Computers & Fluids Vol. 14, No. 4, pp. 393–410, 1986
Printed in Great Britain

0045-7930/86 $3.00 + 0.00
Pergamon Journals Ltd

STABILITY AND PHASE SPEED FOR VARIOUS FINITE ELEMENT FORMULATIONS OF THE ADVECTION EQUATION

BENY NETA† and R. T. WILLIAMS
Naval Postgraduate School, Monterey, CA 93943, U.S.A.

(*Received* 8 *October* 1985; *in revised form* 5 *March* 1986)

Abstract—This paper analyzes the stability and accuracy of various finite element approximations to the linearized two-dimensional advection equation. Four triangular elements with linear basis functions are included along with a rectangular element with bilinear basis functions. In addition, second- and fourth-order finite difference schemes are examined for comparison. Time is discretized with the leapfrog method. The criss-cross triangle formulation is found to be unstable. The best schemes are the isosceles triangles with linear basis functions and the rectangles with bilinear basis functions.

1. INTRODUCTION

Advective processes are dominant in atmospheric and oceanic circulation systems, while diffusive effects are important only in boundary layer regions. Any numerical model for these circulation systems should treat advective effects accurately. In this paper we analyze various finite element formulations of the linearized advection equation in two dimensions, which can be written:

$$\frac{\partial F}{\partial t} + V\cos\theta\frac{\partial F}{\partial x} + V\sin\theta\frac{\partial F}{\partial y} = 0, \tag{1.1}$$

where V is the mean flow speed and θ is the wind direction relative to the x-axis. The quantity F should be interpreted as vorticity or temperature, for example. The analytic solution to (1.1) is

$$F(x, y, t) = F(x - tV\cos\theta, y - tV\sin\theta, 0). \tag{1.2}$$

The schemes considered in this study employ leapfrog time differencing and Galerkin finite element spatial representations. The leapfrog time differencing will tend to increase the phase speeds relative to the exact time variation. Linear elements on the following triangles are treated: isosceles (Fig. 1), biased (Fig. 2), criss-crossed (Fig. 3), and unbiased (Fig. 4). Bilinear basis functions on rectangles (Fig. 5) are also examined. Our results are compared with second- and fourth-order finite differences (see Ref. [1]). The computational stability conditions are derived for each scheme and the computational phase speed is compared with the exact value given by (1.2). The eigensolutions are also provided for comparison. Swartz and Wendroff [2] extended the results of Kreiss and Oliger [3] and Douglas [4]. They compare the efficiency, as defined in [3] and [4], of finite elements and finite differences for the approximation of $F_t = cF_x$.

In Section 2 the schemes are analyzed for the special case where the mean flow is along the x-axis ($\theta = 0$). In Section 4, the isosceles triangles and rectangles are compared for arbitrary θ. In Sections 2 and 4 cyclic boundary conditions are used in both x and y. In Section 3, the case $\theta = 0$, is again considered, but for flow in a channel. The conclusions are given in Section 5.

†Permanent address: Texas Tech University, Department of Mathematics, Lubbock, TX 79409, U.S.A.

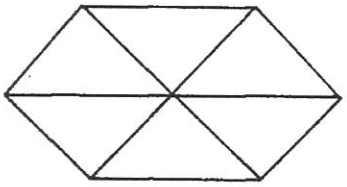

Fig. 1. Isosceles triangles.

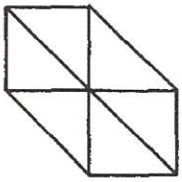

Fig. 2. Biased grid, $\Delta x = \Delta y = d$.

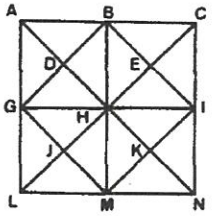

Fig. 3. Criss-cross grid, $\Delta x = \Delta y = d$.

2. ANALYSIS FOR $\theta = 0$

If the basis functions are given by $\{\phi_i(x, y)\}$, the Galerkin approximation to (1.1) becomes

$$\sum_j \dot{F}_j(t) \int \phi_i \phi_j \, dA + V \sum_j F_j(t) \int \phi_i \left(\cos\theta \frac{\partial \phi_j}{\partial x} + \sin\theta \frac{\partial \phi_j}{\partial y} \right) dA = 0, \tag{2.1}$$

for each i. The following expressions for integration over triangles can be found in Zienkiewicz [5]:

$$\int_T \phi_i \phi_j \, dA = \begin{cases} A/6 & i = j \\ A/12 & i \neq j \end{cases} \tag{2.2}$$

$$\int_T \phi_i \frac{\partial \phi_j}{\partial x} \, dA = b_j/6, \tag{2.3}$$

$$\int_T \phi_i \frac{\partial \phi_j}{\partial y} \, dA = a_j/6, \tag{2.4}$$

where T is a triangular element, A is the area of T and the a_j and b_j are defined by

$$a_1 = x_3 - x_2, \quad b_1 = y_2 - y_3,$$
$$a_2 = x_1 - x_3, \quad b_2 = y_3 - y_1,$$
$$a_3 = x_2 - x_1, \quad b_3 = y_1 - y_2.$$

The vertices of the triangle (x_j, y_j) are numbered counter-clockwise. These relations will be used in the first four cases of this section.

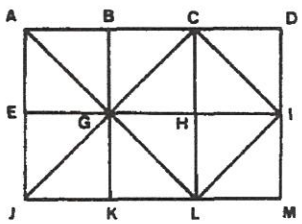

Fig. 4. Unbiased grid, $\Delta x = \Delta y = d$.

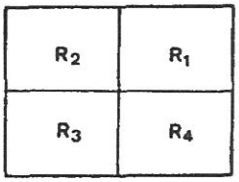

Fig. 5. Bilinear rectangular elements.

In this section (2.1) is simplified by setting $\theta = 0$, which eliminates the last term in (2.1). It is convenient to set

$$V = c \tag{2.5}$$

since V is the analytic phase speed in this case. The initial condition for this section, which is periodic in x and y, is given by

$$F(x, y, 0) = K\, e^{i(\mu x + \nu y)}, \tag{2.6}$$

so that according to (1.2) the analytic solution is

$$F(x, y, t) = K\, e^{i(\mu x + \nu y - \mu c t)}. \tag{2.7}$$

2.1 *Isosceles triangles*

In this case there are six triangles having a vertex in common, as indicated in Fig. 1. The approximate solution at the vertices of the triangles may be obtained by solving the following first-order ordinary differential equation (suppressing the t variable),

$$\begin{aligned}
&\dot{F}(x,y) + \frac{1}{6}\Big\{\dot{F}(x+\Delta x, y) + \dot{F}(x-\Delta x, y) + \dot{F}\Big(x+\frac{\Delta x}{2}, y+\Delta y\Big) \\
&\quad + \dot{F}\Big(x-\frac{\Delta x}{2}, y+\Delta y\Big) + \dot{F}\Big(x+\frac{\Delta x}{2}, y-\Delta y\Big) + \dot{F}\Big(x-\frac{\Delta x}{2}, y-\Delta y\Big)\Big\} \\
&\quad + \frac{1}{3} c \frac{1}{\Delta x}\Big\{2[F(x+\Delta x, y) - F(x-\Delta x, y)] + F\Big(x+\frac{\Delta x}{2}, y+\Delta y\Big) \\
&\quad - F\Big(x-\frac{\Delta x}{2}, y+\Delta y\Big) + F\Big(x+\frac{\Delta x}{2}, y-\Delta y\Big) - F\Big(x-\frac{\Delta x}{2}, y-\Delta y\Big)\Big\} = 0.
\end{aligned} \tag{2.8}$$

The solution to (2.8) should be of the form

$$F(x, y, t) = A(t)\, e^{i(\mu x + \nu y)}, \tag{2.9}$$

where the amplitude $A(t)$ satisfies

$$\dot{A} + i\frac{\sigma}{\Delta t} A = 0, \tag{2.10}$$

and

$$\sigma = 4c\frac{\Delta t}{\Delta x} \frac{\sin \mu \Delta x + \sin \mu \dfrac{\Delta x}{2} \cos \nu \Delta y}{3 + \cos \mu \Delta x + 2 \cos \mu \dfrac{\Delta x}{2} \cos \nu \Delta y}. \tag{2.11}$$

Note that the sign of σ does not depend on the sign of ν.

