JOURNAL OF COMPUTATIONAL PHYSICS 79, 353–372 (1988)

Truncation and Accumulated Errors in Wave Propagation

YI-LING F. CHIANG

Computer and Information Science, New Jersey Institute of Technology, Newark, New Jersey 07102

Received November 7, 1986; revised October 11, 1987

The approximation of the truncation and accumulated errors in the numerical solution of a linear initial-valued partial differential equation problem can be established by using a semi-discretized scheme. This error approximation is observed as a lower bound to the errors of a finite difference scheme. By introducing a modified von Neumann solution, this error approximation is applicable to problems with variable coefficients. To seek an in-depth understanding of this newly established error approximation, numerical experiments were performed to solve the hyperbolic equation

$$\frac{\partial U}{\partial t} = -C_1(x)\, C_2(t) \frac{\partial U}{\partial x},$$

with both continuous and discontinuous initial conditions. We studied three cases: (1) $C_1(x) = C_0$ and $C_2(t) = 1$; (2) $C_1(x) = C_0$ and $C_2(t) = t$; and (3) $C_1(x) = 1 + (x/a)^2$ and $C_2(t) = C_0$. Our results show that the errors are problem dependent and are functions of the propagating wave speed. This suggests a need to derive problem-oriented schemes rather than the equation-oriented schemes as is commonly done. Furthermore, in a wave-propagation problem, measurement of the error by the maximum norm is not particularly informative when the wave speed is incorrect. © 1988 Academic Press, Inc.

1. INTRODUCTION

A difficulty in numerical solutions of a hyperbolic partial differential equation problem is the smearing of solution discontinuities [2, 3, 5, 14–16, 21]. In the neighborhood of a discontinuity, the accuracy of the solution may not be improved by using higher order schemes, regardless of their increasing complexity [13]. In particular, in solving a PDE problem with variable coefficients, even linear ones, the employment of a higher order scheme might be too complex to be practical. It is therefore often asked whether a higher order scheme should be employed or might an error approximation be possible prior to the employment of a scheme?

Error bounds have been studied by many authors [17–19]. As a result of Vichnevetsky's work in the late seventies, it was suggested that the errors in the numerical solution of a problem might be problem dependent rather than equation dependent. He related the errors of a scheme not only to the given equations but also to the given initial conditions. He did not, however, pursue this finding. His

0021-9991/88 $3.00

formula is applicable only to linear hyperbolic equations with constant coefficients [23].

Currently, the numerical results of several authors from solving more general problems than the one considered by Vichnevetsky appear to support the assumption that the errors are problem dependent [7, 18]. Hence, we intend to establish an error estimate technique applicable to a more general class of problems than only linear hyperbolic equations with constant coefficients.

We consider a linear initial-value PDE problem in one space variable [24] as in the problem; the given equation is

$$\frac{\partial U}{\partial t} = \tilde{X} \cdot U \tag{1.1}$$

where the differential operator $\tilde{X}\cdot$ is

$$\tilde{X}\cdot = \sum_{i=0}^{N} C_{i1}(x)\, C_{i2}(t) \frac{\partial^i}{\partial x^i}, \tag{1.2}$$

and the given initial condition $U(x, 0) = U_0(x)$ is either continuous or piecewise continuous.

It is well known that (1.1) is satisfied by the von Neumann solution

$$U(x, t) = \hat{a}_w(t)\, e^{iwx} \tag{1.3}$$

in the frequency domain [25] if the coefficients $C_{i1}(x)$ in $\tilde{X}\cdot$ are all constants. This leads to Vichnevetsky's error approximation. However, (1.3) is not a solution of (1.1) if the coefficients are not all constants. We suggest the modification

$$U(x, t) = \hat{a}_w(t)\, \phi_w(x)\, e^{iwx} \tag{1.4}$$

where $\phi_w(x)$ is arbitrary and will be determined later.

In the following sections, the modified von Neumann solution and the error approximation are described. To further understand this newly derived error approximation, the following linear hyperbolic problem is studied with both continuous and piecewise continuous initial conditions:

$$\frac{\partial U}{\partial t} = -C_1(x)\, C_2(t) \frac{\partial U}{\partial x}. \tag{1.5}$$

We considered three cases: (1) $C_1(x) = C_0$ and $C_2(t) = 1$; (2) $C_1(x) = C_0$ and $C_2(t) = t$; and (3) $C_1(x) = 1 + (x/a)^2$ and $C_2(t) = C_0$, where C_0 and a are constants. Our results support the assumption that the errors are problem dependent. This suggests that an appropriate scheme should be derived problem oriented rather than equation oriented as is commonly done. We observe that the error approximation forms a good lower bound to the errors encountered in a finite difference scheme. Futhermore, the errors are functions of the propagating speed. If the speed of the wave varies with its position, then a wave distortion occurs, and the numerical wave is not propagating along the path of the exact wave. After a time period, an initial point will not travel to the same point by following the numerical or the exact wave. A delay (or an advance) of the numerical wave on the arrival at a fixed point relative to the exact wave raises doubt on the meaning of conventional pointwise error measurement technique. Trefethen [22] has also considered variable speed wave propagation problems, but from a very different point of view.

II. Truncation and Accumulated Errors

In general, the von Neumann solution may not satisfy (1.1). Hence, we consider the modified von Neumann solution

$$U(x, t) = \hat{a}_w(t)\, \phi_w(x)\, e^{iwx} \tag{1.4}$$

where $\phi_w(x)$ satisfies

$$\frac{d \log \hat{a}_w(t)}{dt} = \frac{\tilde{X} \cdot [\phi_w(x)\, e^{iwx}]}{\phi_w(x)\, e^{iwx}}. \tag{2.1}$$

As defined in (2.1), $\phi_w(x)$ will exist if $C_{i2}(t)$ in (1.2) are all connected by a relation of the form $C_{i2}(t) = C_i g(t)$ for some function g and for constants C_i. (This is indicated by the referee.) However, if $\phi_w(x)$ does exist, then a space transformation is possible, i.e.,

$$y = x + \frac{1}{iw} \log \phi_w(x), \tag{2.2}$$

and the exact and numerical solutions of (1.1), $U(x, t)$ and $U_h(x_n, t)$, respectively, may be found as

$$U(x, t) = \int_{-\infty}^{\infty} \hat{a}_w(0)\, \phi_w(x) \cdot \exp\left(iwx + \int_0^t \hat{X}(w, \xi)\, d\xi\right) dw \tag{2.3}$$

and

$$U_h(x_n, t) = \int_{-\infty}^{\infty} \hat{a}_w(0)\, \phi_w(x_n) \cdot \exp\left(iwx_n + \int_0^t \hat{A}(w, \xi)\, d\xi\right) dw, \tag{2.4}$$

where

$$\hat{X}(w, t) = \frac{\hat{X} \cdot [\phi_w(x)\, e^{iwx}]}{\phi_w(x)\, e^{iwx}} \tag{2.5}$$

and

$$\hat{A}(w, t) = \frac{\tilde{A} \cdot [\phi_w(x_n)\, e^{iwx_n}]}{\phi_w(x_n)\, e^{iwx_n}}. \tag{2.6}$$

