JOURNAL OF COMPUTATIONAL PHYSICS 91, 424–440 (1990)

The Modified Equation Approach to Flux-Corrected Transport

N. GRANDJOUAN

Laboratoire de Physique des Milieux Ionisés, Laboratoire du CNRS, Ecole Polytechnique, 91128 Palaiseau Cedex, France

Received July 31, 1989; revised October 25, 1989

For convective problems, the "modified equation" can be considered as the actual partial differential equation solved by a given numerical scheme using finite differences. Such an expression characterizes the dissipative and dispersive properties of the scheme. Adjusting the parameters of flux-corrected-transport (FCT) algorithms to cancel the successive truncation terms in the modified equation can be used in place of Fourier analysis when the velocity is no longer constant and uniform. This technique is used to propose a time-centered FCT algorithm in which diffusion/antidiffusion coefficients are velocity gradient dependent and which has reduced diffusion and noise level. © 1990 Academic Press, Inc.

INTRODUCTION

The flux-corrected-transport (FCT) technique [1] is a way to design convective algorithms which are highly accurate in regions with smooth gradients and are "sufficiently" diffusive around strong gradients and shocks. Different FCT schemes have been proposed [2, 3]. However, in most cases, Fourier analysis (restricted to a uniform and constant velocity) was used to determine their dissipative and dispersive properties. Parasite effects due to space discretization in the presence of velocity gradients were approached with the introduction of ZIP fluxes [4].

Displaying space *and* time errors of a numerical scheme, the "modified equation" technique is a way to answer the question: What analytic equation is actually being solved numerically? Hirt [5] opened the track in 1968 with a heuristic analysis; then Warming and Hyett [6] established the connections between the modified equation method and the von Neuman (Fourier) method when the velocity is constant. More recently, Lerat and Peyret [7–9] developed a method for use with coupled hydrodynamic systems.

In this paper, we use the modified equation to optimize the diffusion–antidiffusion coefficients of two FCT algorithms. This method corroborates the conclusions of Fourier analysis when the velocity is constant and uniform. For a more general velocity, the modified equation method, which can still be used as opposed to Fourier analysis, emphasizes differences between numerical schemes that give almost identical results for passive uniform convection.

0021-9991/90 $3.00

In Section 1, the classical ETBFCT scheme [10, 12] is examined, first by Fourier analysis, and in more detail, through its modified equation. This widespread algorithm [11, 12], when used in classical hydrodynamics, can turn smooth parts of a solution into a succession of steps. A first possible explanation of the staircase appearance of the solutions is the slight linear instability of the unlimited scheme, which is demonstrated by Fourier analysis. In inhomogeneous velocity fields, the fourth order accuracy of the scheme drops to second order: this situation is examined through the modified equation. This method underlines the origin of the dominant remaining truncation errors and suggests the use of an intermediate solution evaluated at half step.

In Section 2, a new time-centered FCT scheme is proposed, examined, and tested. First, its linear properties are examined by Fourier analysis for a uniform velocity. Then, the modified equation method is used to adjust the diffusion–antidiffusion coefficients for minimum truncations errors in the case of an inhomogeneous velocity field. The optimal coefficients are found to be velocity gradient dependent. The lower dissipation and the reduced noise level of this time-centered FCT algorithm are illustrated by simulations of a shock tube problem.

1. Modified Equation of the ETBFCT Scheme

ETBFCT blends three fluxes computed at the cell interfaces. The first flux (f^c) describes convection. It has second order space accuracy. The second (f^d) is a strong diffusion and the third (f^a), the antidiffusion, is "corrected" by the limiter. The main effect of the latter flux is to cancel the diffusion whenever this does not destroy the positivity of the solution. Carefully designed, it can also increase the precision of the whole scheme. We shall refer to the "limited" (or complete) algorithm when the nonlinear flux limiter controls the antidiffusion and to the "unlimited" scheme when the flux limiter is removed. We shall concentrate on the latter situation, which is the only one accessible to a linear or linearized analysis and which already brings a clear insight into many aspects of the complete scheme.