The solution of (2.10) is approximated by the leapfrog scheme

$$\frac{A^{n+1} - A^{n-1}}{2\Delta t} + i\frac{\sigma}{\Delta t} A^n = 0, \tag{2.12}$$

where A^n is the solution at $n\Delta t$. Let

$$A^n = A_0 \lambda^n, \tag{2.13}$$

then

$$\lambda^2 + 2i\sigma\lambda - 1 = 0,$$

and

$$\lambda = -i\sigma \pm \sqrt{(1 - \sigma^2)}. \tag{2.14}$$

In order for the method to be stable, the value of λ should satisfy

$$|\lambda| \leq 1. \tag{2.15}$$

This implies that

$$c\frac{\Delta t}{\Delta x} \leq 0.45144. \tag{2.16}$$

In this case the two values of λ may be written in polar form as

$$\lambda_1 = e^{-i\alpha}, \quad \lambda_2 = e^{i(\alpha+\pi)}, \tag{2.17}$$

where

$$\alpha = \arcsin \sigma. \tag{2.18}$$

The complete discrete solution F_D then takes the form

$$F_D = [Me^{-i\alpha n} + E\, e^{i(\pi+\alpha)n}]\, e^{i(\mu x+\nu y)}, \tag{2.19}$$

where M, E are arbitrary constants such that (by the initial condition)

$$M + E = K.$$

Thus,

$$F_D = (K-E)\, e^{i\mu(x-\alpha n/\mu+\nu y/\mu)} + (-1)^n E\, e^{i\mu(x+\alpha n/\mu+\nu y/\mu)}. \tag{2.20}$$

Note that the approximate solution consists of two waves or modes instead of one in the true solution. The second term, describes a wave propagating in a direction opposite to that of the true solution. This is called a computational mode. The constant in (2.20) will be specified by prescribing the method for making the first time-step (see e.g. Ref. [1], p. 113).

The phase speed C_F of the numerical solution is defined by

$$C_F = \frac{\alpha}{\mu\Delta t} \tag{2.21}$$

(see e.g. Ref. [1]).

The phase speed will be computed for various values of $\mu\Delta x$ and $\nu\Delta y$ and it will be compared to the values obtained by other discretizations. The graphs will appear at the end of the section.

2.2 *Biased*

The biased elements are one possible element when the nodal points are arranged in a rectangular fashion. These elements are shown in Fig. 2.

When the formulae (2.2)–(2.4) are used for the elements in Fig. 2, eqn (2.1) becomes

$$\begin{aligned}\dot{F}(x,y) &+ \tfrac{1}{6}\{\dot{F}(x+d,y)+\dot{F}(x-d,y)+\dot{F}(x,y+d)+\dot{F}(x-d,y+d)+\dot{F}(x,y-d)\\ &+\dot{F}(x+d,y-d)\}+\tfrac{1}{3}c\frac{1}{d}\{2[F(x+d,y)-F(x-d,y)]+F(x,y+d)\\ &-F(x-d,y+d)+F(x+d,y-d)-F(x,y-d)\}=0,\end{aligned} \tag{2.22}$$

where $\Delta x = \Delta y = d$. The solution of this equation is of the form (2.19) where σ is now given by

$$\sigma = 4c\frac{\Delta t}{d}\frac{\sin\mu d+\frac{1}{2}\sin\nu d+\frac{1}{2}\sin(\mu-\nu)d}{3+\cos\mu d+\cos\nu d+\cos(\mu-\nu)d}. \tag{2.23}$$

Note that σ depends on the sign of ν. The condition of stability is

$$c\frac{\Delta t}{d} \leq 0.456. \tag{2.24}$$

2.3 *Criss-cross*

The criss-cross formulation was proposed by Fix [6] and it represents another breakdown from a rectangular array of points. In this case and the next, two types of points involved. Type I points (like point H in Fig. 3) are common to eight triangles and type II points (like D) are common to only four triangles. As a result the equations for each type are different.

The eqn (2.1) at type I points is

$$\dot{F}(x,y)+\frac{1}{8}\left\{\dot{F}(x+d,y)+\dot{F}(x-d,y)+\dot{F}\left(x+\frac{d}{2},y+\frac{d}{2}\right)+\dot{F}\left(x+\frac{d}{2},y-\frac{d}{2}\right)\right.$$
$$\left.+\dot{F}\left(x-\frac{d}{2},y+\frac{d}{2}\right)+\dot{F}\left(x-\frac{d}{2},y-\frac{d}{2}\right)+\dot{F}(x,y+d)+\dot{F}(x,y-d)\right\}$$
$$+\tfrac{1}{4}c\frac{1}{d}\left\{F(x+d,y)-F(x-d,y)+F\left(x+\frac{d}{2},y+\frac{d}{2}\right)-F\left(x-\frac{d}{2},y-\frac{d}{2}\right)\right.$$
$$\left.+F\left(x+\frac{d}{2},y-\frac{d}{2}\right)-F\left(x-\frac{d}{2},y+\frac{d}{2}\right)\right\}=0, \quad (2.25)$$

and at type II points it becomes

$$\dot{F}\left(x-\frac{d}{2},y+\frac{d}{2}\right)+\tfrac{1}{4}\{\dot{F}(x,y)+\dot{F}(x,y+d)+\dot{F}(x-d,y+d)+\dot{F}(x-d,y)\}$$
$$+\tfrac{1}{2}c\frac{1}{d}\{F(x,y+d)-F(x-d,y+d)+F(x,y)-F(x-d,y)\}=0. \quad (2.26)$$

To solve these equations we try

$$F_{\mathrm{I}}=A\mathrm{e}^{\mathrm{i}(\mu x+\nu y)},$$
$$F_{\mathrm{II}}=B\mathrm{e}^{\mathrm{i}(\mu x+\nu y)}. \quad (2.27)$$

Substituting in (2.25) and (2.26) yields two ordinary differential equations for the unknowns $A(t)$, $B(t)$,

$$\dot{A}\{4+\cos\mu d+\cos\nu d\}+\dot{B}\cos\mu\frac{d}{2}\cos\nu\frac{d}{2}$$
$$+2ci\frac{1}{d}A\sin\mu d+2ci\frac{1}{d}B\sin\mu\frac{d}{2}\cos\nu\frac{d}{2}=0, \quad (2.28a)$$

$$\dot{A}\cos\mu d\cos\nu d+\dot{B}+2ci\frac{1}{d}A\sin\mu d\cos\nu d=0. \quad (2.28b)$$

Using the leapfrog scheme to approximate the solution, and letting

$$A^n=A_0\mathrm{e}^{in\alpha\Delta t},\quad B^n=B_0\mathrm{e}^{in\alpha\Delta t}, \quad (2.29)$$

one obtains

$$\sigma=2c\frac{\Delta t}{d}\left[\frac{-b\pm\sqrt{(b^2-4ae)}}{2a}\right] \quad (2.30)$$

where

$$\sigma=\sin\alpha\Delta t,$$
$$a=4+\cos\mu d+\cos\nu d-\cos\mu d\cos\mu\frac{d}{2}\cos\nu d\cos\nu\frac{d}{2},$$
$$b=\sin\mu d-\sin(\tfrac{3}{2}\mu d)\cos\nu d\cos\nu\frac{d}{2},$$
$$e=-\sin\mu d\sin\mu\frac{d}{2}\cos\nu d\cos\nu\frac{d}{2}. \quad (2.31)$$

It can be shown numerically that b^2-4ae is sometimes negative and thus σ can be complex. Thus the method is unstable. This result shows the danger of having two overlapping grids with a large difference in point spacing.

