JOURNAL OF COMPUTATIONAL PHYSICS 3, 176–189 (1968)

A Method for Reducing Dispersion in Convective Difference Schemes

JACOB E. FROMM[1]

IBM Research Laboratory, San Jose, California 95114

Received April 22, 1968

ABSTRACT

Difference schemes that can be implemented through the fractional time step technique of Marchuk are amenable to an expedient which may be termed "a method of zero average phase error". The method involves a linear combination of forward and backward time steps. It is applied to a second-order approximation and compared with a fourth order approximation. There is a substantial reduction in dispersion.

I. INTRODUCTION

With progressing experience in the use of nonlinear difference approximations in the solution of fluid flow problems, it has become increasingly evident that the methods used have fallen far short of covering the range of applicability which was anticipated of them in their early application through the computer. The limitations (to low Reynolds numbers, for example) were not understood, nor was there any known practical means of giving measure to the relative merit of numerically stable schemes. With the appearance of the enlightening work of Roberts and Weiss [1], it is now evident that limitations of the methods relate to the dispersion and dissipation present in the numerical approximation. Numerical dissipation dominates in first-order approximations, while in second-order approximations, dispersion at high wave numbers becomes the most serious shortcoming. Dispersion is common to all methods, but dissipation is absent from time-centered schemes. Unfortunately, the explicit time-centered scheme (leap-frog method) has equally severe dispersion as counterparts of the same order that are not time centered. It appears that one attribute of an artificial viscosity is to eliminate high wave number components in the solution. Because of the adverse phase distortions of high wave number components, coupled with no dissipation, time-centered methods of necessity must use an artificial viscosity if the physical problem does

[1] Also with the IBM Scientific Center, Palo Alto, California.

not itself contain a large viscosity. Roberts and Weiss have shown that, with a fourth-order time-centered method, those wave numbers which behave adversely are more sharply delineated from the meaningful low wave numbers. With this fourth-order method a more disciminating use can be made of artificial viscosity.

In contrast, we shall look here at difference methods that do contain some dissipation but by virtue of their derivation have neither dissipation nor dispersion for mesh-length transfer of a given variable. These methods are often classed under the heading of Lax–Wendroff [2], but notable distinctions exist between them and the two-step Lax–Wendroff method used by Burstein [3]. The methods considered are all subject to the fractional time step procedure of Marchuck [4] and are of the same class as the method employed successfully by Leith [5]. The methods must be designed to optimize both amplitude and phase properties of meaningful wave numbers. Notable in the improvement of such methods and their extension to fourth order is the work of Crowley [6], [7].

We describe some additional modification and extensions which are applicable in terms of fractional time steps. We first give the amplitude and phase properties of the second-order method employed by Leith [5]. With these properties as a reference we proceed with a description of the artifice used to give a sharp reduction in dispersion in this method. Comparison of the resultant phase and aplitude properties are made with the unmodified second-order method and also with the fourth-order method of Crowley [7].

II. The Second-Order Fractional Time Step Method

The fractional time step method involves component-wise addition of convection contributions. That is, the results obtained by a one-dimensional calculation are operated upon successively to extend the dimensionality. It can be shown that this extension of the dimensionality also increases the order of approximation and is always expressible as a single step.

To illustrate, let us consider the Helmholz vorticity equation for ideal fluid flow in two space dimensions

$$\frac{\partial \omega}{\partial t} + u\frac{\partial \omega}{\partial x} + v\frac{\partial \omega}{\partial y} = 0, \tag{1}$$

where ω is the vorticity, u and v are the velocities in the x and y directions, respectively, and t is the time. In the incompressible case we obtain a closed set of equations by including the velocity divergence equation

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{2}$$

and the definition of vorticity, namely,

$$\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \omega. \tag{3}$$

Implementation of this system of equations in finite-difference calculations is usually achieved by defining a streamfunction ψ so that

$$u = \frac{\partial \psi}{\partial y}, \qquad v = -\frac{\partial \psi}{\partial x}. \tag{4}$$

