JOURNAL OF COMPUTATIONAL PHYSICS 15, 251–265 (1974)

Phase Error and Stability of Second Order Methods for Hyperbolic Problems. II

DAVID GOTTLIEB*

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

AND

ELI TURKEL

Mathematical Applications Group Inc.

Received October 30, 1973

Two new second order schemes are presented for the solution of hyperbolic systems in two space dimensions. The first method is a modification of the Lax–Wendreff method that dramatically reduces the numerical phase error compared with all other schemes that do not use data beyond a nine point lattice. The second scheme is a one parameter generalization of the rectangular form of Richtmyer's method. The scheme is shown to be stable (the stability proof is only formal since it is a nonlinear scheme) for all symmetric hyperbolic equations. The phase error and allowable time steps are functions of the free parameter. At one extreme the parameter can be chosen so as to yield maximal allowable time steps but with a phase error that is large compared with other second order methods. Alternately we can choose the free parameter so that the phase error is smaller than that of the Richtmyer type schemes but at the cost of a smaller permissible time step. Numerical experiments with the equations of dynamic elasticity are presented that confirm these conclusions.

1. INTRODUCTION

In [7] many second order schemes were considered for the equation

$$w_t + Aw_x + Bw_y = 0, \tag{1}$$

where A and B are simultaneously symmetrizable, or in divergence free form

$$w_t + f_x + g_y = 0. \tag{1'}$$

* Partially supported by the Office of Scientific Research of the U.S. Air Force, Grant No. AF-AFOSR-44620-71-C-0110.

These schemes were compared with respect to stability and phase error properties. When we limit the schemes to those that do not use mesh data that extend beyond a nine point rectangular lattice then the Leapfrog and Lax–Wendroff methods together with a nine point splitting scheme have the smallest phase errors. For many of the methods considered a small phase error was coupled with a small permissible time step. As an example a family of schemes was introduced that generalizes the rotated Richtmyer and Burstein [11] schemes. As we increase the free parameter α the phase error decreases but the stability properties of the scheme also decrease. In fact except for the rotated Richtmyer scheme all members of this family (including the Burstein scheme) are weakly unstable for particular choices of the matrices A and B appearing in Eq. (1) even for symmetric hyperbolic systems (see [7]).

In this paper we shall present a modification of the Lax–Wendroff method which greatly decreases the phase error without using data beyond a nine point lattice. This is of importance for problems where higher order schemes (which must use data beyond the nine point lattice) are difficult to use because of curved boundaries. We shall also investigate a new family of schemes for stability properties and phase errors. The family is linearly stable but requires a small time step and has a poor phase representation. However, by considering a nonlinear variation of these schemes one can choose the parameter to either decrease the phase error or, conversely, increase the time step. In contrast to the Burstein type scheme this family appears to be stable for all matrices A and B that can be simultaneously symmetrized. However, due to the nonlinearity of the scheme only a formal presentation of this stability can be presented together with some numerical results.

2. Improved Lax–Wendroff Method

The phase error for a numerical method is defined only when the matrices A and B commute. In this case we have

$$\begin{aligned} E(\xi, \eta) &= \text{numerical phase} - \text{analytic phase} \\ &= \arctan((\operatorname{Im} G)(\operatorname{Re} G)^{-1}) + (\lambda A\xi + \sigma B\eta), \end{aligned} \tag{2}$$

where $G(\xi, \eta)$ is the amplification matrix associated with the scheme and where $\lambda = \Delta\tau/\Delta x$, $\sigma = \Delta\tau/\Delta y$. As discussed in [7] the phase error is of importance only when the Fourier variables ξ and η are small. Thus, we neglect terms involving the Fourier variables in Eq. (2) of fourth order and higher.

For the Lax–Wendroff method [5] we have

$$\begin{aligned} E(\xi, \eta) &= \tfrac{1}{6}[\lambda A\xi^3 + \sigma B\eta^3 - (\lambda A\xi + \sigma B\eta)^3] + O(\xi^4 + \eta^4) \\ &= \tfrac{1}{6}[\lambda A(1 - (\lambda A)^2)\, \xi^3 + \sigma B(1 - (\sigma B)^2)\, \eta^3 \\ &\quad - 3\lambda\sigma^2 AB^2\xi\eta^2 - 3\lambda^2\sigma A^2B\xi^2\eta] + O(\xi^4 + \eta^4). \end{aligned} \tag{3}$$

The stability criterion for this method is $\rho(\lambda A, \sigma B) \leqslant 1/(8)^{1/2}$, where $\rho(A, B)$ is the larger of the spectral radii of A and B. It is readily seen that if $\lambda\rho(A) + \sigma\rho(B) \leqslant 1$ and ξ and η are positive then E is positive, i.e., there is a phase lag. This seems to be the smallest phase error thus far achieved for schemes that do not use data beyond a nine point rectangular lattice. Should one wish to change the coefficient of the ξ^3 term in the expansion of the phase error then we must use a one-dimensional operator that requires more than three mesh points at the previous time step. In particular if we wish $E(\xi, \eta)$ to be of fourth order for all ξ and η then we choose a higher order method which entails using mesh points beyond the nine point lattice. However, it is possible to force $E(\xi, \eta)$ to be of fourth order for particular values of the Fourier variables ξ, η. A logical choice is to minimize the phase error when $\xi = \eta$. When $\xi = \eta$ the phase error for the Lax–Wendroff method becomes (assuming as before that A and B commute)

$$E(\xi, \xi) = (\xi^3/6)[\lambda A(1 - (\lambda A)^2 - 3(\sigma B)^2) + \sigma B(1 - (\sigma B)^2 - 3(\lambda A)^2)] + O(\xi^4). \quad (4)$$