2.4 *Unbiased*

The unbiased elements, which are given in Fig. 4, represent a modification of the biased elements given in Fig. 2. These elements have symmetry in the x-direction when viewed in blocks such as A–C–L–J in the figure below. In this case we also have two types of points: type I points like G and type II points like H. As before eqn (2.1) is replaced by two ordinary differential equations for F_{I} and F_{II}:

$$\begin{aligned}\dot{F}_{\mathrm{I}}(x,y)&+\tfrac{1}{8}\{\dot{F}_{\mathrm{II}}(x+d,y)+\dot{F}_{\mathrm{II}}(x-d,y)+\dot{F}_{\mathrm{I}}(x+d,y+d)+\dot{F}_{\mathrm{I}}(x-d,y+d)\\&+\dot{F}_{\mathrm{I}}(x+d,y-d)+\dot{F}_{\mathrm{I}}(x-d,y-d)+\dot{F}_{\mathrm{II}}(x,y+d)+\dot{F}_{\mathrm{II}}(x,y-d)\}\\&+\tfrac{1}{4}c\frac{1}{d}\{2[F_{\mathrm{II}}(x+d,y)-F_{\mathrm{II}}(x-d,y)]+F_{\mathrm{I}}(x+d,y+d)-F_{\mathrm{I}}(x-d,y+d)\\&+F_{\mathrm{I}}(x+d,y-d)-F_{\mathrm{I}}(x-d,y-d)\}=0 \qquad (2.32)\end{aligned}$$

$$\begin{aligned}\dot{F}_{\mathrm{II}}(x+d,y)&+\tfrac{1}{4}\{\dot{F}_{\mathrm{I}}(x+2d,y)+\dot{F}_{\mathrm{I}}(x+d,y+d)+\dot{F}_{\mathrm{I}}(x,y)+\dot{F}_{\mathrm{I}}(x+d,y-d)\}\\&+c\{F_{\mathrm{I}}(x+2d,y)-F_{\mathrm{I}}(x,y)\}=0. \qquad (2.33)\end{aligned}$$

Now introduce (2.27) into (2.32) and (2.33), which gives

$$\dot{A}\{4+2\cos\mu d\cos\nu d\}+\dot{B}\{\cos\mu d+\cos\nu d\}+4ci\frac{1}{d}A\sin\mu d\cos\nu d$$

$$+4ci\frac{1}{d}B\sin\mu d=0, \qquad (2.34)$$

$$\dot{A}\{\cos\mu d+\cos\nu d\}+2\dot{B}+4ci\frac{1}{d}A\sin\mu d=0. \qquad (2.35)$$

The solution of this system is obtained as in the previous subsection,

$$\sigma_{\pm}=\sin\alpha_{\pm}\Delta t=4c\frac{\Delta t}{d}\frac{\sin\mu d\,[\cos\mu d\pm\sqrt{(8+2\cos\mu d\cos\nu d-\cos^2\nu d)}]}{8+2\cos\mu d\cos\nu d-\cos^2\mu d-\cos^2\nu d}. \qquad (2.36)$$

The solution for F_{I} and F_{II} can now be written

$$F_{\mathrm{I}}=\mathrm{e}^{i(\mu x+\nu y)}\{C_1\mathrm{e}^{in\alpha_-\Delta t}+C_2(-1)^n\mathrm{e}^{-in\alpha_-\Delta t}+C_3\mathrm{e}^{in\alpha_+\Delta t}+C_4(-1)^n\mathrm{e}^{-in\alpha_+\Delta t}\}, \qquad (2.37)$$

$$F_{\mathrm{II}}=\mathrm{e}^{i(\mu x+\nu y)}\{R_+C_1\mathrm{e}^{in\alpha_-\Delta t}+R_+C_2(-1)^n\mathrm{e}^{-in\alpha_-\Delta t}+R_-C_3\mathrm{e}^{in\alpha_+\Delta t}+R_-C_4(-1)^n\mathrm{e}^{-in\alpha_+\Delta t}\}, \qquad (2.38)$$

where

$$R_{\pm}=-\frac{\sigma_{\pm}(\cos\mu d+\cos\nu d)+4c\dfrac{\Delta t}{d}\sin\mu d}{2\sigma_{\pm}}.$$

The constants C_i depend on the initial solution and the method for making the first time step. If one uses the forward approximation (see e.g. Ref. [1], p. 113), then it can be shown that

$$C_1=\frac{K-R_+}{R_--R_+}\frac{1-ic\dfrac{\Delta t}{d}\sin\mu d+\mathrm{e}^{i\alpha_-\Delta t}}{2\cos\alpha_-\Delta t},$$

$$C_2=\frac{K-R_+}{R_--R_+}\frac{1-ic\dfrac{\Delta t}{d}\sin\mu d-\mathrm{e}^{i\alpha_-\Delta t}}{2\cos\alpha_-\Delta t},$$

$$C_3 = \frac{R_- - K}{R_- - R_+} \frac{1 - ic\dfrac{\Delta t}{d}\sin \mu d + e^{-i\alpha_+ \Delta t}}{2\cos \alpha_+ \Delta t},$$

$$C_4 = \frac{R_- - K}{R_- - R_+} \frac{1 - ic\dfrac{\Delta t}{d}\sin \mu d - e^{i\alpha_+ \Delta t}}{2\cos \alpha_+ \Delta t}. \tag{2.39}$$

The first terms in (2.37) and (2.38) represent the physical mode and the remaining three terms are computational modes. Note that the previously considered stable schemes only have one computational mode.

2.5 *Bilinear rectangular elements*

In this case one cannot use formulae (2.2)–(2.4), but similar expressions can be developed. Let the common vertex of all rectangles with sides $2a$ and $2b$ be in the origin in Fig. 5.

It is easy to show that the entries of the mass matrix are independent of the rectangle and

$$\iint_{R_i} \phi_1^2 \, dx \, dy = \frac{\Delta x \, \Delta y}{9},$$

$$\iint_{R_i} \phi_1 \phi_2 \, dx \, dy = \iint_{R_i} \phi_1 \phi_4 \, dx \, dy = \frac{\Delta x \, \Delta y}{18},$$

$$\iint_{R_i} \phi_1 \phi_3 \, dx \, dy = \frac{\Delta x \, \Delta y}{36}, \tag{2.40}$$

where ϕ_i are bilinear basis functions. The integrals containing $\partial \phi_i / \partial x$ are easily computed and we quote here the results for R_1. The values for other rectangles are just a permutation of these values.

$$\iint_{R_1} \phi_1 \frac{\partial \phi_1}{\partial x} \, dx \, dy = -\iint_{R_1} \phi_1 \frac{\partial \phi_2}{\partial x} \, dx \, dy = -\frac{\Delta y}{6},$$

$$\iint_{R_1} \phi_1 \frac{\partial \phi_3}{\partial x} \, dx \, dy = -\iint_{R_1} \phi_1 \frac{\partial \phi_4}{\partial x} \, dx \, dy = \frac{\Delta y}{12}. \tag{2.41}$$

With these basis functions, eqn (2.1) becomes

$$\begin{aligned} &\dot{F}(x,y) + \tfrac{1}{4}\{\dot{F}(x,y+\Delta y) + \dot{F}(x,y-\Delta y) + \dot{F}(x-\Delta x,y) + \dot{F}(x+\Delta x,y)\} \\ &\quad + \tfrac{1}{16}\{\dot{F}(x-\Delta x,y+\Delta y) + \dot{F}(x+\Delta x,y+\Delta y) + \dot{F}(x-\Delta x,y-\Delta y) \\ &\quad + \dot{F}(x+\Delta x,y-\Delta y)\} + \tfrac{3}{16} c \frac{1}{\Delta x}\{F(x+\Delta x,y+\Delta y) - F(x-\Delta x,y+\Delta y) \\ &\quad + F(x+\Delta x,y-\Delta y) - F(x-\Delta x,y-\Delta y) + 4[F(x+\Delta x,y) \\ &\quad - F(x-\Delta x,y)]\} = 0. \end{aligned} \tag{2.42}$$

Proceeding as before,

$$F(x,y) = A(t) e^{i(\mu x + \nu y)} \tag{2.43}$$

then

$$\dot{A}\{2 + \cos \mu \Delta x\} + 3ci \frac{1}{\Delta x} A \sin \mu \Delta x = 0. \tag{2.44}$$