In (2.3) and (2.4), $\hat{a}_w(0)$ is the Fourier transform of $U_0(x)$ in the transformed y-space. In (2.6), $\tilde{A}\cdot$ is a finite difference operator defined also in the y-space and the numerical solution $U_h(x_n, t)$ satisfies

$$\frac{dU_h(x_n, t)}{dt} = \tilde{A}\cdot U_h(x_n, t),$$
$$U_h(x_n, 0) = U_0(x_n). \tag{2.7}$$

Details on the existence of $\phi_w(x)$ may be found in Ref. [9].

Let $E_{\mathrm{T}}(x_n, t)$ and $e(x_n, t)$ be the truncation and acummulated errors at the mesh point $x = x_n$, respectively. Then we have

$$E_{\mathrm{T}}(x_n, t) = \tilde{X}\cdot U(x_n, t) - \tilde{A}\cdot U(x_n, t) \tag{2.8}$$

and

$$e(x_n, t) = U(x_n, t) - U_h(x_n, t). \tag{2.9}$$

By considering that $U(x, t)$ is band-limited, i.e., its Fourier transform vanishes for $|w| \geqslant \pi/h$, where h is a constant [5, 6, 20], and by using the sampling theory and Whittaker's theory [26], the truncation and accumulated errors in a wave propagation problem can be estimated as

$$e^*(x, t) = \int_{-\pi/h}^{\pi/h} \hat{a}_w(0)\, \phi_w(x)\, e^{iwx} \left\{ \exp\left(\int_0^t \hat{X}(x, \xi)\, d\xi \right) - \exp\left(\int_0^t \hat{A}(w, \xi)\, d\xi \right) \right\} dw \tag{2.10}$$

and

$$E_T^*(x, t) = \int_{-\pi/h}^{\pi/h} \hat{a}_w(0)\, \phi_w(x)\, e^{iwx} \{ \hat{X}(w, t) - \hat{A}(w, t) \} \times \exp\left(\int_0^t \hat{X}(w, \xi)\, d\xi \right) dw. \tag{2.11}$$

In (2.10) and (2.11), both $\hat{a}_w(0)$ and $\hat{A}(w, t)$ are defined in the transformed y-space. However, this restriction may be removed as described in Ref. [9]. Furthermore, it can be shown that at the mesh point $x = x_n$,

$$e^*(x_n, t) = e(x_n, t) \tag{2.12}$$

and

$$E_{\mathrm{T}}^*(x_n, t) = E_{\mathrm{T}}(x_n, t). \tag{2.13}$$

Hence, $e^*(x, t)$ and $E_T^*(x, t)$ can be called the accumulated and truncation errors everywhere, respectively.

III. The Problem

We studied three wave propagation problems in the general form

$$\begin{gathered} \frac{\partial U}{\partial t} = -C_1(x)\, C_2(t) \frac{\partial U}{\partial x}, \\ U(x, 0) = U_0(x) \quad \text{is given,} \end{gathered} \tag{1.5}$$

where the initial wave is either continuous or piecewise continuous. As described in (1.5), the initial wave $U_0(x)$ is propagating with a speed $v = C_1(x)\, C_2(t)$. In the study, the continuous wave is given to be a sinusoidal wave

$$U_0^1(x) = \sin k\pi x, \qquad -1 \leqslant k \leqslant 1, \tag{3.1}$$

and the discontinuous wave is

$$U_0^2(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geqslant 0. \end{cases} \tag{3.2}$$

However, $U_0^2(x)$ is not band-limited, hence, the newly established error approximation may not be applied. For this reason, we used the discrete Fourier series of $U_0^2(x)$ in the interval $[-N\,\Delta x, N\,\Delta x]$ instead, i.e., for an even N,

$$U_0^3(x_n) = \frac{1}{2} + \frac{1}{2N} \sum_{k=-N/2}^{N/2-1} \left\{ \frac{\cos \dfrac{2k+1}{2N}\pi}{\sin \dfrac{2k+1}{2N}\pi} \sin \frac{2k+1}{N} n\pi + \cos \frac{2k+1}{N} n\pi \right\}. \tag{3.3}$$

$U_0^3(x_n)$ is defined at the mesh point $x = x_n$. It has a period of $2N\,\Delta x$. Hence, in the wave propagation, a reflected wave occurs. To avoid the disturbance of the reflected wave near the discontinuity, N is taken to be a large number. This, however, causes an increase in computing time. This disadvantage may be reduced to a minimum with the application of the fast Fourier transform [12].

First Problem

The first problem we considered is that $C_1(x) = C_0$ and $C_2(t) = 1$; hence, the initial wave is propagating with a constant speed $v = C_0$ [1] toward the right along the characteristic

$$\frac{dx}{dt} = C_0, \tag{3.4}$$

which represents a family of straight lines. With the initial waves being $U_0^1(x)$ and $U_0^3(x)$, respectively, the propagating waves correspondingly are

$$U^{(1)}(x, t) = \sin[k\pi(x - C_0 t)] \tag{3.5}$$

and

$$U^{(3)}(x_n, t) = \frac{1}{2} + \frac{1}{2N} \sum_{k=-N/2}^{N/2-1} \left\{ \frac{\cos \frac{2k+1}{2N}\pi}{\sin \frac{2k+1}{2N}\pi} \sin\left[\frac{2k+1}{N}\pi\left(n - \frac{C_0 t}{\Delta x}\right)\right] + \cos\left[\frac{2k+1}{N}\pi\left(n - \frac{C_0 t}{\Delta x}\right)\right]\right\}. \tag{3.6}$$