As previously indicated within a nine point lattice we cannot introduce terms whose Fourier transform can be expanded with a leading term of the form ξ^3, η^3 for small ξ, η. However, by considering mixed derivatives we can introduce terms whose Fourier transform begins with terms of the form $\xi\eta^2$ or $\xi^2\eta$ for ξ, η small. Thus, we are able to add terms to the basic Lax–Wendroff method which will decrease the phase error and will even be of third order for the particular case of $\xi = \eta$. Hence, we consider the scheme

$$\begin{aligned} w^{n+1} = {} & Lw^n + \delta_x\mu_x\delta_y{}^2(1 - (\lambda A)^2 - 3(\sigma B)^2)(\lambda A w^n/6) + \delta_y\mu_y\delta_x{}^2(1 - (\sigma B)^2 \\ & - 3(\lambda A)^2)(\sigma B w^n/6), \end{aligned} \quad (5)$$

where L is the Lax–Wendroff operator, μ is an averaging operator, and δ is a central difference operator (both operators defined over half meshes). As with the original Lax–Wendroff method it is possible to add fourth order terms to increase the permissible time step. We note that the sign of the third order terms introduced in Eq. (5) is opposite that of the odd order stabilizers discussed by Eilon, Gottlieb, and Zwas [2], when $\rho(\lambda A, \sigma B) \leqslant \frac{1}{2}$. This is in agreement with our previous observation that decreasing the phase error frequently also decreases the allowable time step.

When the matrices A and B are constant the amplification matrix for the difference scheme (5) is

$$\begin{aligned} G(\xi, \eta) = {} & I - 2i[\alpha(1 - \alpha^2)^{1/2}(1 + K_1\beta^2)\lambda A + \beta(1 - \beta^2)^{1/2}(1 + K_2\alpha^2)\sigma B] \\ & - 2[(\lambda A)^2\alpha^2 + (\sigma B)^2\beta^2 + (AB + BA)\alpha\beta(1 - \alpha^2)^{1/2}(1 - \beta^2)^{1/2}], \end{aligned} \quad (6)$$

where

$$K_1 = \tfrac{2}{3}(1 - \lambda^2 A^2 - 3\sigma^2 B^2) \qquad K_2 = \tfrac{2}{3}(1 - \sigma^2 B^2 - 3\lambda^2 A^2) \tag{7}$$

and

$$\alpha = \sin(\xi/2) \qquad \beta = \sin(\eta/2).$$

The phase error for this scheme is

$$E(\xi, \eta) = \tfrac{1}{6}[\lambda A\xi(1 - \lambda^2 A^2)(\xi^2 - \eta^2) + \sigma B\eta(1 - \sigma^2 B^2)(\eta^2 - \xi^2) + O(\xi^4 + \eta^4), \tag{8}$$

and, hence,

$$E(\xi, \xi) = O(\xi^4).$$

The mixed derivative terms introduced in Eq. (5) involve cubic polynomials in the matrices A and B. If these matrices are not sparse then the computation will be time consuming. Therefore, we could consider the scheme

$$\begin{aligned} w^{n+1} = Lw^n &+ \delta_x\mu_x\delta_y{}^2(1 - \rho(\lambda A)^2 - 3\rho(\sigma B)^2)(\lambda A w^n/6) \\ &+ \delta_y\mu_y\delta_x{}^2(1 - \rho(\sigma B)^2 - 3\rho(\lambda A)^2)(\sigma B w^n/6), \end{aligned} \tag{9}$$

where $\rho(A)$ denotes the spectral radius of A. With this scheme $E(\xi, \xi)$ is fourth order only for scalar equations. The amplification matrix for this scheme is similar to Eqs. (6) and (7) except that the matrices appearing in (7) are replaced by their spectral radii. We note that if the time step becomes large enough then the coefficients K_1, K_2 appearing in (7) can become negative there is an increase in the phase error. This phenomena is present in the results presented in a later chapter.

The stability of these schemes no longer follows from the proof given by Lax and Wendroff. Indeed, it is not clear that the schemes as given in Eqs. (5) and (9) are stable for any time step. However, by adding an appropriate fourth order term stability can be achieved. As an example should we add the term

$$-\nu\delta_x{}^2\delta_y{}^2 w; \qquad 0 < \nu < \tfrac{1}{8} \tag{10}$$

to Eq. (5) or (9) then it is trivial to show, by a perturbation argument, that the scheme is stable for sufficiently small time steps.