Equation (2) is then identically satisfied and (3) becomes

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = -\omega. \tag{5}$$

We consider a fractional time step difference approximation of (1) and write to second order in the space variables

$$\tilde{\omega}_{l,m} = \omega^n_{l,m} + \frac{\alpha^n_{l,m}}{2}(\omega^n_{l-1,m} - \omega^n_{l+1,m}) + \frac{(\alpha^n_{l,m})^2}{2}(\omega^n_{l-1,m} - 2\omega^n_{l,m} + \omega^n_{l+1,m}),$$

and

$$\omega^{n+1}_{l,m} = \tilde{\omega}_{l,m} + \frac{\beta^n_{l,m}}{2}(\tilde{\omega}_{l,m-1} - \tilde{\omega}_{l,m+1}) + \frac{(\beta^n_{l,m})^2}{2}(\tilde{\omega}_{l,m-1} - 2\tilde{\omega}_{l,m} + \tilde{\omega}_{l,m+1}). \tag{6}$$

Here l and m are mesh indices in the x and y direction respectively and n is the time index such that $\omega^n_{l,m} = \omega(l\,\Delta x, m\,\Delta y, n\,\Delta t)$; α and β are defined by

$$\alpha^n_{l,m} \equiv \frac{u^n_{l,m}\,\Delta t}{\Delta x}$$

and

$$\beta^n_{l,m} \equiv \frac{v^n_{l,m}\,\Delta t}{\Delta y}. \tag{7}$$

The tilda in (6) indicates fractional time step values resulting from x component convection. Expansion of (6) leads to a 9-point formula which may also be derived geometrically. The 9-point formula is stable for $\alpha \leqslant 1$ and $\beta \leqslant 1$ while the associated 5-point truncated formula (an extension of dimensionality without an increase in order) is unstable for all but the highest wave numbers. This was shown by Leith.

To correctly examine the stability and phase properties of (6) we must do the analysis in two space dimensions since a one-dimensional analysis will not show the stability gained by increasing the order of approximation.

We proceed by linearizing (6), i.e., let α and β be assumed constant (to later be varied).

Substituting the Fourier component solution

$$\omega_{l,m}^{n} = Ar^{n} \exp i(k_x l\,\Delta x + k_y m\,\Delta y), \tag{8}$$

into (6) we may write

$$r = \tilde{r}_{l,m} + \frac{\beta}{2}(\tilde{r}_{l,m-1} - \tilde{r}_{l,m+1}) + \frac{\beta^2}{2}(\tilde{r}_{l,m-1} - 2\tilde{r}_{l,m} + \tilde{r}_{l,m+1}),$$

where

$$\begin{aligned} \tilde{r}_{l,m} &= 1 - i\alpha \sin k_x \Delta x + \alpha^2(\cos k_x \Delta x - 1), \\ \tilde{r}_{l,m+1} &= (\cos k_y \Delta y + i \sin k_y \Delta y)\,\tilde{r}_{l,m}\,, \end{aligned} \tag{9}$$

and

$$\tilde{r}_{l,m-1} = (\cos k_y \Delta y - i \sin k_y \Delta y)\,\tilde{r}_{l,m}\,.$$

Numerical stability requires that

$$r\bar{r} \leqslant 1, \tag{10}$$

where $\bar{r}$ is the complex conjugate of r.

Following Roberts and Weiss, we shall examine the properties of $r\bar{r}$ over the stable range of α and β, i.e., $0 \leqslant \alpha \leqslant 1$ and $0 \leqslant \beta \leqslant 1$. If $r\bar{r} = 1$ a disturbance will neither grow nor damp. This is an ideal property for an approximation of pure convection, hence our interest lies in determining how far short we are of this in terms of numerical damping.

In Fig. 1 we have plotted the squared amplitude property $r\bar{r}$ for a series of wave numbers as a function of α and β in the range $0 \leqslant \alpha \leqslant 1$ and $0 \leqslant \beta \leqslant 1$. The contours are for prescribed values of $r\bar{r}$ with reference level $r\bar{r} = 1$. The plot intervals are a power of 2, chosen such that a near optimum number of readable contour levels occur in the individual plots. We have for simplicity taken $k_x \Delta x = k_y \Delta y = \Theta$ and examined the behavior of $r\bar{r}$ for

$$\Theta = \pi, \frac{2\pi}{3}, \frac{\pi}{2}, \frac{\pi}{3}, \frac{\pi}{4}, \text{ and } \frac{\pi}{5},$$

corresponding respectively to wave lengths

$$\Lambda = 2\Delta x,\ 3\Delta x,\ 4\Delta x,\ 6\Delta x,\ 8\Delta x, \text{ and } 10\Delta x.$$