The ETBFCT finite difference approximation of the 1D Cartesian continuity equation,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \tag{1.1}$$

can be written

$$\frac{\rho_i^{n+1} - \rho_i^n}{\delta t} = -\frac{1}{\delta x}\left[f^c_{i+1/2} - f^c_{i-1/2} - f^d_{i+1/2} + f^d_{i-1/2} + C_{i+1/2} f^a_{i+1/2} - C_{i-1/2} f^a_{i-1/2}\right], \tag{1.2}$$

where δx and δt are the grid spacing and time step, ρ_i^n is the value of the solution in cell i at time $n\delta t$, and ρ_i^{n+1} is the value at the end of the current time step, at time $(n+1)\,\delta t$. The coefficients $C_{i+1/2}$ represent the effect of the nonlinear flux limitation and their values are in the range (0, 1).

If we use $w_{i+1/2}$ to denote the arithmetic mean of any quantity w in cells i and $(i+1)$, the fluxes are defined as

$$f^{\mathrm{c}}_{i+1/2} = \mathbf{u}_{i+1/2}\left[\frac{\rho^n_{i+1}+\rho^n_i}{2}\right], \qquad f^{\mathrm{d}}_{i+1/2} = v_{i+1/2}\left[\frac{\rho^n_{i+1}-\rho^n_i}{\delta x}\right],$$
$$f^{\mathrm{a}}_{i+1/2} = \mu_{i+1/2}\left[\frac{\rho^t_{i+1}-\rho^t_i}{\delta x}\right], \tag{1.3}$$

where v and μ are adjustable coefficients. The "transported" solution ρ^t is an approximate value of ρ^{n+1} given by

$$\rho^t_i = \rho^n_i - \frac{\delta t}{\delta x}\,[f^{\mathrm{c}}_{i+1/2} - f^{\mathrm{c}}_{i-1/2}].$$

To Fourier analyse this scheme, we consider, at time $n\,\delta t$ in cell i a perturbation of the form $\rho_i = \rho_0\, e^{j(ki\,\delta x)}$, where ρ_0 is the amplitude. Writing

$$\beta = k\,\delta x; \qquad a = 1-\cos\beta; \qquad b = \frac{2\,\delta t}{\delta x^2}; \qquad \varepsilon = \frac{\mathbf{u}\,\delta t}{\delta x}, \tag{1.4}$$

we obtain, for the amplification factor $G = \rho_0^{n+1}/\rho_0^n$ of the unlimited ETBFCT scheme,

$$|G|^2 = 1 + a[2\varepsilon^2 - 2b(v-\mu)] + a^2[\varepsilon^2(4\mu b - 1) + b^2(v-\mu)^2]$$
$$+ a^3[2\varepsilon^2\mu b(\mu b - 1)] - a^4(\varepsilon\mu b)^2. \tag{1.5}$$

If we now choose the coefficients v and μ to be [10, 12]

$$v = \frac{\delta x^2}{\delta t}\left[\frac{1}{6}+\frac{\varepsilon^2}{3}\right]; \qquad \mu = \frac{\delta x^2}{\delta t}\left[\frac{1}{6}-\frac{\varepsilon^2}{6}\right] \tag{1.6}$$

the squared module of the amplification factor G reduces to

$$|G|^2 = 1 + \frac{a^2\varepsilon^2}{3}(1-\varepsilon^2)\left[1 - \frac{2a}{3}(2+\varepsilon^2) - \frac{a^2}{3}(1-\varepsilon^2)\right]. \tag{1.7}$$

A contour plot of $|G|$ as a function of the Courant number ε (vertical axis) and of the reduced wavelength $1/\delta x$ of the modes (horizontal axis) is represented in Fig. 1a. It shows that the bracketed expression in Eq. (1.7) changes sign for $a \approx 0.6$ (for $\varepsilon < \frac{1}{2}$) and that $|G|$ is slightly greater than unity for $1/\delta x > 5$–6. This underlines the weak instability of the unlimited scheme for intermediate modes. The complete