3. A Stable Two Step Family of Difference Schemes

The schemes introduced thus far have the disadvantage of being complicated and that it is time consuming to evaluate the matrices. In addition, it is also difficult to find an analytic stability condition for these methods. We, therefore, introduce a family of two step methods that overcomes these difficulties. To be specific consider the scheme

$$\bar{w} = (\mu_x\mu_y + \gamma\delta_x\delta_y)\, w^n + \tfrac{1}{2}\lambda(\delta_x f(\mu_y w^n) + \delta_y g(\mu_x w^n)),$$
$$w^{n+1} = w^n + \lambda\delta_x f(\mu_y \bar{w}) + \lambda\delta_y g(\mu_x \bar{w}) - \nu\delta_x{}^2\delta_y{}^2 w^n, \tag{11}$$

where, $-\frac{1}{4} \leqslant \gamma \leqslant \frac{1}{4}$, $0 \leqslant \nu \leqslant \frac{1}{8}$, $\lambda = \Delta\tau/\Delta x = \Delta\tau/\Delta y$. When $\gamma = 0$ and $\nu = 0$ we have the rotated Richtmyer scheme which is an efficient algorithm (see Wilson [9]) though with a poor phase representation [7]. We note that for all nonzero γ the same number of function evaluations are required as for the particular case of $\gamma = 0$.

Let ξ and η be the Fourier variables and let

$$\alpha = \sin \xi/2 \qquad \beta = \sin \eta/2.$$

Then, the amplification matrix for this family is

$$G(\xi, \eta) = I + 2i\lambda M(1 - \alpha^2)^{1/2}\,(1 - \beta^2)^{1/2} - 4\gamma\alpha\beta) - 2\lambda^2 M^2 - 16\nu\alpha^2\beta^2, \tag{12}$$

where

$$M = A\alpha(1 - \beta^2)^{1/2} + B\beta(1 - \alpha^2)^{1/2}. \tag{13}$$

The amplification matrix G depends only on the matrix M and not on the matrices A and B individually; hence, we can invoke the spectral mapping theorem. We denote the spectral radius of G by g and that of M by m. Then, from the spectral mapping theorem we have that

$$g = 1 + 2i\lambda m((1 - \alpha^2)^{1/2}\,(1 - \beta^2)^{1/2} - 4\gamma\alpha\beta) - 2\lambda^2 m^2 - 16\nu\alpha^2\beta^2, \tag{14}$$

where new equation (14) is a scalar equations. From Eq. (14)

$$|\, g\,|^2 = (1 - 2\lambda^2 m^2 - 16\nu\alpha^2\beta^2)^2 + 4\lambda^2 m^2((1 - \alpha^2)^{1/2}\,(1 - \beta^2)^{1/2} - 4\gamma\alpha\beta)^2. \tag{15}$$

We also have from Eq. (13) that

$$m \leqslant \rho(A, B)(|\,\alpha\,|\,(1 - \beta^2)^{1/2} + |\,\beta\,|\,(1 - \alpha^2)^{1/2}), \tag{16}$$

where as before $\rho(A, B)$ denotes the larger of the spectral radii of the matrices A and B.

A necessary condition for stability is that $|\, g\,|^2 \leqslant 1$ or by using Eq. (15) that

$$16\alpha^2\beta^2\nu(16\alpha^2\beta^2\nu - 2) + 4[16\nu\alpha^2\beta^2 - 1 + ((1 - \alpha^2)^{1/2}\,(1 - \beta^2)^{1/2} - 4\gamma\alpha\beta)^2]\,\lambda^2 m^2 + 4\lambda^4 m^4 \leqslant 0, \tag{17}$$

which is a quadratic inequality in $(\lambda m)^2$. If $\lambda m = 0$ the inequality is satisfied since α, β, ν are all less than or equal to 1 and ν is always positive. Therefore, the inequality is true for $(\lambda m) \leqslant s_0(\alpha, \beta)$ for some s_0. Thus, if the inequality is satisfied for a particular value of λm then it is also satisfied for all smaller values of λm.

Before we try to find $s_0(\alpha, \beta)$ in particular cases we wish to make a general observation. The terms involving ν are

$$16\alpha^2\beta^2\nu(16\alpha^2\beta^2\nu - 2 + 4\lambda^2 m^2).$$

If $2(\lambda m)^2 > 1$ it then obviously pays to choose $\nu = 0$ in order to improve the inequality (17). However, we shall shortly show that for $\nu = 0$ we have that $(\lambda m)^2 \leqslant \frac{1}{2}$ for all γ. Hence, $(\lambda m)^2 \leqslant \frac{1}{2}$ for all γ and all ν. Since this limit is obtained by the rotated Richtmyer method $\gamma = 0$, $\nu = 0$, it shows that this method is optimal within the family of schemes that we are presently considering.