Using the leapfrog scheme and letting $A^n = A(n\Delta t) = A_0^n \lambda$, one obtains λ as in (2.14) where

$$\sigma = 3c \frac{\Delta t}{\Delta x} \frac{\sin \mu \Delta x}{2 + \cos \mu \Delta x}. \tag{2.45}$$

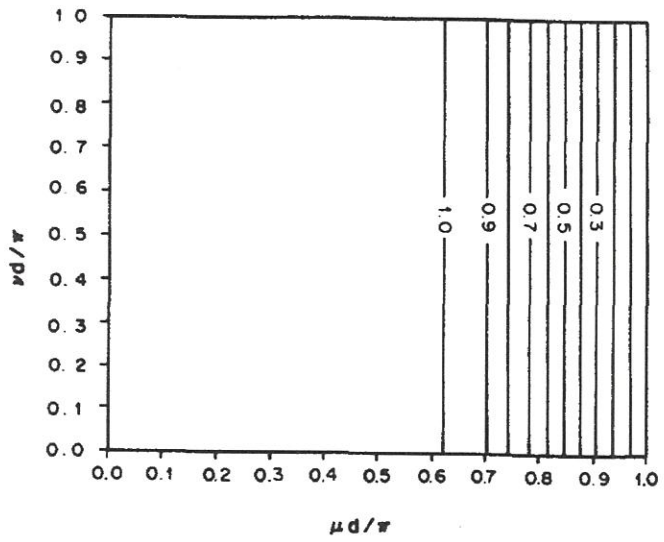

Fig. 6. Ratio C_F/c for bilinear rectangular elements.

Note that this is the only case where σ is independent of the y-wavenumber, ν. The condition for stability is $|\sigma| \leq 1$ which implies (see Ref. [7], p. 202),

$$c\frac{\Delta t}{\Delta x} \leq \frac{1}{\sqrt{3}}. \tag{2.46}$$

For comparison we quote the results in [1] for second and fourth order finite differences. For second order (p. 114)

$$\sigma = c\frac{\Delta t}{\Delta x}\sin \mu \Delta x, \tag{2.47}$$

$$c\frac{\Delta t}{\Delta x} \leq 1. \tag{2.48}$$

For fourth order (p. 136)

$$\sigma = c\frac{\Delta t}{\Delta x}\left(\tfrac{4}{3}\sin \mu \Delta x - \tfrac{1}{6}\sin 2\mu \Delta x\right) \tag{2.49}$$

$$c\frac{\Delta t}{\Delta x} \leq 0.73. \tag{2.50}$$

In order to compare the accuracy of the schemes treated in this section, the various formulae will be evaluated numerically as functions of μd and νd. The time-step for each scheme is selected so that the computational stability parameter will be 80% of the critical value. This procedure is frequently used in numerical prediction models. Figure 6 contains the ratio C_F/c for bilinear rectangular elements from (2.18), (2.21) and (2.45). In this case the phase speed is independent of ν, but the two-dimensional diagram is retained for comparison with the schemes which are obtained for triangular elements. The scheme shows excellent accuracy except for the smallest scales where the phase speed decreases to zero for shortest resolvable wave.

Figure 7 contains C_F/c for the isosceles triangular elements from (2.18), (2.21) and (2.11). With this scheme the phase speed is a function of the y-wavenumber ν, but it is an even function so that only the upper half plane is shown. Comparison of Figs 6 and 7 shows that the isosceles elements give much better phase speeds than the rectangular elements when the y-wavenumber ν is small, similar results were found by Hinsman [7] with the shallow water equations. However, for $\nu d/\pi = 1$ the rectangular elements give better phase speeds than the triangles, which in the upper right-hand corner give negative phase speeds. The difference in behavior between the two schemes can be explained by comparing (2.8) with (2.42). When the fields are independent of y ($\nu = 0$) the approximation for $\partial F/\partial x$ will

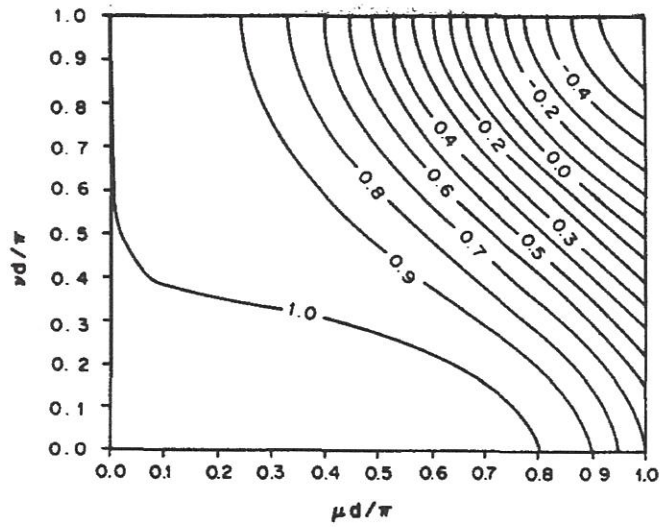

Fig. 7. Ratio C_F/c for isosceles triangular elements.

be better for the isosceles triangles because (2.8) contains terms of the form $F(x+\Delta x/2, y+\Delta y) - F(x-\Delta x/2, y+\Delta y)$ whereas the corresponding terms in (2.42) are $F(x-\Delta x, y+\Delta y) - F(x-\Delta x, y+\Delta y)$. If ν is not small the improvement in $\partial F/\partial x$ by terms at $y \pm \Delta y$ disappears.

Figure 8 contains C_F/c for the biased triangular elements from (2.18), (2.21) and (2.23). In this case the solutions are not symmetric in ν because of the biased form of the elements. Comparison of Fig. 8 with Figs 6 and 7 shows that along $\nu = 0$ the biased triangles are

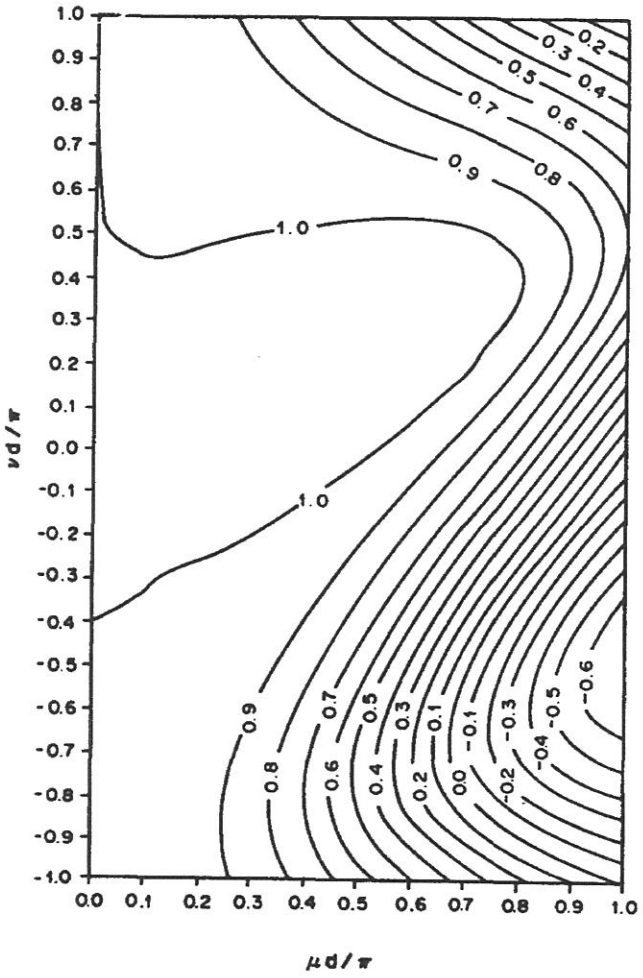

Fig. 8. Ratio C_F/c for biased triangular elements.