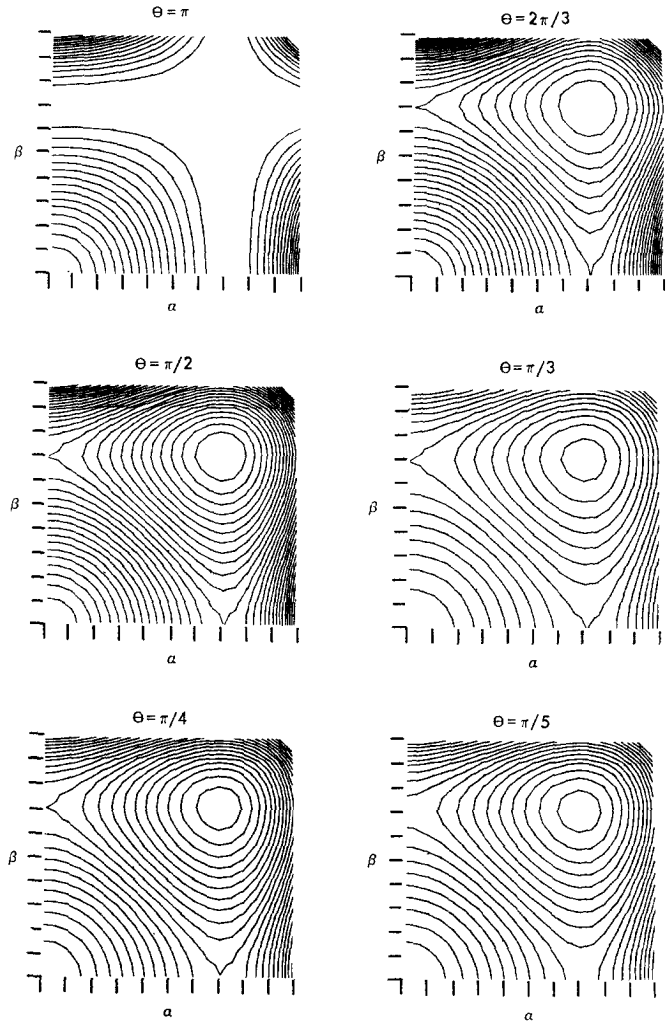

FIG. 1. Contour plots of the square of the modulus of amplification for the second order method of Eq. (6). See Table I.

The reference value, $r\bar{r} = 1$, occurs in each case at $\alpha = \beta = 0$ and at α and (or) $\beta = 1$. This absence of damping in the latter instance is consistent with point to point transfer of grid values and is implicit in the difference method we are considering. The extremum in each of the plots is a point of maximum damping. For $\alpha = \beta \approx 0.75$, we have complete damping for the highest finite grid wave number, $\Theta = \pi$. Lesser damping occurs at succeeding lower wave numbers such

that, in the low wave number limit, the system (6) will be neutrally stable ($r\bar{r} = 1$) everywhere in the range of α and β. The statistics of Figure 1 are given in Table I.

TABLE I

STATISTICS OF FIGS. 1 AND 3 (amplitude)^2
INCLUDING FOURTH-ORDER APPROXIMATION DATA

Λ	Θ	Plot Increment (Fig. 1)	Minimum (Amplitude)^2 (Fig. 1)	Plot Increment (Fig. 3)	Minimum (Amplitude)^2 (Fig. 3)	Minimum (Amplitude)^2 4th Order
$2\Delta x$	π	1/16	0	1/16	0	0
$3\Delta x$	$2\pi/3$	1/32	.1916	1/32	.2234	.3621
$4\Delta x$	$\pi/2$	1/64	.5626	1/64	.6103	.7903
$6\Delta x$	$\pi/3$	1/128	.8789	1/256	.9010	.9742
$8\Delta x$	$\pi/4$	1/512	.9575	1/512	.9665	.9949
$10\Delta x$	$\pi/4$	1/1024	.9765	1/2048	.9859	.9986