Second Problem

The second problem describes a wave propagating in a medium which is under constant pressure in the direction of propagation. We assume that $C_1(x) = C_0$ and $C_2(t) = t$; hence, the initial wave propagates with a variable speed $v = C_0 t$ toward the right. At any time $t = t_0$ the wave moves uniformly with the speed $v = C_0 t_0$; however, v changes from time to time. No distortion of the wave will occur, but in every fixed time interval, the traveling distance of the wave is different. The characteristic [11] is given as

$$\frac{dx}{dt} = C_0 t, \tag{3.7}$$

which represents a family of parabolas. Even though $\phi_w(x) \equiv 1$ in this case, the von Neumann solution is

$$U(x, t) = \exp\left(iw\left(x - \frac{C_0 t^2}{2}\right)\right). \tag{3.8}$$

The propagating waves may be written as

$$U^{(1)}(x, t) = \sin\left[k\pi\left(x - \frac{C_0 t^2}{2}\right)\right] \tag{3.9}$$

and

$$U^{(3)}(x_n, t) = \frac{1}{2} + \frac{1}{2N} \sum_{n=-N/2}^{N/2-1} \left\{ \frac{\cos \frac{2k+1}{2N}\pi}{\sin \frac{2k+1}{2N}\pi} \sin\left[\frac{2k+1}{N}\pi\left(n - \frac{C_0 t^2}{2\,\Delta x}\right)\right] + \cos\left[\frac{2k+1}{N}\pi\left(n - \frac{C_0 t^2}{2\,\Delta x}\right)\right]\right\}. \tag{3.10}$$

Third Problem

This problem shows a wave propagation in a nonuniform medium and is chosen for the special properties of the wave as described below. We assume that $C_1(x) = 1 + (x/a)^2$ and $C_2(t) = C_0$. Hence, the wave is propagating with a speed $v = C_0(1 + (x/a)^2)$ toward the right along the characteristic

$$\frac{dx}{dt} = C_0\left(1 + \left(\frac{x}{a}\right)^2\right). \tag{3.11}$$

Equation (3.11) represents a family of curves,

$$\tan^{-1}\frac{(x - x_0)a}{a^2 + xx_0} = \frac{C_0 t}{a} \tag{3.12}$$

where x_0 indicates the initial position. Each curve in this family is composed of many branches. However, in physical reality, a wave will not return after approaching infinity. Hence, the existence of the wave may not exceed the time period $T_0 = \pi a/2C_0$ to be called the time threshold. Furthermore, since each point on the wavefront has a different speed, the wavefront undergoes a distortion.

In this problem,

$$\phi_w(x) = \exp\left(iwa\left(\tan^{-1}\frac{x}{a} - \frac{x}{a}\right)\right), \tag{3.13}$$

and the exact solution are

$$U^{(1)}(x, t) = \sin\left[ka\pi\frac{\left(x - a\tan\dfrac{C_0 t}{a}\right)}{\left(a + x\tan\dfrac{C_0 t}{a}\right)}\right] \tag{3.14}$$

$$U^{(3)}(x_n, t) = \frac{1}{2} + \frac{1}{2N}\sum_{k=-N/2}^{N/2-1}\left\{\frac{\cos\dfrac{2k+1}{2N}\pi}{\sin\dfrac{2k+1}{2N}\pi}\sin\left[\frac{2k+1}{N\Delta x}\frac{a\left(n\Delta x - a\tan\dfrac{C_0 t}{a}\right)}{\left(a + n\Delta x\tan\dfrac{C_0 t}{a}\right)}\right]\right.$$
$$\left. + \cos\left[\frac{2k+1}{N\Delta x}\pi\frac{a\left(n\Delta x - a\tan\dfrac{C_0 t}{a}\right)}{\left(a + n\Delta x\tan\dfrac{C_0 t}{a}\right)}\right]\right\}. \tag{3.15}$$

IV. Numerical Results

We employed five stable schemes to solve the described problems. Two of them are finite difference schemes and the others are semi-discretized finite difference schemes. The details on the schemes and the numerical solutions of the semi-discretized schemes can be found in Appendices 1 and 2. The performed experiments are described in Table I. In the table, the symbols, "F_i, $i=1, 2$" and "S_i, $i=1, 2, 3$" represent the finite difference and semi-discretized schemes of order i, and "C" and "D" the continuous and discontinuous initial waves, respectively. The results of the experiments are shown in Figs. 1–7. In the figures except Figs. 4 and 5, only the circled points represent the observed data and the lines drawn between the points are mainly used to connect the data points.

Figure 1 shows the errors observed by employing schemes $F1$ and $F2$ in the experiments of a continuous initial wave. In the figure, the two curves show several local maxima and minima. On the first curve, the error curve of the first order finite difference scheme, the minima occur at the points at which the second order derivative of the solution vanishes; however, on the second order scheme curve, the third order derivative vanishes. (*Remark*: the truncation errors of schemes $F1$ and $F2$ are proportional to the second and third order derivatives of the solution, respectively.) This suggests that the errors are problem dependent. Furthermore, at the points $x=0.50$ and 1.50, the errors of the first order scheme are obviously

TABLE I

Experiment	Problem	Scheme	Mesh size	Time interval	Wave speed	Duration	Wave type	Figure
1	1	$F1$	1/16	1/40	1.0	0.5	C	1 and 2
2	1	$F2$	1/16	1/40	1.0	0.5	C	1 and 2
3	1	$S1$	1/16	1/40	1.0	0.5	C	2
4	1	$S2$	1/16	1/40	1.0	0.5	C	2
5	1	$S3$	1/16	1/40	1.0	0.5	C	2
6	2	$F1$	1/16	1/40	t	1.0	D	3
7	2	$F2$	1/16	1/40	t	1.0	D	3
8	2	$S1$	1/16	1/40	t	1.0	D	3
9	3				1.0	0.5	C	4
10	3				$1+x^2$	0.5	C	4
11	3	$F2$	1/16	1/80	1.0	0.5	C	5
12	3	$F2$	1/16	1/80	$1+x^2$	0.5	C	5
13	1	$F1$	1/16	1/160	1.0	0.125	C	6
14	1	$F1$	1/16	1/160	1.0	0.250	C	6
15	1	$F1$	1/16	1/160	1.0	0.375	C	6
16	1	$F2$	1/16	1/160	1.0	0.125	C	7
17	1	$F2$	1/16	1/160	1.0	0.250	C	7
18	1	$F2$	1/16	1/160	1.0	0.375	C	7