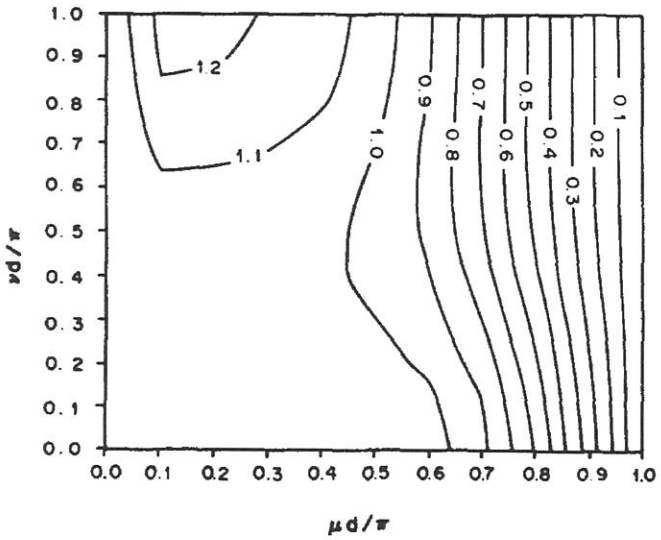

Fig. 9. Ratio C_F/c for unbiased triangular elements.

clearly inferior to the isosceles triangles and slightly inferior to the rectangles. Along the line $v = \mu$ the biased elements are clearly superior to the isosceles triangles, while along the line $v = -\mu$ the biased elements are very poor. The behavior of the biased elements can be explained in terms of the relation of the wave orientation to the element bias.

Figure 9 contains C_F/c for the unbiased triangular elements from (2.18) (2.21) and (2.36). As a result of the unbiased formulation, the phase speed is independent of the sign of v. Comparison of Figs 6 and 9 shows that the unbiased triangles give nearly the accuracy of the rectangles except in the upper left-hand corner where the unbiased triangles over-predict the phase speed. The major disadvantages of this scheme are the presence of extra computational modes and the possibility that each node can have different amplitudes on the two grids. It can be seen from (2.37) and (2.38) that the ratio of the amplitudes of the physical modes on the two grids is given by R_-. Figure 10 gives R_- as a function of μd and vd. For small values of μd and vd the R_- is close to 1 so that the solutions on the two grids have about the same amplitude, as would be expected for large-scale fields. For larger values of μd and vd, R_- differs from 1.0 so that the solution on the two grids will have different amplitudes. In this case the total solution will have grid scale noise of amplitude [R_ − 1]. An examination of (2.39) shows that the amplitude

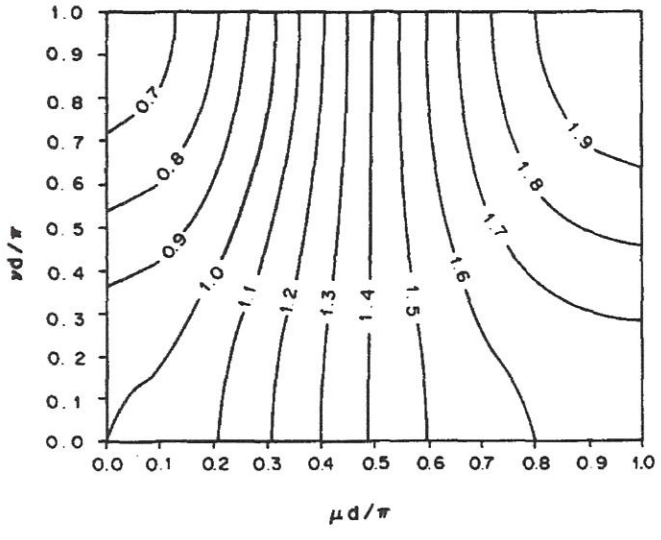

Fig. 10. R_- as a function of μd and vd.

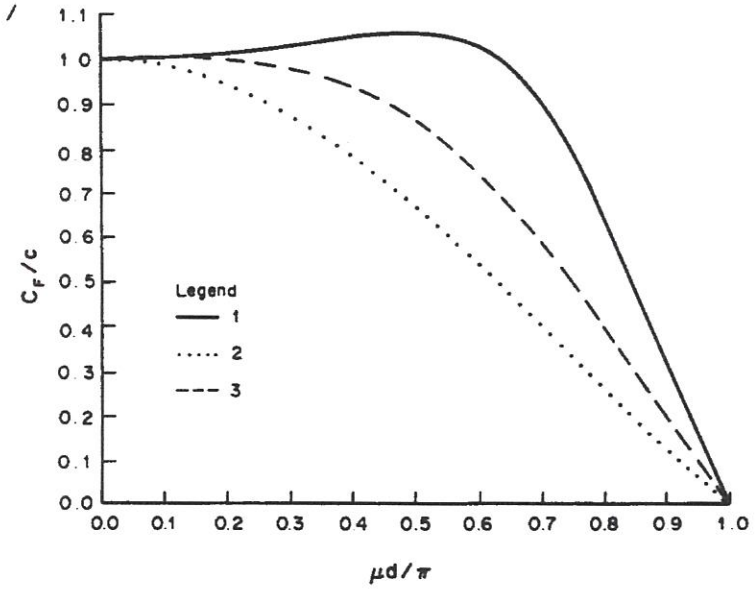

Fig. 11. Ratio C_F/c for bilinear rectangular elements (1), second-order (2) and fourth-order (3) finite differences.

of the computational modes C_2, C_3, and C_4 are small for large-scale waves (small μd and vd) and they become important for small-scale waves.

Figure 11 gives C_F/c for the following schemes: (1) bilinear rectangular elements from (2.18), (2.21) and (2.45), (2) second-order finite difference from (2.18) (2.21) and (2.47), (3) fourth-order finite difference from (2.18), (2.21) amd (2.19). With those schemes the solutions are independent of v. The finite element scheme is superior to both of the finite difference schemes except at $\mu d = \pi$.

The results of this section show that the scheme with isosceles triangle elements is superior to all of the schemes which use right triangles for elements. For small y wavenumber the triangular basis functions give better accuracy than the rectangular basis functions, but the rectangles became superior for the larger y wavenumbers. In addition, the rectangles give phase speeds which are independent of v, so that there will be no computational dispersion in y. The fourth-order finite difference scheme is better than the second order scheme and they are both inferior to finite elements.

3. FLOW IN A CHANNEL

In this section we consider flow in a channel between walls at $y = 0, W$. The mean flow is along the x-axis so that again $\theta = 0$. The boundary conditions are

$$F(x, 0, t) = F(x, W, t) = 0, \tag{3.1}$$

which according to (1.1) will be maintained if they are satisfied initially. The analytic solution with these boundary conditions can be obtained by combining (2.7) with $+v$ and $-v$ exponents, which gives

$$F = \frac{2K}{i}\left[e^{i(\mu x + vy - \mu ct)} - e^{i(\mu x - vy - \mu ct)}\right] = K \sin vy\, e^{i\mu(x - ct)}. \tag{3.2}$$

The boundary conditions are then satisfied by

$$v = \frac{m\pi}{W}, \tag{3.3}$$

where m is a positive integer.

The finite element and finite difference schemes can be combined in the manner of (3.2) to satisfy the boundary conditions as long as the phase speed C_F is independent of the sign of v. This independence condition is satisfied by all of the schemes which have symmetric dependence on points in y. All of the schemes treated in Section 2 are of this type except for the biased triangles (see 2.23). The solutions for biased triangles with the boundary condition (3.1) will be derived in this section.

