75(-2)	2.71(-3)	0.227
$t = 2.0$				$t = 4.0$				
C11	6.65(-1)	5.00(-2)	1.07(-1)	0.218	6.65(-1)	5.00(-2)	1.07(-1)	0.425
C20	6.59(-1)	8.61(-2)	1.56(-1)	0.249	6.59(-1)	8.61(-2)	1.56(-1)	0.487
C21	6.64(-1)	2.07(-1)	1.07(-1)	0.264	6.64(-1)	2.07(-1)	1.07(-1)	0.517
C22	6.64(-1)	1.97(-1)	1.04(-1)	0.279	6.64(-1)	1.97(-1)	1.04(-1)	0.547
D11	7.06(-1)	1.00(-2)	2.92(-2)	0.305	7.06(-1)	1.00(-2)	2.43(-1)	0.600
D20	7.00(-1)	7.30(-2)	1.27(-1)	0.312	7.00(-1)	7.30(-2)	1.27(-1)	0.689
D21	7.05(-1)	1.90(-2)	2.76(-2)	0.347	7.05(-1)	1.90(-2)	2.76(-2)	0.791
D22	7.05(-1)	1.75(-3)	2.71(-3)	0.445	7.06(-1)	1.75(-3)	2.71(-3)	0.877

Table III: Numerical results for Problem 2 at time $t = 10$.

Method	$\left\ \underset{\sim}{\mathbf{V}} \right\ _2$	$\left\ \underset{\sim}{\mathbf{z}} \right\ _2$	$\left\ \underset{\sim}{\mathbf{z}} \right\ _\infty$	CPU(sec)
C11	5.61(-1)	1.86(-1)	4.00(-1)	1.049
C20	5.29(-1)	2.87(-1)	5.73(-1)	1.121
C21	5.54(-1)	3.82(-1)	3.85(-1)	1.278
C22	5.59(-1)	3.72(-1)	3.75(-1)	1.372
D11	7.04(-1)	8.05(-2)	1.94(-1)	1.483
D20	6.58(-1)	2.46(-1)	4.81(-1)	1.590
D21	6.96(-1)	4.17(-1)	6.47(-2)	1.697
D22	7.07(-1)	8.28(-3)	4.48(-2)	2.178

Table IV: Numerical results for Problem 3 at time $t = 1.0$

Method	$\left\ \underset{\sim}{\mathbf{V}} \right\ _2$	$\left\ \underset{\sim}{\mathbf{z}} \right\ _2$	$\left\ \underset{\sim}{\mathbf{z}} \right\ _\infty$	CPU(sec)
C11	1.78	2.01(-2)	9.64(-2)	0.007
C20	1.83	5.76(-2)	1.22(-1)	0.007
C21	1.80	1.59(-1)	1.01(-1)	0.008
C22	1.78	1.67(-1)	9.55(-2)	0.008
D11	1.76	1.75(-2)	4.00(-2)	0.008
D20	1.82	4.82(-2)	7.78(-2)	0.008
D21	1.79	1.62(-2)	3.72(-2)	0.009
D22	1.78	4.51(-3)	2.78(-3)	0.010

Table V: Numerical results for Problem 4 at time $t = 1.0$

Method	$\left\ \underset{\sim}{\mathbf{V}} \right\ _2$	$\left\ \underset{\sim}{\mathbf{z}} \right\ _2$	$\left\ \underset{\sim}{\mathbf{z}} \right\ _\infty$	CPU(sec)
C11	5.97(-1)	1.40(-1)	5.76(-1)	0.009
C20	5.98(-1)	2.51(-1)	5.79(-1)	0.010
C21	5.99(-1)	2.38(-1)	5.50(-1)	0.011
C22	5.97(-1)	2.34(-1)	5.62(-1)	0.012
D11	5.83(-1)	9.04(-2)	5.40(-1)	0.012
D20	5.90(-1)	9.78(-2)	5.48(-1)	0.012
D21	5.82(-1)	8.53(-2)	5.34(-1)	0.013
D22	5.79(-1)	8.60(-2)	5.18(-1)	0.016

Table VI: Numerical results for Problem 5 at time $t = 10.0$

Method	$\ z_{\sim}\ _2$	$\ z_{\sim}\ _{\infty}$
C11	1.73(-7)	4.01(-7)
C20	6.61(-6)	8.94(-6)
C21	1.41(-3)	2.86(-6)
C22	7.46(-4)	1.20(-6)
D11	4.26(-6)	8.94(-6)
D20	6.37(-6)	9.65(-6)
D21	2.36(-6)	8.60(-6)
D22	2.74(-6)	7.96(-6)

REFERENCES

1. S. Abarbanel, D. Gottlieb and E. Turkel, *Difference schemes with fourth order accuracy for hyperbolic equations*, SIAM J. Appl. Math., 29 (1975), 329-351.
2. J.P. Boris and D.L. Book, *Flux corrected transport I. SHASTA A fluid transport algorithm that works*, J. Computational Phys., 11 (1973), 38-64.
3. M. Goldberg and E. Tadmor, *Scheme-independent stability criteria for difference approximations of hyperbolic initial-boundary value problems. I*, Math. Comp., 32 (1978), 1097-1107.
4. A.R. Gourlay and J. L1, Morris, *The extrapolation of first order methods for parabolic partial differential equations. II*, SIAM J. Numer. Anal., 17 (1980), 641-655.
5. B. Gustafsson, *The convergence rate for difference approximations to mixed initial boundary value problems*, Math. Comp., 29 (1975), 396-406.
6. B. Gustafsson, H.-O. Kreiss and A. Sundström, *Stability theory of difference approximations for mixed boundary value problems. II*. Math. Comp. 26 (1972), 649-686.
7. A.Q.M. Khaliq, *Ph.D. Thesis*, Brunel University, 1983.
8. A.Q.M. Khaliq and E.H. Twizell, *The extrapolation of stable finite difference schemes for first order hyperbolic equations*, Intern. J. Computer Math., 11 (1982), 155-167
9. H.-O. Kreiss and J. Olinger, *Comparison of accurate methods for the integration of hyperbolic equations*, Tellus, 24 (1972), 199-215.
10. J.D. Lambert, *Computational Methods in Ordinary Differential Equations*, John Wiley and Sons, Chichester, 1973.
11. J.D. Lawson and J.LI. Morris, *The extrapolation of first order methods for parabolic partial differential equations. I*, SIAM J. Numer. Anal. 15(1978), 1212-1224.
12. A.R. Mitchell and D.F. Griffiths, *The Finite Difference Method in Partial Differential Equations*, John Wiley and Sons, Chichester, 1980.
13. J. Olinger, *Fourth order difference methods for the initial boundary-value problem for hyperbolic equations*, Math. Comp. 28 (1974), 15-25.
14. R.D. Richtmyer, *A survey of finite difference methods for nonsteady fluid dynamics*, NCAR Research Tech. Notes 63-2, 1963.
15. E. H. Twizell and A..Q.M. Khaliq, *Onestep multiderivative methods for first order ordinary differential equations*, BIT, 21(1981), 518-527. (Also Brunel University Department of Mathematics Technical Report TR/02/82.)

**NOT TO BE
REMOVED**
FROM THE LIBRARY

XB 2261247 5