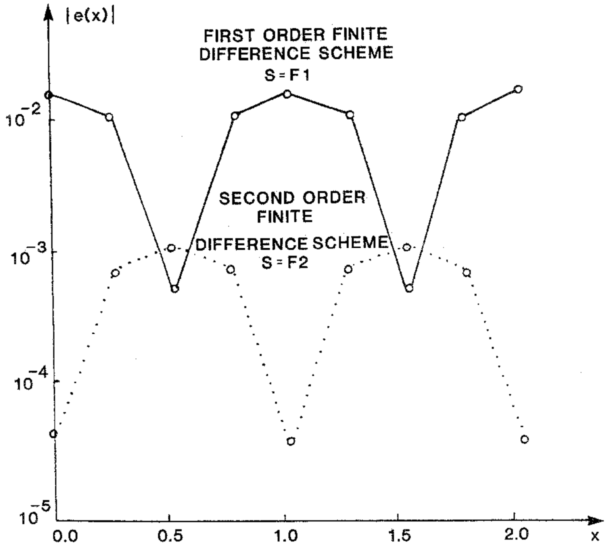

FIG. 1. The errors accumulated in a continuous wave propogation of constant speed by applying a first- and a second-order finite-difference scheme. As shown, in regions, the first-order scheme operates smaller errors than the second-order one. In the experiments, the wave speed $c = 1.0$, the mesh size $h = 1/16$, the time period $T = 0.5$, after 20 time steps. The initial wave is given as $\bar{U}_0(x) = \sin(\pi x)$.

smaller than those of the second order one. Hence, in certain regions, a lower order scheme may generate more accurate numerical solution than a higher order one. Figure 2 shows that the error approximations from $S1$ and $S2$ are indeed the lower bounds to the errors of the finite difference schemes $F1$ and $F2$, respectively. Figure 3 shows the data observed by employing the schemes $F1$, $F2$, and $S1$ in the experiments of a discontinuous initial data which is represented in a discrete Fourier series with $N = 32$ as given in (3.3). As shown, near a discontinuity, the errors are independent of the order of the schemes used. However, the estimated error still serves as a lower bound to the finite difference schemes on the right-hand side of the propagating discontinuity. The explanation on the irregularity of the errors on the left-hand side of the discontinuity may be found from Ref. [8].

Figure 4 shows two initial sinusoidal waves after 40 time steps in a time period $T = 0.5$. The dotted wave is traveling with a constant speed $v = 1.0$ and the solid one with a variable speed $v = 1 + x^2$. As shown, the solid wave suffers distortion. At the left farther end, the solid wave oscillates rapidly, hence, that part of wave is omitted. As shown in Fig. 5, large error occurs in the region with large wave distortion. Due to the instability of the schemes, outside the interval $(-\frac{6}{8}, \frac{6}{8})$, the numerical wave may not be found.

The occurrence of large errors may be explained. Consider two points, x_1^0 and x_2^0, $x_1^0 \neq x_2^0$ on the initial wave. In a time period T, those points will travel either with the wave of constant speed C_0 to x_1^1 and x_2^1 or the one of variable speed $C_0(1 + x^2)$

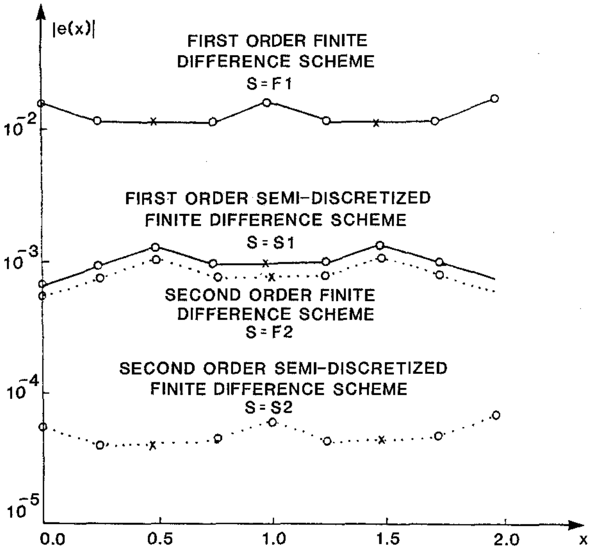

FIG. 2. The error approximations of the semi-discretized schemes were observed as lower bounds to the errors of finite-difference schemes in a continuous wave propagation. The parameters used in the experiments are the same as used in Fig. 1. The data points at the locations marked by × are deleted to make the point.

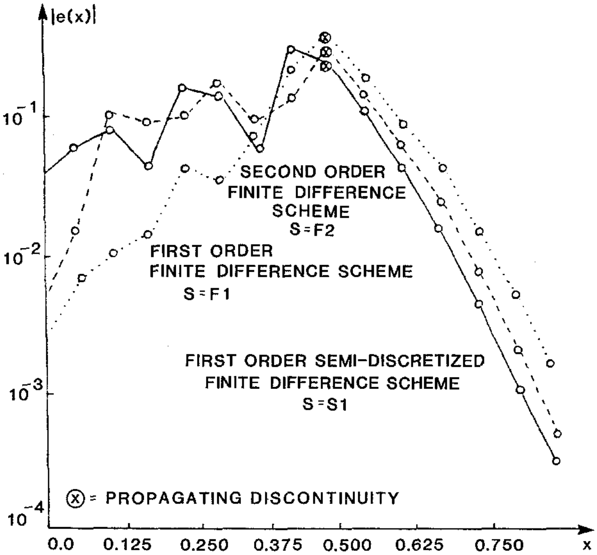

FIG. 3. The errors near the discontinuity are independent of the order of the employed schemes. However, the error approximations are still observed as lower bounds of the errors of the finite-difference scheme. In this case, the wave is propagating with a variable speed $v = t$. The initial wave is in the form of a discrete Fourier transform with $N = 32$ as given in (3.3). ○–––○, $S = F2$; ○ ··· ○, $S = F1$; ○—○, $S = S1$.

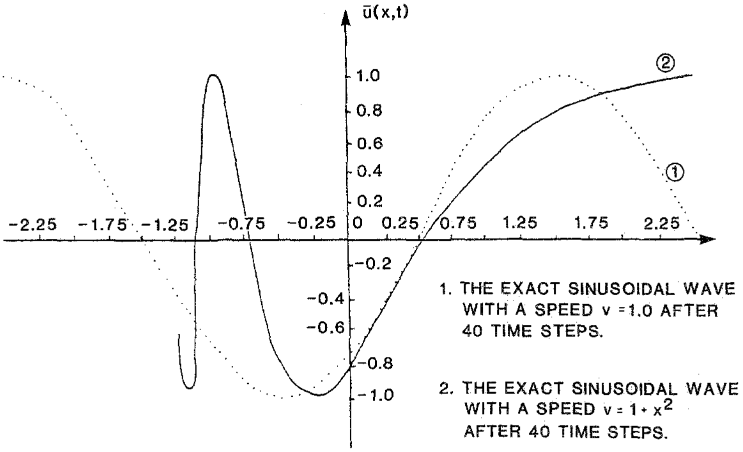

FIG. 4. The distortion of an initial sinusoidal wave with the propagation speed being a function of the space. $\bar{U}_0(x) = \sin((\pi/2)x)$, $T = 0.5$.