In order to include the boundary conditions we let

$$F(x, y, t) = A(y, t)\, e^{i\mu x}, \tag{3.4}$$

where

$$A(y, t) = 0 \quad y = 0, W. \tag{3.5}$$

Upon substitution into (2.22) and simplification, one obtains the following differential equation for $A(y, t)$:

$$\dot{A}(y-\Delta y)e^{i\mu\Delta x/2}\cos\mu\Delta x/2 + \dot{A}(y)(3+\cos\mu\Delta x) + \dot{A}(y+\Delta y)e^{-i\mu\Delta x/2}\cos\mu\Delta x/2$$

$$+ 2ic\frac{1}{\Delta x}[A(y-\Delta y)e^{i\mu\Delta x/2}\sin\mu\Delta x/2 + 4A(y)\sin\mu\Delta x/2\cos\mu\Delta x/2$$

$$+ A(y+\Delta y)e^{-i\mu\Delta x/2}\sin\mu\Delta x/2] = 0, \quad 0 < y < W. \tag{3.6}$$

Now let

$$\begin{aligned} a &= e^{i\mu\Delta x/2}, \\ b &= a\cos\mu\Delta x/2, \\ f &= 3+\cos\mu\Delta x, \\ d &= 4\cos\mu\Delta x/2 \text{ and} \\ e &= 2\sin\mu\Delta x/2; \end{aligned} \tag{3.7}$$

then (3.6) can be written as a tridiagonal system of ordinary differential equations.

$$\begin{bmatrix} f & b^* & & \\ b & f & b^* & \\ & & b & f \end{bmatrix} \begin{bmatrix} \dot{A}(\Delta y) \\ \vdots \\ \dot{A}(W-\Delta y) \end{bmatrix} + ic/\Delta x \begin{bmatrix} ed & ea^* & & \\ ea & ed & ea^* & \\ & & ea & ed \end{bmatrix} \begin{bmatrix} A(\Delta y) \\ \vdots \\ A(W-\Delta y) \end{bmatrix} = 0, \tag{3.8}$$

where * indicates the complex conjugate. To obtain the eigenvalues and eigenfunctions of this system we let

$$A(y, t) = e^{i\lambda ct/\Delta x}B(y). \tag{3.9}$$

Then

$$\begin{bmatrix} \lambda f+ed & \lambda b^*+ea^* & & \\ \lambda b+ea & \lambda f+ed & \lambda b^*+ea & \\ & & \lambda b+ea & \lambda f+ed \end{bmatrix} \begin{bmatrix} B(\Delta y) \\ \vdots \\ B(W-\Delta y) \end{bmatrix} = 0. \tag{3.10}$$

Let $D(\alpha, \beta, \gamma)$ be the determinant of the coefficient matrix in (3.10). Then generalizing a result of Ford and Ruttan [8] we can show that the eigenvalues λ_k and eigenvectors $\underline{B}_k$ are given by

$$\lambda_k = 4\sin\mu\frac{\Delta x}{2}\frac{-2\cos\mu\dfrac{\Delta x}{2}[r^2-f] + r\left[f-4\cos^2\mu\dfrac{\Delta x}{2}\right]}{4r^2\cos^2\mu\dfrac{\Delta x}{2}-f^2}, \quad k = 1, 2, \ldots, n \tag{3.11}$$

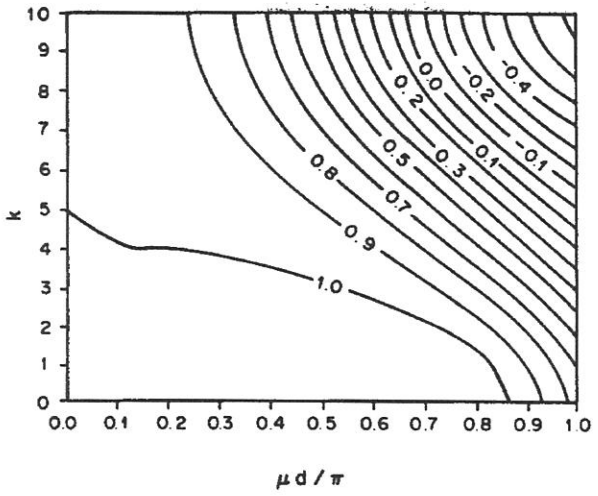

Fig. 12. Ratio C_F/c for biased triangular elements in a channel.

$$(\underline{B}_k)_j = a^j \sin\frac{kj\pi}{n+1}, \quad k = 1, 2, \ldots, n, \tag{3.12}$$

where

$$r = \cos\frac{k\pi}{n+1}. \tag{3.13}$$

The eigenfunction (3.12) tilts in y due to the presence of the term a^j, as can be seen by substitution into (3.4). If the leapfrog time differences are introduced, it can be shown that

$$C_F = -\frac{1}{\mu\Delta t}\arcsin\left[\frac{\Delta t}{\Delta x}c\lambda\right]. \tag{3.14}$$

The computational stability condition is given by

$$\left|\frac{c\Delta t}{\Delta x}\right| \leq 1. \tag{3.15}$$

Figure 12 gives the ratio C_F/c from (3.14). When compared with the plane wave solutions in Fig. 8, the phase speeds in Fig. 12 are much better along $k = 0$. However, if the initial condition is of the form

$$F(x, y, 0) = K \sin\frac{k\pi y}{W} e^{i\mu x}, \tag{3.16}$$

it will be projected on all of the eigenfunctions (3.12). Since these modes move at different phase speeds, The initial disturbance will become distorted in y. This will not happen with the other schemes considered in the paper, because their eigenfunctions do not tilt for flow in a channel.

4. TWO DIMENSIONS

In this section we consider the more general case where $\theta \neq 0$ so that the $\partial F/\partial y$ term in eqn (1.1.) will not be zero. Only isosceles triangles and bilinear rectangular elements are compared since the other elements are inferior to these. Clearly, one now requires the

evaluation of the integral (2.4). With the initial condition (2.6), the exact solution (1.2) becomes

$$F(x, y, t) = Ke^{i[\mu\Delta x + \nu y - (\mu V \cos\theta + \nu V \sin\theta)t]}. \tag{4.1}$$

Using isosceles triangles whose base angle is also θ one has

$$\sigma = 4V \cos\theta \frac{\Delta t}{\Delta x} \frac{3 \sin \nu\Delta y \cos \mu \frac{\Delta x}{2} + \sin \mu\Delta x + \cos \nu\Delta y \sin \mu \frac{\Delta x}{2}}{3 + \cos \mu\Delta x + 2 \cos \mu \frac{\Delta x}{2} \cos \nu\Delta y}, \tag{4.2}$$

and the approximate solution is

$$F_D = (K - E)e^{i(\mu x + \nu y - \alpha n\Delta t)} + (-1)^n Ee^{i(\mu x + \nu y + \alpha n\Delta t)}, \tag{4.3}$$

where $\alpha\Delta t = \arcsin\sigma$.

The relative phase speed is now given by

$$\frac{C_F}{c} = \frac{\alpha}{\mu V \cos\theta + \nu V \sin\theta} = \frac{\arcsin\sigma}{2\frac{\Delta t}{\Delta x} V \cos\theta \left(\mu \frac{\Delta x}{2} + \nu\Delta y\right)}, \tag{4.4}$$

with the use of $\sin\theta = 2(\Delta y/\Delta x)\cos\theta$. For the rectangles, the results are

$$\sigma = 3V \cos\theta \frac{\Delta t}{\Delta x}\left[\frac{\sin \mu\Delta x}{2 + \cos \mu\Delta x} + \frac{\sin \nu\Delta y}{2 + \cos \nu\Delta y}\right], \tag{4.5}$$

and where θ is the angle between the x-axis and the diagonal, and the phase speed becomes

$$\frac{C_F}{c} = \frac{\arcsin\sigma}{V \cos\theta \frac{\Delta t}{\Delta x}(\mu\Delta x + \nu\Delta y)}. \tag{4.6}$$

It is known that $c\Delta t/\Delta x \leq 1/\sqrt{6}$ for both finite elements (see Cullen [9]). Clearly now the rectangles do not have the advantage mentioned previously with respect to the time-step. The corresponding formulae for finite differences are easily obtained from symmetry and (2.47), (2.49). For second order, one has

$$\sigma = V \frac{\Delta t}{\Delta x} \cos\theta \,(\sin \mu\Delta x + \sin \nu\Delta y), \tag{4.7}$$

$$\frac{C_F}{c} = \frac{\arcsin\sigma}{V \frac{\Delta t}{\Delta x} \cos\theta \,(\mu\Delta x + \nu\Delta y)}, \tag{4.8}$$

and for the fourth order,

$$\sigma = V \frac{\Delta t}{\Delta x} \cos\theta \,[\tfrac{4}{3}(\sin \mu\Delta x + \sin \nu\Delta y) - \tfrac{1}{6}(\sin 2\mu\Delta x + \sin 2\nu\Delta y)], \tag{4.9}$$

where again θ is the angle between the x-axis and the diagonal.