Since inequality (17) is quite difficult to analyse we shall examine several special cases. When the viscosity coefficient ν is zero the constant term in the inequality vanishes and we immediately can reduce (17) to

$$(\lambda m)^2 \leqslant 1 - ((1 - \alpha^2)^{1/2} (1 - \beta^2)^{1/2} - 4\gamma\alpha\beta)^2 \tag{18}$$

or by using inequality (16) we have that (18) is certainly satisfied if

$$(\lambda \rho(A, B)) = \min_{\substack{-1 \leqslant \alpha \leqslant 1 \\ -1 \leqslant \beta \leqslant 1}} \frac{1 - ((1 - \alpha^2)^{1/2} (1 - \beta^2)^{1/2} - 4\gamma\alpha\beta)^2}{(|\alpha| (1 - \beta^2)^{1/2} + |\beta| (1 - \alpha^2)^{1/2})^2} = \frac{1 - 4|\gamma|}{2}, \tag{19}$$

where the minimum for γ positive occurs when α and β are negative and for γ negative the minimum occurs when α, β are positive. Also, when $A = \pm B$ this condition is also necessary. Thus, for general A and B the necessary condition for stability, when $\nu = 0$, is

$$\lambda\rho(A, B) \leqslant (\tfrac{1}{2}(1 - 4|\gamma|))^{1/2}, \qquad -\tfrac{1}{4} \leqslant \gamma \leqslant \tfrac{1}{4}. \tag{20}$$

Hence, as we have previously indicated, the largest time step occurs when $\gamma = 0$ and in that case $\lambda\rho(A, B) = 1/(2)^{1/2}$. Also, as previously, shown this cannot be improved upon by choosing a nonzero viscosity coefficient.

However, for nonzero γ we can improve the stability condition by introducing viscosity. For example if $\gamma = -\frac{1}{4}$ then formula (20) indicates that the scheme is unconditionally unstable. We shall now show that with $\nu = \frac{1}{8}$ the scheme is stable with $\lambda\rho(A, B) \leqslant \frac{1}{2}$. If we substitutes $\nu = \frac{1}{8}$ and $\gamma = -\frac{1}{4}$ into (17) then we get the inequality

$$\alpha^2\beta^2(\alpha^2\beta^2 - 1) + [2\alpha^2\beta^2 - 1 + ((1 - \alpha^2)^{1/2} (1 - \beta^2)^{1/2} + \alpha\beta)^2](\lambda m)^2 + (\lambda m)^4 \leqslant 0. \tag{21}$$

From our previous remarks it follows that it is sufficient to show that this inequality holds for a particular value of λm and it then follows that it is true for all smaller values of λm. Choosing $\lambda\zeta = \frac{1}{2}$ inequality (21) becomes

$$4\alpha^2\beta^2(\alpha^2\beta^2 - 1) + \tfrac{1}{4}[8\alpha^2\beta^2 - 1 + ((1-\alpha^2)^{1/2}(1-\beta^2)^{1/2} + \alpha\beta)^2] + \tfrac{1}{16} \leqslant 0. \quad (22)$$

Furthermore, we need only consider α, β positive since the left-hand side of (22) is more negative for α, β negative. In that case inequality (22) is equivalent to

$$\tfrac{1}{16}[(\alpha^2 - \beta^2)^4 + 8(\alpha^4 - \beta^4)^2] + (\alpha^2 + \beta^2)(\alpha\beta)[(\alpha^2 + \beta^2) - 4\alpha^2\beta^2] \geqslant 0. \quad (23)$$

But the first term is the sum of squares and is always positive while the second term is positive by the Schwarz inequality. Hence (22) is verified and so we have shown that we have a stability condition that

$$\lambda\rho(A, B) \leqslant \tfrac{1}{2} \qquad \text{when} \quad \gamma = -\tfrac{1}{4}, \quad \nu = \tfrac{1}{8}.$$

The phase error associated with (11) is

$$\begin{aligned} E(\xi, \eta) = {} & \tfrac{1}{6}[\lambda A\xi^3 + \lambda B\eta^3 - (\lambda A\xi + \lambda B\eta)^3] + \lambda A(\tfrac{1}{4}\xi\eta^2 + \gamma\xi^2\eta) \\ & + \lambda B(\tfrac{1}{4}\xi^2\eta + \gamma\xi\eta^2) + O(\xi^4 + \eta^4). \end{aligned} \quad (24)$$

We see that if γ is positive then the phase lag is large for ξ, η positive while if γ is negative the phase lag is large for ξ positive and η negative. Thus, nothing is gained over the rotated Richtmyer method which itself has a poor phase representation.