To investigate the dispersion properties of (6) we consider phase errors at the higher wave numbers as above. The true phase shift of our component type of procedure is given by

$$\phi_1 = -\alpha k_x \Delta x - \beta k_y \Delta y. \tag{11}$$

The phase shift of the numerical approximation is

$$\phi_2 = -\tan^{-1} \frac{\mathrm{Im}(r)}{\mathrm{Re}(r)}. \tag{12}$$

Here we plot the phase error

$$\Delta\phi \equiv \phi_1 - \phi_2. \tag{13}$$

In Fig. 2 we give contour plots of $\Delta\phi$ in the range $0 \leqslant \alpha \leqslant 1$ and $0 \leqslant \beta \leqslant 1$ for the approximation (6). The contours are for prescribed values of $\Delta\phi$ with reference level $\Delta\phi = 0$. Statistics of the plots of Figure 2 are given in Table II.

We examine the behavior of $\Delta\phi$ for successive values

$$\Theta = \frac{2\pi}{3}, \frac{\pi}{2}, \frac{\pi}{3}, \frac{\pi}{4}, \text{ and } \frac{\pi}{5}.$$

The reference value, $\Delta\phi = 0$, occurs in each case at $\alpha = \beta = 0$ and at α and (or) $\beta = 1$. As with $r\bar{r}$ this is implicit in the difference method since, for α and (or) $\beta = 1$, mesh values are passed from point to point. This is true even for $\Theta = \pi$, but if

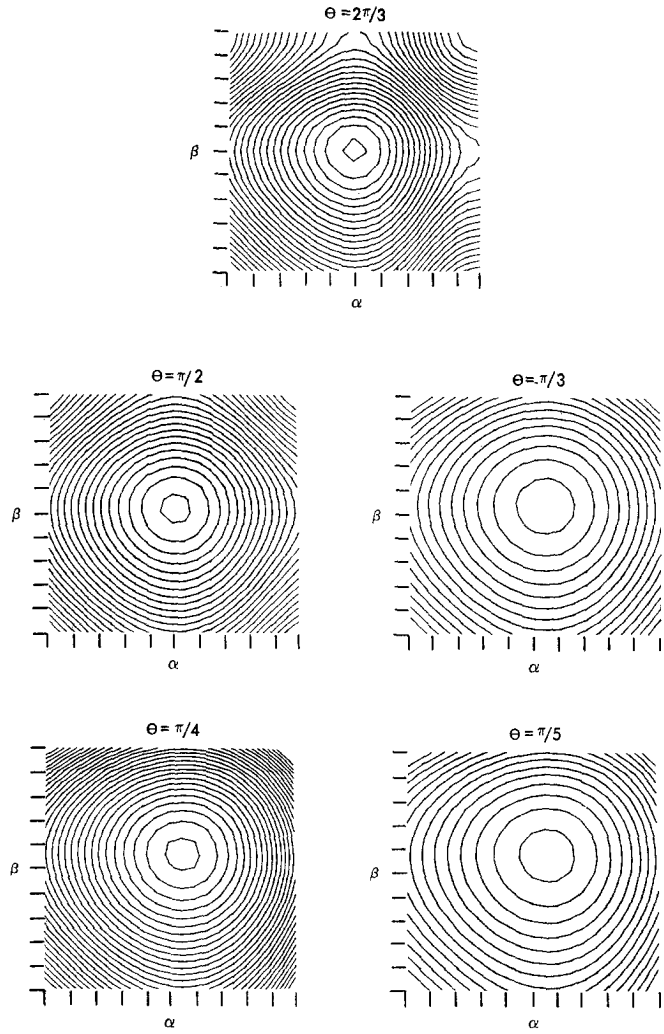

FIG. 2. Contour plots of phase error, $\Delta\phi$, in radians per time step Δt for the second-order method of Eq. (6). See Table II.

α and (or) $\beta \neq 1$, disturbances for $\Theta = \pi$ do not propogate at all. Because of this nonpropogation, which incidentally is common to all difference approximations, we delete this highest wave number from consideration as a meaningful part of any numerical solution. Fortunately, as Crowley [7] has pointed out, the inherent dissipation of the numerical approximation becomes an asset in that it tends to remove this nonpropogating mode.