The ratio C_F/c for the isosceles triangle elements from (4.4) is presented in Fig. 13 for $\theta = 45°$. Figure 14 contains this ratio for rectangular elements from (4.6). The phase speeds for the second- and fourth-order finite difference schemes [see (4.7)–(4.9)] are given in Figs 15 and 16, respectively. All four schemes give the best phase speeds along the line $\mu = \nu$, where the wave phase lines are normal to the wind vector. With this orientation, the

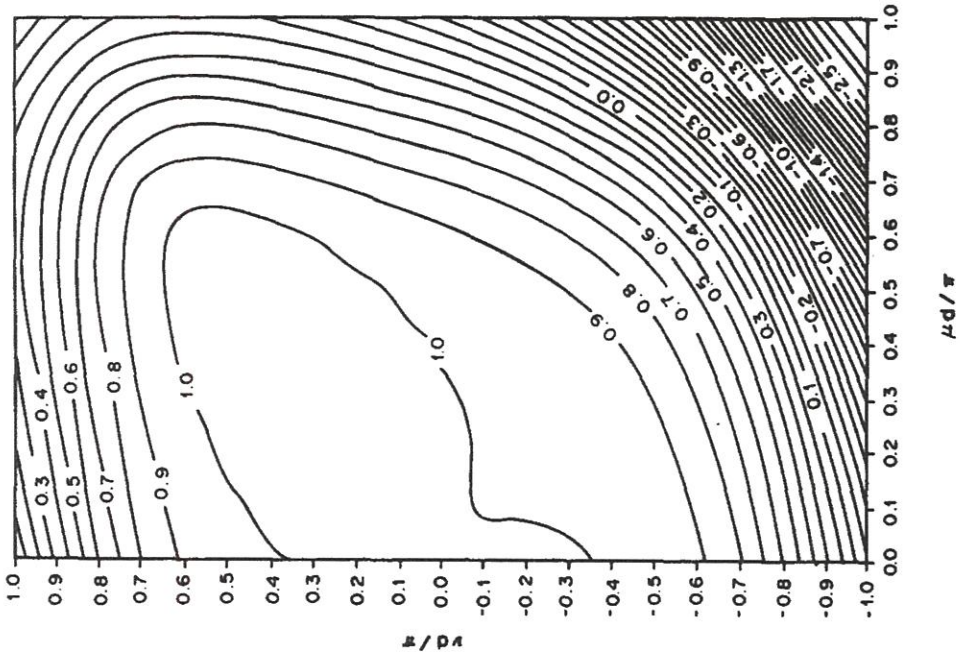

Fig. 14. Ratio C_f/c for bilinear rectangular elements in two dimensions.

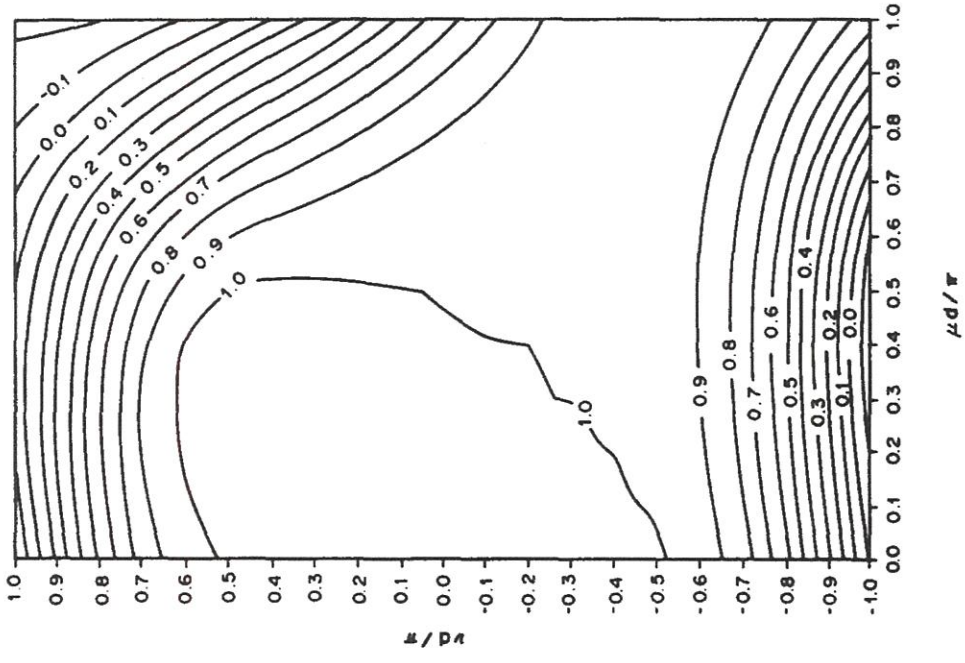

Fig. 13. Ratio C_f/c for isosceles triangular elements in two dimensions.

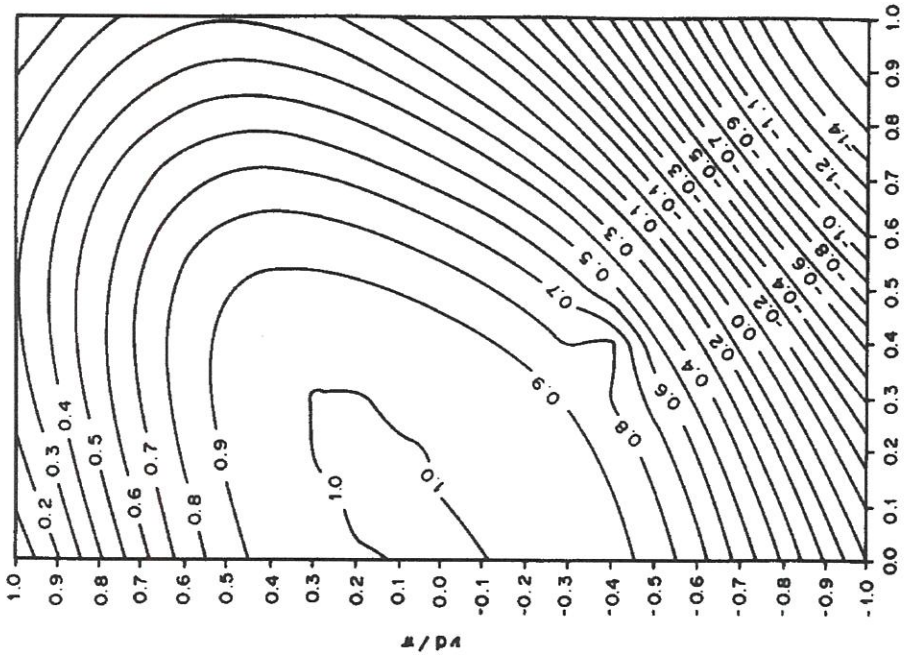

Fig. 16. Ratio C_p/c for fourth-order finite differences in two dimensions.