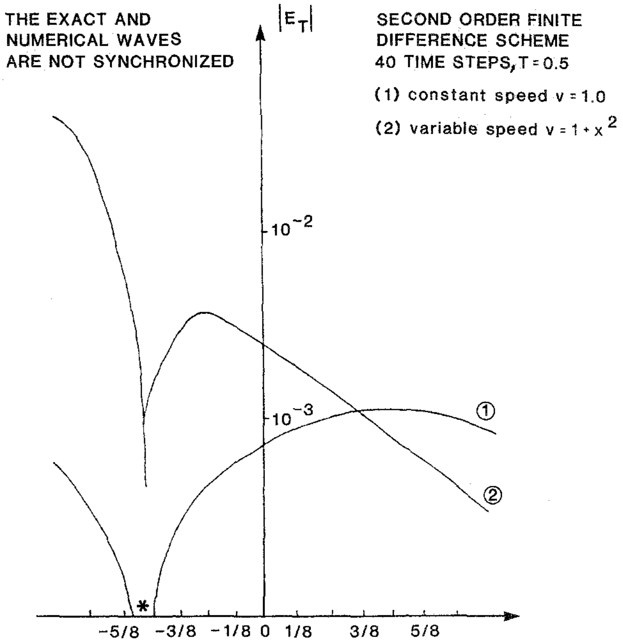

FIG. 5. The error distribution of a continuous wave propagation after 40 time steps. (1) Shows the error on the numerical wave as the initial wave is propagating with a constant speed, and (2) with a variable speed being a function of the space. In this case, the error is proportional to the third-order derivation of the solution. At $*$ the third derivative of $u(x, t)$ vanishes, hence, the error is small. The initial wave is the same as in Fig. 4.

to x_1^2 and x_2^2, respectively. Let d_1 and d_2 be the distances x_1^1 and x_2^1, and x_1^2 and x_2^2, respectively. Then literally,

$$d_1 = |x_1^0 - x_2^0| \tag{4.1}$$

and

$$d_2 = \left| (x_1^0 - x_2^0) \frac{\sec^2 C_0 T}{(1 - x_1^0 \tan C_0 T)(1 - x_2^0 \tan C_0 T)} \right|. \tag{4.2}$$

For arbitrary x_1^0, x_2^0, C_0, and T, $d_1 \neq d_2$, and the wave of variable speed suffers a distortion.

Consider again the traveling paths of the numerical and the exact waves. Assume that the initial point x_1^0 is traveling with the numerical wave. Hence, in a time interval Δt, x_1^0 will move along a straight line

$$\frac{dx}{dt} = C_0[1 + (x_1^0)^2] \tag{4.3}$$

with a constant speed $s_N = C_0[1 + (x_1^0)^2]$ to x_1^N. Its traveling distance in Δt is

$$\Delta d_1^N = s_N \Delta t = C_0 \Delta t[1 + (x_1^0)^2]. \tag{4.4}$$

In general, let $x_{1,j}^N$ be the location of x_1^0 traveling with the numerical wave at the time $t = t_j$, where $x_{1,0}^N = x_1^0$. Then in the time period $T = J \Delta t$, x_1^0 moves to $x_{1,J}^N$, and

$$d_1^N = |x_{1,J}^N - x_1^0| = \left| \sum_{j=0}^{J-1} C_0 \Delta t(1 + (x_{1,j}^N)^2) \right|, \tag{4.5}$$

with

$$x_{1,j}^N = x_{1,j-1}^N + C_0 \Delta t(1 + (x_{1,j-1}^N)^2). \tag{4.6}$$

By using (2.2), we have

$$y = \tan^{-1} x - \tan^{-1} x_1^0, \tag{4.7}$$

and in the y-space, the discretization is equally spaced, i.e.,

$$y_0 = y(x_1^0) = 0$$

and

$$y_j = jh = \tan^{-1} x_{1,j}^N - \tan^{-1} x_1^0, \tag{4.8}$$

where h is a constant. Substituting (4.8) into (4.5), we have

$$d_1^N = \left| C_0 T \left[1 + \frac{1}{J} \sum_{j=0}^{J-1} \left(\frac{\tan jh + x_1^0}{1 - x_1^0 \tan jh} \right)^2 \right] \right|. \tag{4.9}$$

However, in the same time period, if x_1^0 is following the exact wave with a speed $s=C_0(1+x^2)$, then x_1^0 will move a distance d_1^E to x_1^1, i.e.,

$$d_1^E=|x_1^0-x_1^1|=\left|[1+(x_1^0)^2]\frac{\tan C_0T}{1-x_1^0\tan C_0T}\right|, \tag{4.10}$$

and

$$\tan C_0T<\frac{1}{x_1^0}. \tag{4.11}$$

Equations (4.9) and (4.10) give the ratio $R=d_1^N/d_1^E$, and

$$R=\left|\frac{C_0T(1-x_1^0\tan C_0T)}{[1+(x_1^0)^2]\tan C_0T}\left[1+\frac{1}{J}\sum_{j=0}^{J-1}\left(\frac{\tan jh+x_1^0}{1-x_1^0\tan jh}\right)^2\right]\right|. \tag{4.12}$$

In (4.12), for fixed C_0, T, x_1^0, and J, R varies with the mesh size h. (*Remark*: the mesh size h can never be zero because the computer is a finite machine.) Hence, in T, the point x_1^0 will arrive at different points by following the exact and the numerical waves. In this case, a pointwise error measurement is measuring not only the accumulated error from the application of the numerical method, but also the

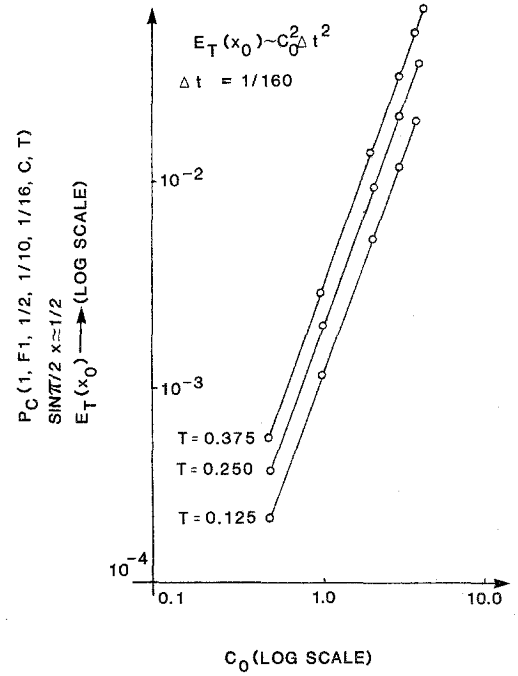

FIG. 6. The truncation error of a first-order finite-difference scheme is proportional to the square of the wave speed.

traveling difference between the numerical and exact waves. For fixed C_0, T, and x_1^0, this traveling difference increases with increasing time steps J and mesh size h.