TABLE II

STATISTICS OF FIGS. 2 AND 4 (phase error in radians/time step)
INCLUDING FOURTH-ORDER APPROXIMATION DATA

Λ	Θ	Plot Increment (Fig. 2)	Max Phase Error (Fig. 2)	Plot Increment (Fig. 4)	Max Phase Error (Fig. 4)	Max Phase Error 4th Order
$3\Delta x$	$2\pi/3$	1/32	$-.8826$	1/128	$\mp.1415$	$-.4673$
$4\Delta x$	$\pi/2$	1/64	$-.3947$	1/1024	$\mp.0122$	$-.1406$
$6\Delta x$	$\pi/3$	1/128	$-.1291$	1/1024	$\pm.0075$	$-.0236$
$8\Delta x$	$\pi/4$	1/512	$-.0573$	1/2048	$\pm.0051$	$-.0061$
$10\Delta x$	$\pi/5$	1/512	$-.0301$	1/4096	$\pm.0031$	$-.0021$

In the phase error plots of Fig. 2 the extremum in each case is the point of maximum error. By definition (13) a phase lag is represented by a negative number and we see that, neglecting roundoff, we always have a lagging error for the approximation (6). The phase error diminishes rapidly at successively lower wave numbers such that in the limit of small wave numbers the error vanishes. Borrowing the terminology from physical optics, we say the dispersion is "normal", i. e., the wave velocity increases with increased wavelength. The effects are upstream steepening of waves.

It is almost a general feature of difference methods that they produce dispersion of the "normal" type. Assuming that this is to be expected we ask: what can be done to reverse the phase error to produce a lead in phase? The answer, of course, is that if we knew an advanced time solution and calculated a solution from it for an earlier time, the inherent lag of the difference approximation would represent a lead in phase viewed in terms of forward time.

III. THE ZERO AVERAGE PHASE ERROR METHOD

To effectively include an advanced time solution to reverse the phase error, one immediately considers implicit methods. Unfortunately the presently known implicit methods also exhibit a lagging phase error. In the present report, we shall make use of the point-to-point transfer of grid values that is inherent in fractional time step methods to affect an advanced time. We may think of a local advanced time solution corresponding to α and (or) $\beta = \pm 1$. We assume that a directional difference approximation can be written and because of the componentwise treatment, we can affect an advanced time solution in any direction. Thus, con-

sidering an approximation backward in time (backward in time from a projected solution for $\alpha = \beta = 1$), we can write for α and $\beta > 0$

$$\tilde{\omega}_{l,m} = \omega^n_{l-1,m} + \frac{\alpha^n_{l,m} - 1}{2}(\omega^n_{l-2,m} - \omega^n_{l,m}) + \frac{(\alpha^n_{l,m} - 1)^2}{2}(\omega^n_{l-2,m} - 2\omega^n_{l-1,m} + \omega^n_{l,m})$$

and[2]

$$\omega^{n+1}_{l,m} = \tilde{\omega}_{l,m-1} + \frac{\beta^n_{l,m} - 1}{2}(\tilde{\omega}_{l,m-2} - \tilde{\omega}_{l,m}) + \frac{(\beta^n_{l,m} - 1)^2}{2}(\tilde{\omega}_{l,m-2} - 2\tilde{\omega}_{l,m-1} + \tilde{\omega}_{l,m}). \tag{14}$$

Now Eqs. (14) are a perfectly valid difference procedure and may be derived by interpolation upstream. The physical interpretation, inherent in the given form, anticipates the leading phase which would not otherwise be self-evident. The magnitude of the phase errors are related in their distribution to those of Eq. (6). The dissipation likewise is closely related to that of the system (6).

To take advantage of now having two difference forms of opposite phase error we make a linear combination of (6) and (14). While any linear combination will improve the phase error a simple average will probably give the best accuracy. Thus, for the simple average we obtain for α and $\beta > 0$

$$\tilde{\omega}_{l,m} = \omega^n_{l,m} + \frac{\alpha^n_{l,m}}{4}(\omega^n_{l-1,m} - \omega^n_{l+1,m} + \omega^n_{l-2,m} - \omega^n_{l,m}) + \frac{(\alpha^n_{l,m})^2}{4}(\omega^n_{l-1,m} - 2\omega^n_{l,m} + \omega^n_{l+1,m}) + \frac{(\alpha^n_{l,m})^2 - 2\alpha^n_{l,m}}{4}(\omega^n_{l-2,m} - 2\omega^n_{l-1,m} + \omega^n_{l,m})$$