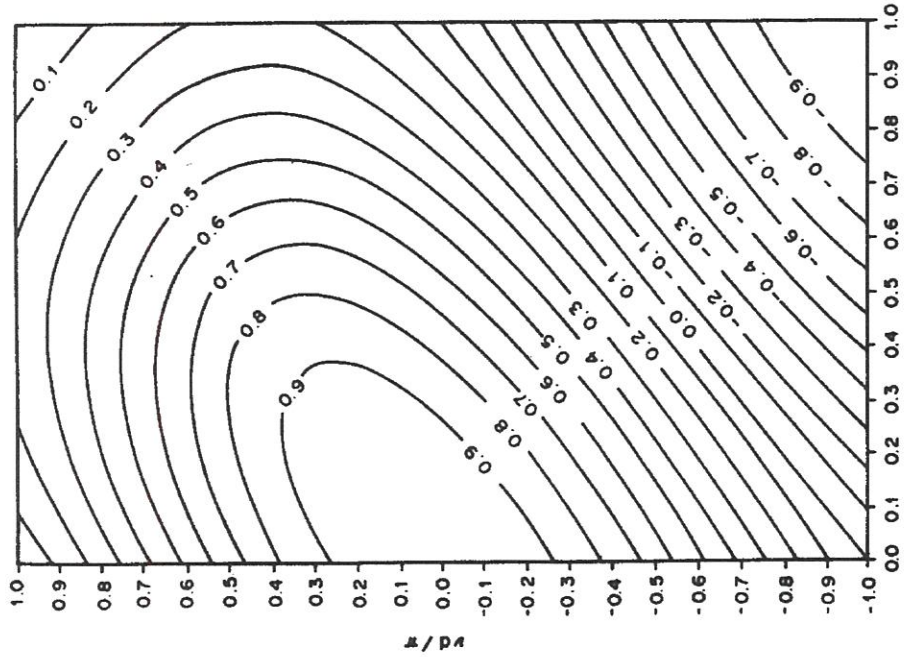

Fig. 15. Ratio C_p/c for second-order finite differences in two dimensions.

rectangular elements and the finite difference schemes can calculate the space derivative in the direction of wave propagation more accurately, because this direction is across the diagonal of the grid. Comparison with the earlier solutions for $\theta = 0$. (Fig. 11) shows that the current results along $\mu = \nu$ are more accurate as a function of $(\mu^2 + \nu^2)^{1/2}$ than those in Fig. 11 as a function of μ. The phase speeds along $\nu = -\mu$ appear to be poorer, but according to (4.1) the phase speed along this line is zero so there is actually no error. Figures 14–16 are generally similar, with the rectangular finite element being the best and the second-order finite difference the worst. The isosceles triangle finite element scheme (Fig. 13) does not do as well as the rectangular finite element scheme (Fig. 14) along the line $\nu = \mu$.

5. SUMMARY AND CONCLUSIONS

This paper has examined various finite element approximations to the linearized two-dimensional advection equation. Four triangular elements with linear basis functions were included, along with a rectangular element with bilinear basis functions. In addition, second- and fourth-order finite difference schemes were examined for comparison. Leapfrog time differences were used in all cases, although the results could easily be generalized to other time differencing schemes.

The various finite element and finite difference schemes were compared for the case where the mean flow was directed along the x-axis. The finite element formulation which is based on isosceles triangles was clearly superior to the formulations which are based on right triangles that are obtained from a rectangular array of nodal points. Staniforth (private communication) pointed out that if the mean flow is along the y-axis (i.e. $\theta = 90°$) then the phase speed when $\nu = 0$ becomes equal to the phase speed for the rectangles. The biased triangles produced a highly distorted phase speed field as would be expected from their non-symmetric form. The criss-cross arrangement was unstable, perhaps due to the juxtaposition of elements of different size. The unbiased elements, which are obtained by reversing alternate biased triangles, were found to give poorer phase speeds than the isosceles triangles. The analysis of the unbiased scheme required the separation of the nodal points into two types. The resulting solutions had three computational modes as compared with one for most of the other schemes. The major disadvantages of this scheme are the presence of extra computational modes and the likelihood that each mode will have a different amplitude on each of the two grids. The latter effect will manifest itself as small-scale noise. The phase speed for the rectangular finite element scheme was found to be independent of the y-wavenumber. The isosceles triangle scheme gave better phase speeds than the rectangle scheme for low y-wavenumbers, but the situation reversed for the higher y-wavenumbers. The finite difference schemes gave similar but poorer behavior compared to the rectangular finite elements scheme with the second-order finite difference scheme being the poorest.

The schemes were also examined for flow in a channel and it was found that the solutions for each scheme could be combined to satisfy the wall boundary conditions. The only exception to this was the biased triangle scheme, for which a separate solution was obtained. The eigenfunctions were found to tilt in the same sense as the elements.

The isosceles triangles and the rectangles were also compared for plane waves with the mean flow at a 45° angle to the x-axis. It was found that the rectangles were best for waves whose phase lines were normal to the mean wind direction, but considering all wave orientations there was no great difference between the isosceles triangles and the rectangles.

The results of this paper show little difference in overall accuracy between the finite element formulation with isosceles triangles or rectangles. If the resolution is uniform the isosceles triangles may have a small advantage. However, if the resolution varies appreciably, the triangles will be changed so that they are no longer isosceles; then the good properties of the isosceles triangles would be lost. On the other hand, the rectangles can retain their shape when the resolution is changed (see Staniforth and Mitchell [10]. Also, with rectangles the equations can be solved more efficiently, as has been discussed by Staniforth [11].

Since eqn (1.1) is hyperbolic, it is expected that many of the results from this paper will hold for other hyperbolic systems, such as the shallow water equations. For example, Kelley and Williams [12] found considerable small-scale noise in a study of flow in a channel with the shallow water equations. They used the unbiased triangular elements with non-staggered basis functions. Schoenstadt [13] and Williams [14] showed that the use of the shallow water equations with non-staggered elements can lead to small-scale noise. From our present analysis it appears that the noise would be much worse due to the extra computational modes with the unbiased elements.

Acknowledgements—The authors would like to thank Professor C.-S. Liou for many helpful discussions, and Dr A. N. Staniforth for his careful review of the paper. The first author was supported by a National Research Council Associateship. The NRC Associate Program is sponsored by the Naval Postgraduate School Foundation Research Program which is funded by the Chief of Navy Material. This research was also supported by the Naval Air Systems Command through the Naval Environmental Prediction Research Facility. The expert typing of this manuscript by Penny Jones was greatly appreciated. The numerical computations were performed by the W. R. Church Computer Center.

REFERENCES

1. G. J. Haltiner and R. T. Williams, *Numerical Prediction and Dynamic Meteorology*. John Wiley, New York (1980).
2. B. Swartz and B. Wendroff, The relative efficiency of finite difference and finite element methods I: hyperbolic problems and splines. *SIAM J. numer. Anal.* **11**, 979–993 (1974).
3. H.-O. Kreiss and J. Oliger, Comparison of accurate methods for the integration of hyperbolic equations. *Tellus* **24**, 199–215 (1972).
4. J. Douglas, A survey of numerical methods for partial differential equations. *Advances in Computers*, Vol. 2, pp. 1–54. Academic Press, New York (1961).
5. O. C. Zienkiewicz, *The Finite Element Method in Engineering Science*. McGraw-Hill, New York (1977).
6. G. J. Fix, On the structure of the errors made in mixed finite element method. *Nonlinear PDEs in Engineering and Applied Science* (Edited by R. L. Sternberg, A. J. Kalinowski and J. S. Papadakis), pp. 53–85. Marcel Dekker, New York (1980).
7. D. E. Hinsman, Numerical simulation of atmospheric flow on variable grids using the Galerkin finite element method. Ph.D. thesis, Naval Postgraduate School, Department of Meteorology, Monterey, CA (1983).
8. W. T. Ford and A. Ruttan, Eigenvalues of tri-diagonal Toeplitz matrices. *Proc. 2nd Regional Conf. Numer. Anal.* (Edited by B. Neta), Texas Tech University, Lubbock, 23–25 September. INTT Rep. 26 (1983).
9. M. J. P. Cullen, A simple finite element method for meteorological problems. *J. Inst. math Applices.* **11**, 15–31 (1973).
10. A. N. Staniforth and H. L. Mitchell, A variable resolution finite-element technique for regional forecasting with the primitive equations. *Mon. Weath. Rev.* **106**, 439–447 (1978).
11. A. N. Staniforth, On the formulation of efficient finite-element codes for flows in regular domains: some considerations. *Int. J. numer. Meths Fluids* (in press).
12. R. G. Kelley and R. T. Williams, A finite element prediction model with variable element sizes. Naval Postgraduate School Rep. NPS-63 Wu76101 (1976).
13. A. L. Schoenstadt, A transfer function analysis of numerical schemes used to simulate geostrophic adjustment. *Mon. Weath. Rev.* **108**, 1248–1259 (1980).
14. R. T. Williams, On the formulation of finite-element prediction models. *Mon. Weath. Rev.* **109**, 463–466 (1981).