We also observed the accumulated errors of the finite-difference schemes being functions of the wave speed. In both Figs. 6 and 7, the errors are measured at a fixed point on the propagating continuous wavefront after 20, 40, and 60 time steps. Figure 6 shows the data observed from $F1$ and initially, the fixed point x^0 satisfies the relationship $\sin(\pi/2)\, x^0 = \frac{1}{2}$. Correspondingly,

$$|E_{T1}(x_n)| = \alpha_1 \left| C_0^2 \Delta t^2 \left. \frac{\partial^2 U}{\partial x^2} \right|_{x=\xi_1} \right|, \qquad x_n \leqslant \xi_1 \leqslant x_{n+1}, \quad (4.13)$$

and

$$|E_{T2}(x_n)| = \alpha_2 \left| C_0 \left[C_0^2 - \left(\frac{\Delta x}{\Delta t} \right)^2 \right] \Delta t^3 \left. \frac{\partial^3 U}{\partial x^3} \right|_{x=\xi_2} \right|, \qquad x_n \leqslant \xi_2 \leqslant x_{n+1}, \quad (4.14)$$

where α_1 and α_2 are independent of Δt. Since the errors are measured at a fixed point on the wavefront, the derivatives $\partial^2 U/\partial x^2$ and $\partial^3 U/\partial x^3$ may be considered as constants at the measured points. This gives

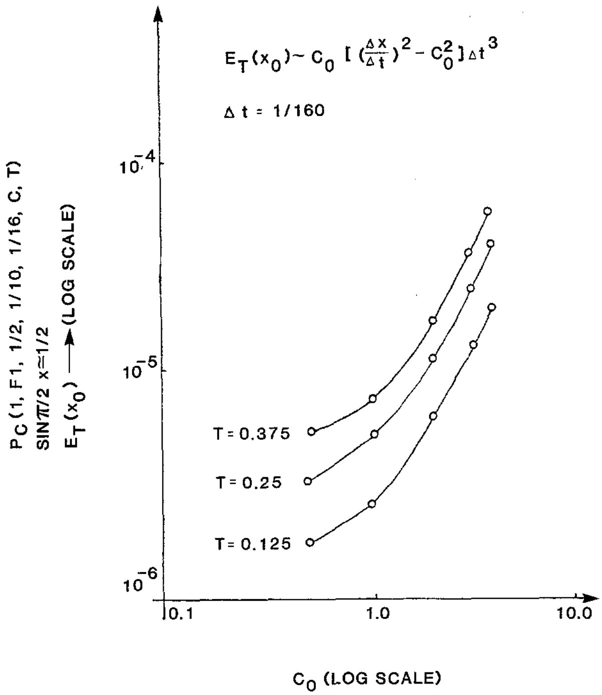

FIG. 7. The truncation error of a second-order finite-difference scheme is a function of the wave speed.

$$\log |E_{T1}(x_n)| = 2\log|C_0| + \text{constant} \tag{4.15}$$

and

$$\log |E_{T2}(x_n)| = \log|C_0| + \log\left|C_0^2 - \left(\frac{\Delta x}{\Delta t}\right)^2\right| + \text{constant} \tag{4.16}$$

Equation (4.15) represents a set of straight lines with a slope 2, and Eq. (4.16) a set of curves whose slopes are closer to 1 as C_0 is small and increases to 3 as C_0 increases. The curves in both Figs. 6 and 7 have these slopes.

VIII. Conclusion

The established error approximations are observed as lower bounds to the errors of finite difference schemes. More importantly, they relate the errors to both the equations and the conditions of a problem; hence, they may be used as tools to derive problem-oriented schemes. We are now looking into the derivation of such schemes.

The applications of the error approximation are not as limited as they appear to be. The computations are also not as hard and expansive as imagined. The assumption that the solution of the problem is band limited will be satisfied by the discrete Fourier series of a function. In a finite interval, the discrete Fourier series of a continuous or a piecewise continuous function is always possible. Furthermore, the Fourier series of the sine or the cosine function is simple; hence, the evaluation of the error approximation is really easy. Nevertheless, the technique of fast Fourier transform [10] could further reduce the computation cost. Hence, this approach is worthy of exploration.

Our observations suggest that the errors of finite difference schemes are functions of the wave speed and that in certain regions, a lower order scheme may generate a more accurate numerical solution. This suggests the need to solve wave propagation problems by using more than one scheme in parallel. In a multischeme computation, the numerical solution forms a network. If a numerical switch is possible at every node of this network, then a mixed solution will be more accurate.

An important consequence of this study is that the traveling paths of the numerical and the exact waves are different. If the propagating speed of a wave varies with position, then a point on the exact wave will travel along a curve, but on the numerical wave straight lines, which are tangent lines to the path of the exact wave. Hence, a delay (or an advance) of the numerical wave to arrive at a fixed point relative to the exact wave will occur. This puts doubt in the conventional pointwise error measurement technique. Hence, there is a need to define a new error measurement technique to replace the existing pointwise measurement.

APPENDIX 1. The Finite Difference Schemes

1. *Constant Speed*: $s = C_0$

(a) First order:

$$U_n^{j+1} = U_n^j(1 - 2C_0^2\lambda^2) + U_{n+1}^j\left(C_0^2\lambda^2 - \frac{C_0}{2}\lambda\right) + \left(C_0^2\lambda^2 + \frac{C_0}{2}\lambda\right)U_{n-1}^j, \tag{A1.1}$$

where $\lambda = \Delta t/\Delta x$.