4. A Nonlinear Scheme

The general conclusion is that with the family of schemes considered in (11) the rotated Richtmyer scheme is the best. As previously noted when γ is positive there are difficulties when ξ and η have the same sign while when γ is negative there are difficulties when ξ and η have differing signs. Another approach to this problem is to examine the special case $\gamma = \pm\frac{1}{4}$. When $\gamma = \frac{1}{4}\bar{w}_{i+1/2,j+1/2}$ is approximated by $\frac{1}{2}(w_{i+1,j+1} + w_{i,j})$ while when $\gamma = -\frac{1}{4}$ it is approximated by $\frac{1}{2}(w_{i+1,j} + w_{i,j+1})$. Obviously in either case trouble will occur if the wave is traveling perpendicular to the line connecting the points used in the approximation. The obvious solution is to introduce a nonlinearity that will determine the sign of γ depending on the direction of the wave. One way to accomplishing this is to consider the scheme

$$\begin{aligned} \bar{w} &= \mu_x\mu_y w^n + \gamma\,\text{sign}(\mu_x\mu_y w^n)\,|\,\delta_x\delta_y w^n\,| + \tfrac{1}{2}\lambda(\delta_x f(\mu_y w^n) + \delta_y g(\mu_x w^n)) \\ w^{n+1} &= w^n + \lambda\delta_x f(\mu_y\bar{w}) + \lambda\delta_y g(\mu_x\bar{w}) - \nu\delta_x^2\delta_y^2 w^n. \end{aligned} \quad (25)$$

For the purposes of the Fourier transform we replace $\operatorname{sign}(\mu_x\mu_y w) \mid \delta_x\delta_y w \mid$ by $\operatorname{sign}(\operatorname{Re} \mu_x\mu_y w) \operatorname{sign}(\operatorname{Re} \delta_x\delta_y w)\, \delta_x\delta_y w$; these formulations coincide when w is real. With this replacement we can formally construct an amplification matrix which has the form

$$G(\xi, \eta) = I + 2i\lambda M((1 - \alpha^2)^{1/2} (1 - \beta^2)^{1/2} - 4\gamma \mid \alpha \mid \mid \beta \mid) - 2\lambda^2 M^2 - 16\nu\alpha^2\beta^2, \tag{26}$$

where, as before,

$$M = A\alpha(1 - \beta^2)^{1/2} + B\beta(1 - \alpha^2)^{1/2}.$$

Using the same techniques as with the linear scheme we get the formal stability condition, for $\nu = 0$,

$$\lambda\rho(A, B) \leqslant (\tfrac{1}{2}(1 - 4\gamma))^{1/2}. \tag{27}$$

Hence for γ negative we have a large permissable time step than that allowed by the linear scheme as given in Eq. (20). In fact with $\gamma = -\frac{1}{4}$ we now have a maximal allowable time step without using a Strang type splitting scheme. The phase error for the nonlinear scheme is given by

$$\begin{aligned} E(\xi, \eta) = {} & \tfrac{1}{6}(\lambda A\xi^3 + \lambda B\eta^3 - (\lambda A\xi + \lambda B\eta)^3) + \lambda A(\tfrac{1}{4}\xi\eta^2 - \gamma\xi \mid \xi\eta \mid) \\ & + B(\tfrac{1}{4}\xi^2\eta - \gamma\eta \mid \xi\eta \mid) + O(\xi^4 + \eta^4). \end{aligned} \tag{28}$$

As γ approaches $\frac{1}{4}$ the phase error decreases though by (27) the permissible time step also decreases. This is in agreement with what is found in many schemes, that improved phase representation frequently goes together with a small permissible time step (the Strang splitting method L_xL_y is an exception to this rule).

5. Results

As a computational test of the schemes considered we have chosen a sample problem from dynamic elasticity. This problem has the advantage of having constant coefficients and a known analytic solution but it is still a realistic nontrivial system of equations with noncommuting matrices A and B. In addition the x and y directions are not treated symmetrically. Therefore, there is no built in bias towards the modified Lax–Wendroff method for which the phase error is fourth order when $\xi = \eta$.

The equations considered are those of linear elasticity, i.e.,

$$\begin{aligned} \rho u_t &= \tau_{11,x} + \tau_{12,y}, \\ \rho v_t &= \tau_{12,x} + \tau_{22,y}, \\ \tau_{11,t} &= (2\mu + \lambda) u_x + \lambda v_y, \\ \tau_{12,t} &= \mu(u_y + v_x), \\ \tau_{22,t} &= \lambda u_x + (2\mu + x) v_y, \end{aligned} \tag{29}$$

where ρ, λ, μ are positive constants. The particular values chosen for the test problem are $\rho = 0.175$, $\lambda = 0.3$, $\mu = 0.2$. These equations are integrated over the unit square $0 \leqslant x \leqslant 1$. $0 \leqslant y \leqslant 1$. The solution is assumed to be periodic in the x variable with a period of 1. μ, τ_{11}, τ_{22} are symmetric about the axis $y = 0$ while v and τ_{12} are antisymmetric. At the boundary $y = 1$ we use the free surface condition that $\tau_{12} = \tau_{22} = 0$.