and

$$\omega^{n+1}_{l,m} = \tilde{\omega}_{l,m} + \frac{\beta^n_{l,m}}{4}(\tilde{\omega}_{l,m-1} - \tilde{\omega}_{l,m+1} + \tilde{\omega}_{l,m-2} - \tilde{\omega}_{l,m}) + \frac{(\beta^n_{l,m})^2}{4}(\tilde{\omega}_{l,m-1} - 2\tilde{\omega}_{l,m} + \tilde{\omega}_{l,m+1}) + \frac{(\beta^n_{l,m})^2 - 2\beta^n_{l,m}}{4}(\tilde{\omega}_{l,m-2} - 2\tilde{\omega}_{l,m-1} + \tilde{\omega}_{l,m}) \tag{15}$$

[2] Note for example if $\alpha < 0$ then replace $(\alpha^n_{l,m} - 1)$ by $(\alpha^v_{l,m} + 1)$ and permute l indices on the right of the first equation by $+2$. This gives a backward-in-time difference approximation for an $\alpha = -1$ projection of the solution.

Again we will consider a two-dimensional analysis because there exists a related form which is identical to (15) in one dimension but is unstable in two dimensions. The relation appropriate for analyzing the amplitude and phase properties of (15) is

$$r = \tilde{r}_{l,m} + \frac{\beta}{4}(\tilde{r}_{l,m-1} - \tilde{r}_{l,m+1} + \tilde{r}_{l,m-2} - \tilde{r}_{l,m}) + \frac{\beta^2}{4}(\tilde{r}_{l,m-1} - 2\tilde{r}_{l,m} + \tilde{r}_{l,m+1})$$

$$+ \frac{\beta^2 - 2\beta}{4}(\tilde{r}_{l,m-2} - 2\tilde{r}_{l,m-1} + \tilde{r}_{l,m}), \tag{16}$$

where

$$\tilde{r}_{l,m} = 1 - \frac{i\alpha}{2}\sin k_x \Delta x - \frac{i\alpha}{2}(\cos k_x \Delta x - i \sin k_x \Delta x)\sin k_x \Delta x$$

$$+ \frac{\alpha^2}{2}(\cos k_x \Delta x - 1)$$

$$+ \frac{\alpha^2 - 2\alpha}{2}(\cos k_x \Delta x - i \sin k_x \Delta x)(\cos k_x \Delta x - 1),$$

$$\tilde{r}_{l,m+1} = (\cos k_y \Delta y + i \sin k_y \Delta y)\tilde{r}_{l,m},$$

$$\tilde{r}_{l,m-1} = (\cos k_y \Delta y - i \sin k_y \Delta y)\tilde{r}_{l,m},$$

and

$$\tilde{r}_{l,m-2} = (\cos 2k_y \Delta y - i \sin 2k_y \Delta y)\tilde{r}_{l,m}.$$

In Fig. 3 we display the squared amplitude properties of (15) corresponding to Fig. 1 for (6). The statistics of Fig. 3 are included in Table I. We note, in particular, that the minima in amplification (decay) have been moved to $\alpha = \beta = \frac{1}{2}$ and that the minima represent slightly less decay except for the case $\Theta = \pi$. The improvement near the minimum is negligible, but nevertheless the improvement elsewhere is a bonus since we are primarily seeking a reduction in in phase errors. The increased decay for $\Theta = \pi$ is also desirable as has been pointed out above.

In Fig. 4 we give contour plots of phase errors incurred with (15), to be compared with Fig. 2 for Eq. (6). Of particular significance is that, in addition to having zero phase error at α and (or) $\beta = 1$, we have zero phase error for $(\alpha + \beta) = 1$, for $(\alpha - \beta) \approx \frac{1}{2}$ and for $(\beta - \alpha) \approx \frac{1}{2}$. This is true for all wave numbers except $\Theta = \pi$, which we discard for reasons previously indicated. The maximum phase error has been reduced by about a factor of 10 with the largest and most significant improvement at the high wave numbers.