The stability region is

$$C_0\lambda \leqslant \tfrac{1}{2}. \tag{A1.2}$$

(b) Second order:

$$U_n^{j+1} = U_n^j(1 - C_0^2\lambda^2) + U_{n+1}^j \cdot \tfrac{1}{2}(C_0^2\lambda^2 - C_0\lambda) + U_{n-1}^j \cdot \tfrac{1}{2}(C_0^2\lambda^2 + C_0\lambda). \tag{A1.3}$$

The stability region is

$$C_0\lambda \leqslant 1. \tag{A1.4}$$

2. *Function of time*: $s = C_0C_2(t)$

(a) First order:

$$\begin{aligned} U_n^{j+1} = U_n^j[1 - 2C_0^2\lambda^2C_2^2(t_j)] + U_{n+1}^j\left[C_0^2\lambda^2C_2^2(t_j) - \frac{C_0\lambda}{2}\left(C_2(t_j) + \frac{dC_2}{dt}\Delta t\right)\right] \\ + U_{n-1}^j\left[C_0^2\lambda^2C_2^2(t_j) + \frac{C_0\lambda}{2}\left(C_2(t_j) + \frac{dC_2}{dt}\Delta t\right)\right]. \end{aligned} \tag{A1.5}$$

The stability region is

$$C_0\lambda \leqslant 1/(2 \cdot \max_{0 \leqslant t \leqslant T} |C_2(t)|). \tag{A1.6}$$

(b) Second order:

$$\begin{aligned} U_n^{j+1} = U_n^j(1 - C_0^2\lambda^2C_2^2(t_j)) + \frac{1}{2}\left[C_0^2\lambda^2C_2^2(t_j) - C_0\lambda\left(C_2(t_j) + \frac{dC_2}{dt}\frac{\Delta t}{2}\right)\right]U_{n+1}^j \\ + \frac{1}{2}\left[C_0^2\lambda^2C_2^2(t_j) + C_0\lambda\left(C_2(t_j) + \frac{dC_2}{dt}\frac{\Delta t}{2}\right)\right]U_{n-1}^j. \end{aligned} \tag{A1.7}$$

The stability region is

$$C_0\lambda \leqslant 1/\max_{0 \leqslant t \leqslant T} |C_2(t)|. \tag{A1.8}$$

3. *Function of Space*: $s = C_0 C_1(x)$

(a) First order:

$$\begin{aligned} U_n^{j+1} &= (1 - 2C_0^2 \lambda^2 C_1^2(x_n))\, U_n^j \\ &+ \left[C_0^2 \lambda^2 C_1^2(x_n) - \frac{\lambda C_0}{2} C_1(x_n) \left(1 - C_0\, \Delta t \frac{dC_1}{dx}\right)\right] U_{n+1}^j \\ &+ \left[C_0^2 \lambda^2 C_1^2(x_n) + \frac{\lambda C_0}{2} C_1(x_n) \left(1 - C_0\, \Delta t \frac{dC_1}{dx}\right)\right] U_{n-1}^j. \end{aligned} \tag{A1.9}$$

The stability region is

$$C_0 \lambda \leqslant 1/(2 \cdot \max |C_1(x)|). \tag{A1.10}$$

(b) Second order:

$$\begin{aligned} U_n^{j+1} &= \frac{1}{2}\left[\lambda^2 C_0^2 C_1^2(x_n) - \lambda C_0 C_1(x_n) \left(1 - \frac{C_0 \Delta t}{2} \frac{dC_1}{dx}\right)\right] U_{n+1}^j \\ &+ (1 - \lambda^2 C_0^2 C_1^2(x_n))\, U_n^j \\ &+ \frac{1}{2}\left[\lambda^2 C_0^2 C_1^2(x_n) + \lambda C_0 C_1(x_n) \left(1 - \frac{C_0\, \Delta t}{2} \frac{dC_1}{dx}\right)\right] U_{n-1}^j. \end{aligned} \tag{A1.11}$$

The stability region is

$$C_0 \lambda \leqslant 1/\max |C_1(x)|. \tag{A1.12}$$

APPENDIX 2.
Semi-discretized Schemes and Numerical Solutions

The semi-discretized schemes considered are

$$\tilde{A}_1 \cdot = -C_1(x)\, C_2(t) \frac{\tilde{E}\cdot - \tilde{E}\cdot^{-1}}{2\, \Delta x}, \tag{A2.1}$$

$$\tilde{A}_2 \cdot = C_1(x)\, C_2(t) \frac{\tilde{E}\cdot^2 - 6\tilde{E}\cdot + 3\tilde{I}\cdot + 2\tilde{E}\cdot^{-1}}{6\, \Delta x}, \tag{A2.2}$$

and

$$\tilde{A}_3 \cdot = C_1(x)\, C_2(t) \frac{\tilde{E}\cdot^2 - 8\tilde{E}\cdot + 8\tilde{E}\cdot^{-1} - \tilde{E}\cdot^{-2}}{12\, \Delta x}, \tag{A2.3}$$

where $\tilde{I}\cdot$ is the identity operator and $\tilde{E}\cdot$ the shift operator. As defined, $\tilde{A}_1\cdot$ is first order, $\tilde{A}_2\cdot$ second order, and $\tilde{A}_3\cdot$ third order.

Let $U_{h1}(x_n, t)$, $U_{h2}(x_n, t)$, and $U_{h3}(x_n, t)$ be the numerical approximations at $x = x_n$ with respect to the schemes $\tilde{A}_1\cdot$, $\tilde{A}_2\cdot$, and $\tilde{A}_3\cdot$, respectively. Then at $t = T$, the error is computed as

$$|E^{(j)}_{Ti}(x_n)| = |U^{(j)}(x_n, T) - U^{(j)}_{hi}(x_n, T)| \qquad \text{for} \quad i = 1, 2, 3 \tag{A2.4}$$

where $j = 1$ indicates the continuous wave, and $j = 3$ the discontinuous one.

We give the numerical approximations for the following examples.