A particular solution to this problem is

$$\begin{aligned} u(x, y, t) &= \omega \left(\xi \cos \alpha y + \frac{(\beta^2 - \xi^2) \cos \alpha}{2\xi \cos \beta} \cos \beta y\right) \cos \xi(x - vt), \\ v(x, y, t) &= -\omega \left(\alpha \sin \alpha y - \frac{(\beta^2 - \xi^2) \cos \alpha}{2\beta \cos \beta} \sin \beta y\right) \sin \xi(x - vt), \\ (x, y, t) &= \mu \left((2\alpha^2 - \beta^2 - \xi^2) \cos \alpha y - \frac{(\beta^2 - \xi^2) \cos \alpha}{\cos \beta} \cos \beta y\right) \cos \xi(x - vt), \\ (x, y, t) &= 2\mu\alpha\xi \left(\sin \alpha y - \frac{\sin \alpha}{\sin \beta} \sin \beta y\right) \sin \xi(x - vt), \\ (x, y, t) &= \mu(\xi^2 - \beta^2) \left(\cos \alpha y - \frac{\cos \alpha}{\cos \beta} \cos \beta y\right) \cos \xi(x - vt), \end{aligned} \tag{30}$$

where

$$\xi = 2\pi \qquad \omega = \xi v \tag{31}$$

and

$$\begin{aligned} \alpha^2 &= \xi^2((v^2/c_1^2) - 1) \qquad \beta^2 = \xi^2((v^2/c_2^2) - 1) \\ c_1^2 &= (\partial\mu + \lambda)/\rho = 4.0 \qquad c_2^2 \simeq \mu/\rho = 1.143; \end{aligned} \tag{32}$$

hence, $\alpha = \alpha(v)$, $\beta = \beta(v)$.

The boundary condition at $y = 1$ is satisfied if v is a solution to the transcendental equation

$$\frac{\tan \beta}{\tan \alpha} + \frac{4\xi^2\alpha\beta}{(\xi^2 - \beta^2)^2} = 0. \tag{33}$$

The particular solution chosen was the third mode which yields v approximately equal to 1.6975 for the parameters used. When the initial conditions are chosen using Eq. (30) with $t = 0$, then the system (30) is the unique solution to the problem (see [6] and [7] for further details about this problem).

In Tables I and II we compare the various schemes over the time interval

TABLE I

Scheme	ν	CFL	WCHG	Phase
LW	0.0	0.314	0.0957	0.030
MAT	0.0	0.314	−0.0402	0.0057
SPEC	0.0	0.314	−0.002	0.012
LW	0.0	0.471	0.1299	0.027
MAT	0.0	0.471	−0.0238	0.0079
MAT	0.01	0.471	0.0720	0.0077
SPEC	0.0	0.471	0.1061	0.024
LW	0.0	0.754	−0.7484	0.020
LW	0.05	0.754	−0.0639	0.018
LW	0.0625	0.754	unstable	
MAT	0.0	0.754	−0.946	0.0148
SPEC	0.0	0.754	0.4639	0.057

TABLE II

Two Step Phase Improvement as Given in Eq. (34)

ν	CFL	WCHG	Phase
0.0	0.314	−0.0580	0.0014
0.01	0.314	0.0876	0.0012
0.0	0.471	−0.1122	−0.0013
0.02	0.471	0.0898	−0.0017
0.0	0.754	unstable	—

required for the analytic solution to complete four periods. The time step was kept constant throughout the time integration. The time steps were chosen so that a complete period required an integer number of time steps. Thus the CFL numbers ($\text{CFL} = (\Delta\tau/\Delta x)((\lambda + 2\mu)/\rho)^{1/2}$) are not simple fractions. With a CFL condition of about 0.471 160 time steps are needed to complete the four periods. Let WCHG be a measure of the energy growth in the numerical solution, i.e.,

$$\text{WCHG} = \frac{\text{initial energy} - \text{energy at time } t}{\text{initial energy}},$$

where the elastic energy of the system is given by

$$w = \frac{1}{2}\rho(\mu^2 + v^2) + \frac{1}{2\mu(3\lambda + 2\mu)}[(\lambda + \mu)(\tau_{11}^2 + \tau_{22}^2) - \lambda\tau_{11}\tau_{22} + (3\lambda + 2\mu)\tau_{12}^2].$$

WCHG measures the L stability of the solution. The analytic solution has the property that the energy is independent of time, i.e., WCHG $= 0$ for all time. Computationally, when WCHG is negative then the norm of the numerical solution is increasing in time which indicates some instability. When the energy has doubled within the four periods we label the scheme as unstable for that CFL number.

Let PHASE be a measure of the phase error in the numerical solution, i.e.,

$$\text{PHASE} = \frac{\text{pos(analytic)} - \text{pos(computational)}}{\text{pos(analytic)}}$$

where pos denotes the position of a particular zero of the variable u. The computational zero is calculated by linear interpolation on each y coordinate line and then the result is averaged over all these y coordinate lines. A 16×16 mesh was chosen for this problem. However, extra coordinate lines were added for convenience in handling the periodic and symmetry boundary conditions. We have also included a fourth order viscosity with a coefficient ν as given by Eqs. (9), (10), and (25). For the nonlinear schemes we have also listed the parameter γ that appears in Eq. (25).

Table I gives a comparison of the Lax–Wendroff method (LW) together with the improved matrix version (MAT) as given by Eq. (5) and also the improved version with the matrices in Eq. (7) replaced by their respective spectral radii (SPEC) as given in Eq. (9).