In Fig. 5 we compare the short wavelength phase error magnitude of the zero average phase error method with the unmodified second-order method and the fourth-order method of Crowley. The very sharp delineation of the mode $\Theta = \pi$

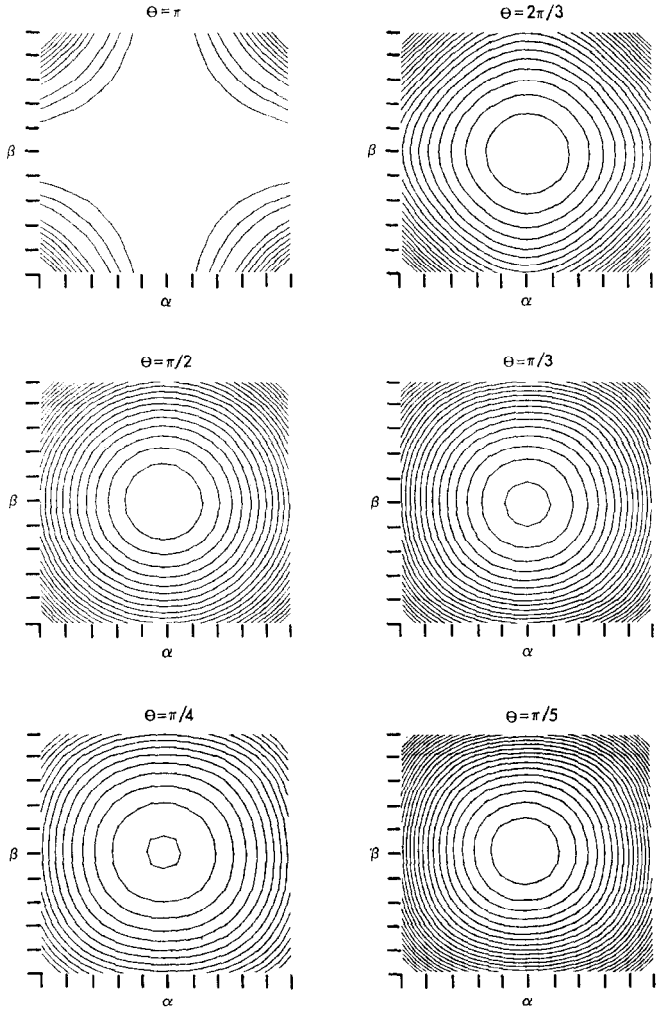

FIG. 3. Contour plots of the square of the modulus of amplification for the second-order, zero average phase error method of Eq. (15). See Table I.

is evident in the zero average phase error result. With the fourth-order method there is still a large error for $\Lambda = 4\Delta x$, however, the improvement over second order is considerable.

In Tables I and II we have included properties of Crowley's fourth-order method with the second-order results here obtained. Damping is less in the fourth-order case. One expects that application of the method of zero average phase error to

fourth-order approximations is appropriate for a reduction in damping but not for a significant improvement in phase error at high wave numbers.

Data of comparison tests under uniform flow conditions are available in movie form as are the data that has been reduced by contour plots. For nonlinear application one knows that (6) (see [5]) and, hence (14), are valid. Therefore, the