Case 1. The numerical continuous wave in the *first problem*:

$$U^{(1)}_{h1}(x_n, t) = \sin\left[k\pi\left(x_n - C_0 t \frac{\sin k\pi\,\Delta x}{k\pi\,\Delta x}\right)\right], \tag{A2.5}$$

$$U^{(1)}_{h2}(x_n, t) = \exp\left(\frac{4C_0 t}{3\,\Delta x}\sin^4\frac{k\pi\,\Delta x}{2}\right)\sin\left[k\pi x_n - \frac{4C_0 t}{3\,\Delta x}\sin\frac{k\pi\,\Delta x}{2}\right.$$
$$\left.\times\cos\frac{k\pi\,\Delta x}{2}\left(\frac{3}{2} + \sin^2\frac{k\pi\,\Delta x}{2}\right)\right], \tag{A2.6}$$

$$U^{(1)}_{h3}(x_n, t) = \sin\left[k\pi x_n - \frac{C_0 t}{3\,\Delta x}(\sin k\pi\,\Delta x)(4 - \cos k\pi\,\Delta x)\right]. \tag{A2.7}$$

Case 2. The numerical discontinuous wave in the *second problem*:

$$U^{(3)}_{h1}(x_n, t) = \frac{1}{2} + \frac{1}{2N}\sum_{k=-N/2}^{N/2-1}\left\{\frac{\cos\dfrac{2k+1}{2N}\pi}{\sin\dfrac{2k+1}{2N}\pi}\right.$$
$$\times\sin\left[\frac{2k+1}{N}n\pi - \frac{C_0 t^2}{2\,\Delta x}\sin\left(\frac{2k+1}{N}\pi\right)\right]$$
$$\left.+\cos\left[\frac{2k+1}{N}n\pi - \frac{C_0 t^2}{2\,\Delta x}\sin\left(\frac{2k+1}{N}\pi\right)\right]\right\}, \tag{A.2.8}$$

$$U^{(3)}_{h2}(x_n, t) = \frac{1}{2} + \frac{1}{2N}\sum_{k=-N/2}^{N/2-1}\exp\left(\frac{2C_0 t^2}{3\,\Delta x}\sin^4\left(\frac{2k+1}{2N}\pi\right)\right)$$
$$\times\left\{\frac{\cos\dfrac{2k+1}{2N}\pi}{\sin\dfrac{2k+1}{2N}\pi}\sin\left[\frac{2k+1}{N}n\pi - \frac{C_0 t^2}{3\,\Delta x}\sin\left(\frac{2k+1}{2N}\pi\right)\right.\right.$$

$$\times \cos\left(\frac{2k+1}{2N}\pi\right)\left(3+2\sin^2\left(\frac{2k+1}{2N}\pi\right)\right)\Bigg]$$

$$+\cos\Bigg[\frac{2k+1}{N}n\pi-\frac{C_0t^2}{3\,\Delta x}\sin\left(\frac{2k+1}{2N}\pi\right)$$

$$\times \cos\left(\frac{2k+1}{2N}\pi\right)\left(3+2\sin^2\left(\frac{2k+1}{2N}\pi\right)\right)\Bigg]\Bigg\} \qquad \text{(A2.9)}$$

$$U_{h3}^{(3)}(x_n,t)=\frac{1}{2}+\frac{1}{2N}\sum_{k=-N/2}^{N/2-1}\left\{\frac{\cos\dfrac{2k+1}{2N}\pi}{\sin\dfrac{2k+1}{2N}\pi}\right.$$

$$\times\sin\left[\frac{2k+1}{N}n\pi-\frac{C_0t^2}{3\,\Delta x}\sin\left(\frac{2k+1}{2N}\pi\right)\cos\left(\frac{2k+1}{2N}\pi\right)\right.$$

$$\left.\times\left(4-\cos\left(\frac{2k+1}{N}\pi\right)\right)\right]+\cos\Bigg[\frac{2k+1}{N}\pi n-\frac{C_0t^2}{3\,\Delta x}$$

$$\left.\times\sin\left(\frac{2k+1}{2N}\pi\right)\cos\left(\frac{2k+1}{2N}\pi\right)\left(4-\cos\left(\frac{2k+1}{N}\pi\right)\right)\Bigg]\right\}. \qquad \text{(A2.10)}$$

REFERENCES

1. W. F. AMES, *Numerical Methods for Partial Differential Equations* (Academic Press, New York, 1977).
2. MATS Y. T. APELKRANS, *Math. Comput.* **22**, 525 (1968).
3. MATS Y. T. APELKRANS, Report 15A, Upsala University, Department of Computer Science, 1969 (unpublished).
4. P. BRENNER AND V. THOMEE, *Math. Sci.* **29**, 329 (1971).
5. L. L. CAMPBELL, *SIAM J. Appl. Math.* **12**, No. 1, 117 (1964).
6. L. L. CAMPBELL, *SIAM J. Appl. Math.* **16**, No. 3, 626 (1968).
7. Y. F. CHIANG, Ph.D. thesis, Rutgers University, 1981 (unpublished).
8. Y. F. CHIANG, CIS Research Report 6, NJIT, 1983 (unpublished).
9. Y. F. CHIANG, CIS Research Report 7, NJIT, 1985 (unpublished).
10. J. W. COOLEY AND J. W. TUKEY, *Math. Comput.* **19**, 297 (1965).
11. P. R. GARABEDIAN, *Partial Differential Equations* (Wiley, New York, 1964).
12. W. M. GENTLEMAN AND G. SANDE, *Proceedings, Spring Joint Comput. Conf. AFIPS* (1966).
13. A. HARTEN, AEC Research and Development Report, NYU, COO-3077-50, 1974 (unpublished).
14. H. O. KRESS AND J. OLIGER, World of Meteorological Organizational International Council of Scientific Unions (1973).
15. A. MAJDA AND S. OSHER, *Commun. Pure Appl. Math.* **30**, 671 (1977).
16. M. S. MOCK AND P. D. LAX, *Commun. Pure Appl. Math.* **31**, 423 (1978).

17. S. A. ORSZAG, *J. Fluid Mech.* **49**, 75 (1971).
18. G. R. RICHTER AND S. R. FALK, *Advances in Computer Methods for Partial Differential Equations*, V, 297 (1984).
19. R. D. RICHTMYER AND K. W. MORTON, *Difference Methods for Initial Value Problems* (Interscience, New York, 1967).
20. D. SLEPIAN AND H. O. POLLAK, *Bell Syst. Tech. J.* **40**, 43 (1961).
21. G. A. SOD, *J. Comput. Phys.* **27**, 1 (1978).
22. C. N. TREFETHEN, *SIAM Rev.* **24**, 113 (1982).
23. R. VICHNEVETSKY AND J. B. BOWLES, *Fourier Analysis of Numerical Approximations of Hyperbolic Equations* (SIAM, Philadelphia, 1982).
24. R. VICHNEVETSKY, *Computer Methods for Partial Differential Equations* (Prentice-Hall, Englewood Cliffs, NJ, 1981).
25. J. VON NEUMANN AND R. D. RICHTMYER, *J. Appl. Phys.* **21**, 232 (1950).
26. J. M. WHITTAKER, *Proc. Edinburgh Math. Soc.* **1**, 169 (1929).