For the dynamic elastic equations the relevant powers of the matrices A and B are sparse and in fact are very close in form to the original matrices A and B. Hence little computer time is added in the evaluation of these matrices. However, this need not be true for all physically relevant equations. As seen in Table I the spectral form of the improvement reduces the phase lag for sufficiently small time steps. However, with a CFL number of about $\frac{1}{2}$ there is virtually no improvement over the Lax–Wendroff method. When the time step is chosen large enough the third order terms included in Eq. (9) become negative. This improves the stability of the scheme but increases the phase error. With the matrix form of the improvement the decrease in phase error is much more noticeable. Even with a time step of CFL $= 0.471$ the phase error of MAT is less than one-third of the phase error for the Lax–Wendroff method. Thus, in this case the extra computational effort is compensated for by increased accuracy.

Both forms of improvement discussed above work better with small time steps.

As Δt increases the factors K_1 and K_2 in Eq. (7) become less positive and hence produce less of a phase error reduction. Furthermore, even with our improvement there still exists a phase lag of almost one percent after four periods. It therefore seems advisable to further decrease the phase lag by eliminating the negative terms in K_1 and K_2. Thus, we replace Eqs. (5) and (9) by the simpler scheme

$$w^{n+1} = Lw^n + (\lambda/6)\, \delta_x\mu_x\delta_y{}^2 Aw + (\sigma/6)\, \delta_y\mu_y\delta_x{}^2 Bw - \nu\delta_x{}^2\delta_y{}^2 w, \tag{34}$$

where, as before, L is the Lax–Wendroff operator. The amplification matrix for this scheme with $\nu = 0$ is still given by Eq. (6) where now both K_1 and K_2 are equal to $\frac{2}{3}$. The phase error for this scheme is given by

$$E(\xi, \eta) = \tfrac{1}{6}[\lambda A\xi(1 - \lambda^2 A^2)(\xi^2 - \eta^2) + \sigma B\eta(1 - \sigma^2 B^2)(\eta^2 - \xi^2) - 3\lambda\sigma\xi\eta AB(\lambda A\xi + \sigma B\eta)] + O(\xi^4 + \eta^4). \tag{35}$$

We note that at $\xi = \eta$ the phase error is no longer of fourth order. Nevertheless, this scheme seems to yield lower phase errors than the scheme given in Eq. (5). Possibly, the phase gain exhibited in Eq. (34) balances the higher order phase lag which until now has been neglected. As seen in Table II the phase lag for this new scheme is lower by a factor of greater than 10 over the Lax–Wendroff method. Also in contrast to the other improvements this scheme maintains a very small phase error as the time step increases.

The scheme introduced in (34) has the additional advantage that it can be written as a two step method. Thus, considering the divergence free form given in Eq. (1′) we present a modification of the two step method in Thommen [8].

$$\begin{aligned} w^{(1)} &= \mu_x w^n - (\lambda/2)\, \delta_x f - (\sigma/2)\, \mu_x\delta_y\mu_y g, \\ w^{(2)} &= \mu_y w^n - (\lambda/2)\, \mu_y\delta_x\mu_x f - (\sigma/2)\, \delta_y g, \\ w^{n+1} &= w^n - \lambda\delta_x f^{(1)} - \sigma\delta_y g^{(2)} + \tfrac{1}{6}[\lambda\delta_x\mu_x\delta_y{}^2 f + \sigma\delta_y\mu_y\delta_x{}^2 g] - \nu\delta_x{}^2\delta_y{}^2 w. \end{aligned}$$

The linearized version of these equations is identical with (34) and so Eq. (36) also produces small phase errors.

In Table III we show a comparison between different members of the nonlinear scheme as given by Eq. (25). For the sake of comparison we have also included the Burstein scheme (see [7]) which is denoted by a B in the first column. As expected the phase error with $\gamma = -\frac{1}{4}$ or $\gamma = 0$ (rotated Richtmyer) is quite large and hence these schemes should not be used in calculations of the phases of waves. However, with $\gamma = -\frac{1}{4}$ there is the compensating feature of a maximal permissible time step coupled together with few function evaluations. This should make the scheme very competetive for problems where only steady state solutions are of importance.

TABLE III

γ	ν	CFL	WCHG	Phase
−0.25	0.0	0.471	0.2140	0.096
0.0	0.0	0.471	0.1714	0.068
0.125	0.0	0.471	0.0478	0.053
B	0.0	0.471	0.0345	0.047
0.248	0.0	0.471	−0.1784	0.033
0.248	0.0625	0.471	0.4849	0.037
−0.25	0.0	0.754	0.3158	0.088
0.0	0.0	0.754	0.2086	0.060
0.125	0.0	0.754	0.0571	0.049
B	0.0	0.754	−0.0278	0.042
0.248	0.0625	0.754	0.1691	0.028
−0.25	0.0	0.942	0.2419	0.077
0.0	0.0	0.942	−0.5203	0.052
0.0	0.0625	0.942	0.0835	0.059
0.125	0.0625	0.942	unstable	—
B	0.0625	0.942	unstable	—

As we increase the parameter γ the phase error decreases and at $\gamma = \frac{1}{8}$ we have phase errors close to that of the Burstein scheme though with a slightly larger time step. However, as shown in [7], the Burstein scheme is unstable for some choices of the matrices A and B but is stable for the dynamic elastic equations. With γ close to $\frac{1}{4}$ the phase error is almost as small as that given by the Lax–Wendroff method in agreement with our previous analysis. Thus, the nonlinear scheme offers a useful general purpose method with the option that there exists a free parameter γ, which can be chosen for the individual problem to either maximize the permissible time step or else to achieve a small phase error at the expense of a small time step.