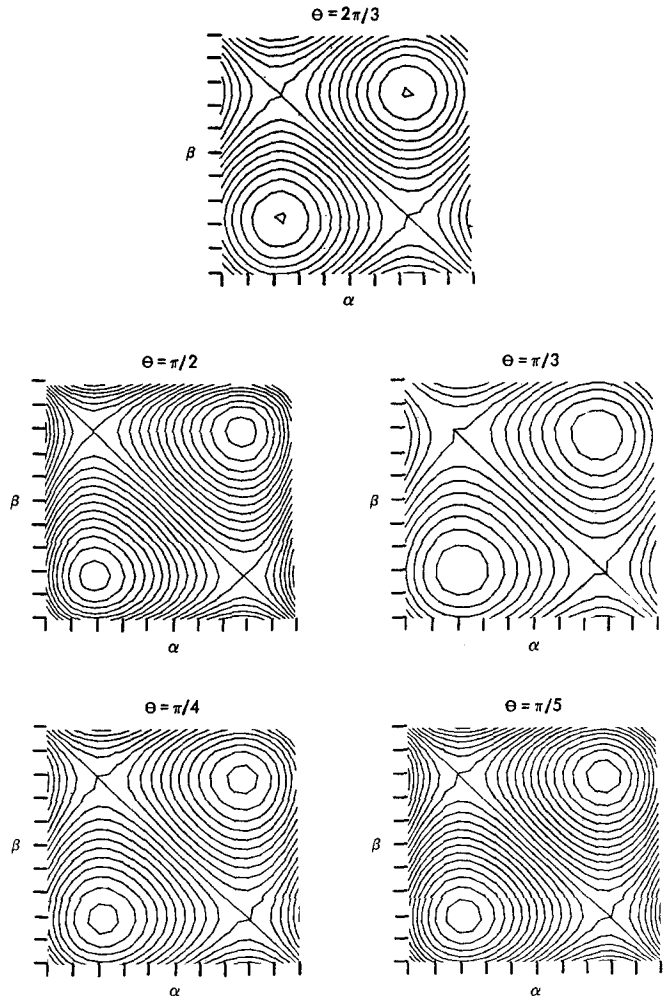

FIG. 4. Contour plots of phase error, $\Delta\phi$, in radians per time step Δt for the second-order, zero average phase error method of Eq. (15). See Table II. The missing contour line for $\Theta = \pi/3$ is the result of an improbable ambiguity of a pure zero saddle point intersected by a pure zero contour line.

linear combination of (6) and (14) is also valid. The means of programming the method in conservative form and considerations that must be given to boundary conditions are the subject of a future paper.

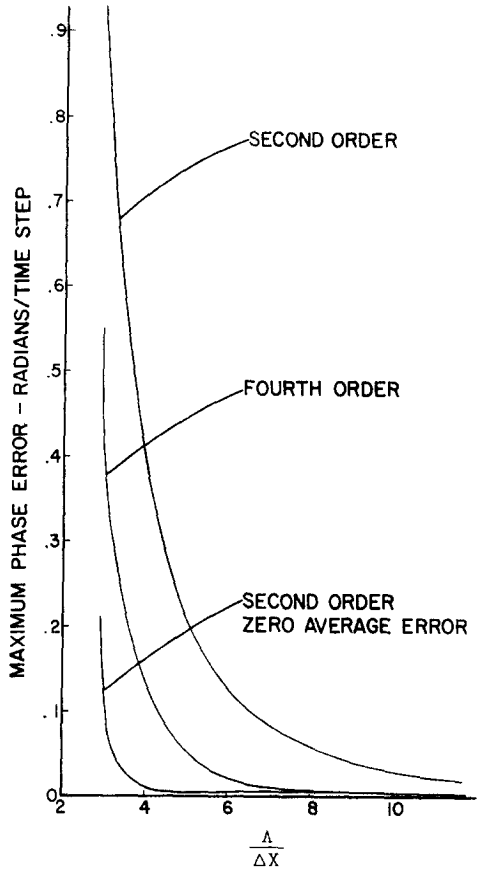

FIG. 5. Comparison of the second-order, zero average phase error method with second- and fourth-order unmodified fractional time step methods.

IV. CONCLUSION

It has been shown that it is possible to design a difference approximation in a rational way to reduce phase error. This is in contrast to extensions to higher order which improve phase errors less dramatically. The method has been shown to be applicable to the "fractional time step" forms and makes use of the property

of grid point to grid point transfer of values without dispersion or dissipation. In the present report we have applied the technique to a second-order form. The resulting phase properties are a considerable improvement over unmodified fourth-order forms in the important high wave number region.

ACKNOWLEDGMENTS

The author wishes to express his thanks to Donald E. Schreiber for the development and implementation of the contour graphics package for the IBM 2250, which was indispensible in the work.

REFERENCES

1. K. ROBERTS and N. WEISS, *Math. Computations* **20**, 272 (1966).
2. P. LAX and B. WENDROFF, *Commun. Pure Appl. Math.* **17**, 381 (1964).
3. S. BURSTEIN, *J. Computational Phys.* **1**, 198 (1967).
4. G. I. MARCHUK, *Dokl. Akad. Nauk SSSR* **155**, 1062 (1964).
5. C. E. LEITH, "Methods of Computational Physics", Vol. 4. Academic Press, New York (1965).
6. W. P. CROWLEY, *J. Computational Phys.* **1**, 471 (1967).
7. W. P. CROWLEY, *Monthly Weather Rev.* **96**, 1 (1968).