6. Conclusions

Two new schemes of second order have been presented for hyperbolic systems of equations in two space dimensions. The first method is a modification of the Lax–Wendroff method. Adding appropriate third order mixed derivatives one can insure that the phase error is of fourth order for the special case that the two Fourier variables ξ and η are equal. Numerical tests indicate that the phase error with this method is about one quarter that of the Lax–Wendroff method; while the Lax–Wendroff method has as small a phase error as any other nine point scheme (see [7]). Replacing the powers of the matrices that appear in the correction term, by their spectral radii yields a scheme which small phase errors only for sufficiently small time steps and hence there is apparently little benefit compensating for the additional computational effort. Should one eliminate the powers of the matrices that appear in the correction term then the scheme can be written as a two step method with only the vectors f and g appearing and not the matrices A and B (Eq. (36)). The phase factor for this scheme displays a phase gain rather than the usual phase lag. In computational tests the phase error for this scheme was smaller than even that of the matrix form of the correction term. Thus, Eq. (36) requires little more work than the two step Thommen scheme but yields a phase error that is smaller by a factor of ten over any other nine point scheme. Hence, the phase error for this scheme is comparable with that of higher order methods but requires less computational time and even more important this method uses only a nine point lattice and so presents fewer problems near boundaries especially curved boundaries. The scheme, Eq. (36), seems to be mildly unstable and requires a small fourth order viscosity to stabilize it.

The second method considered is actually a family of schemes. The linear version of this method has a large phase error and small allowable time steps compared with the rotated Richtmyer method. Therefore, a nonlinear version of the method was introduced and an amplification matrix formally constructed. Choosing the free parameter, γ, equal to $\frac{1}{4}$ yields a scheme which allows (formally) maximal time steps for an explicit scheme but produces a large phase error. Choosing γ slightly larger than $-\frac{1}{4}$ yields a scheme with small phase error but also with a small permissible time step. For general problems one would choose an intermediate value of γ (e.g., approximately $-\frac{1}{8}$) to achieve a moderate phase error coupled with a reasonable time step.

Wilson [10] has observed that the most efficient schemes, in terms of speed, are the Strang splitting methods. It should be noted that with $\gamma = \frac{1}{4}$ no additional evaluations of the vectors f and g are required over the rotated Richtmyer method, $\gamma = 0$. Yet the allowable time step is greater by a factor of $(2)^{1/2}$. Therefore, this particular method, i.e., Eq. (25) with $\gamma = \frac{1}{4}$ would be the most efficient nine point second order scheme presently available. Hence, for problems that do not involve wave propagation, this method is to be recommended.

This two step family of schemes is another demonstration of the benefits that can accrue by the introduction of nonlinearities into a linear scheme, even for linear problems. Fromm [3] has also introduced nonlinearities for the purpose of reducing phase errors. Similarly Boris and Book [1], Harten and Zwas [4], and Van Leer [9] have introduced nonlinearities into the difference schemes in order to preserve monotonicity properties for the numerical solution. There is, therefore, an important need for additional theoretical work justifying the convergence of nonlinear schemes, at least for linear differential equations, under suitable hypotheses.

References

1. J. P. Boris and D. L. Book, *J. Computational Phys.* **11** (1973), 38.
2. B. Eilon, D. Gottlieb and G. Zwas, *J. Computational Phys.* **9** (1972), 387.
3. J. E. Fromm, *J. Computational Phys.* **3** (1968), 176.
4. A. Harten and G. Zwas, *J. Computational Phys.* **9** (1972), 568.
5. P. D. Lax and B. Wendroff, *Comm. Pure Appl. Math.* **17** (1964), 381.
6. R. D. Mindlin, Waves and vibrations in isotropic plates, *in* "Proceedings of the First Symposium on Naval Structural Mechanics" (J. N. Goodier, N. J. Hoff, Ed.).
7. E. Turkel, Phase error and stability of second order methods for hyperbolic problem I, *J. Comp. Physics*, at press.
8. H. U. Thommen, *Z. Agnew. Math. Phys.* **17** (1966), 369.
9. B. Van Leer, Towards the Ultimate Conservative Difference Scheme. I. The Quest for Monotonicity, *in* "Proceedings of the Third Int. Conference on Number Methods in Fluid Mech." (J. Ehlers *et al.*, Eds.), Springer-Verlag, Berlin/New York, 1972.
10. J. C. Wilson, *J. Inst. Math. Appl.* **10** (1972), 238.
11. S. Z. Burstein, *J. Comp. Phys.* **1** (1966), 198.