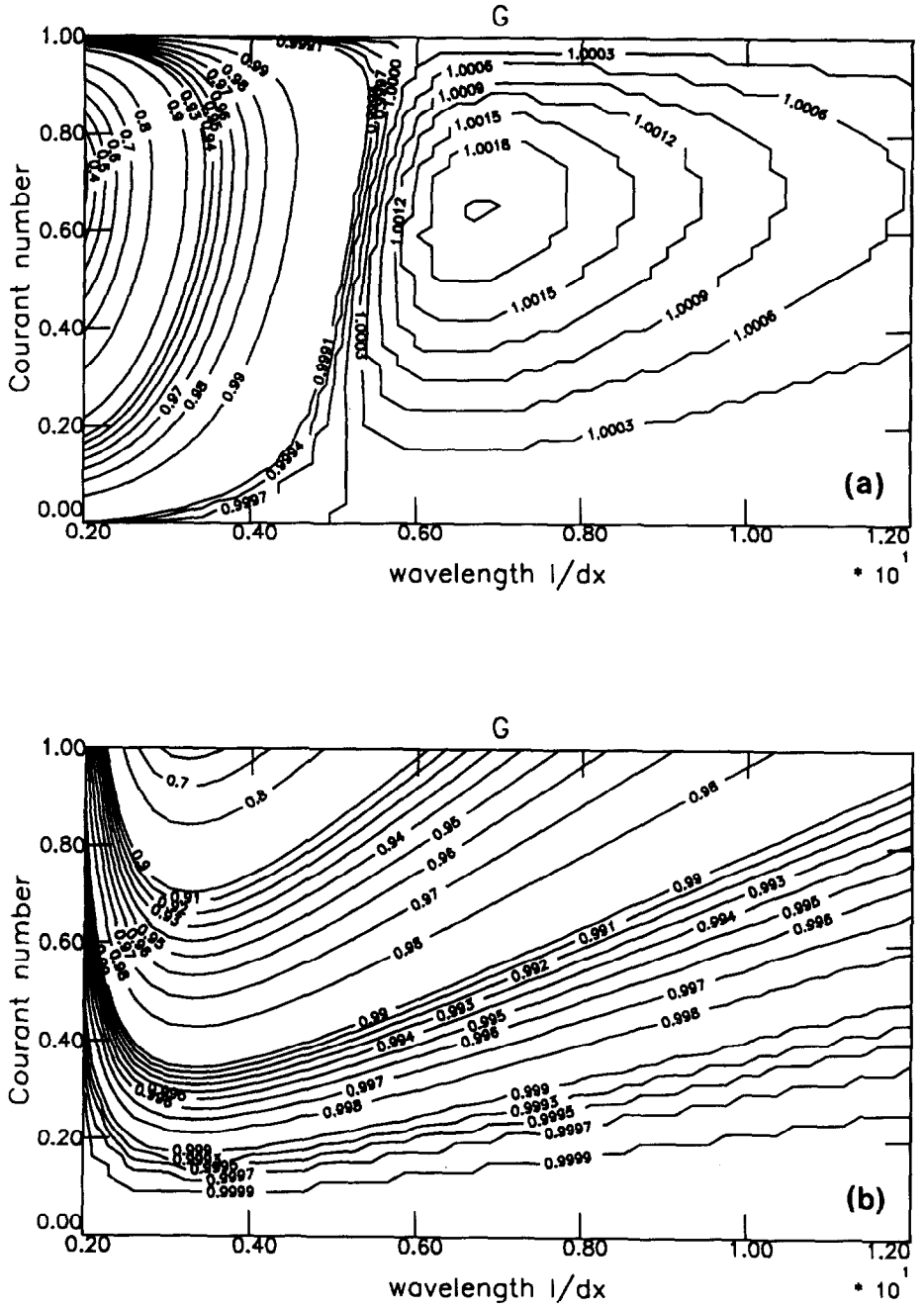

FIG. 1. Amplification factor for (a) ETBFCT and (b) time-centered FCT, as a function of Courant number ε (vertical axis) and reduced wavelength $1/\delta x$ of the modes (horizontal axis). Note that the amplification factor of time-centered FCT is everywhere less than unity.

scheme is stabilized by the properties of the flux limiter that prevents any point of the simulation from growing higher than its greater neighbour, or becoming smaller than its smaller neighbour. In smooth slopes, there is nothing to stop intermediate modes from slowly growing until some local "terraces" are created and then controlled by the limiter.

To derive the modified equation of the ETBFCT scheme, we follow the method described by Warming and Hyett [6] for a constant velocity, but here we keep the time and space derivatives of $\mathbf{u}$ in the modified equation.

The unlimited scheme (1.2), with the flux definitions of Eqs. (1.3), can be written as the sum of three terms,

$$\begin{aligned}
&\text{(I)} \quad \frac{\rho_i^{n+1}-\rho_i^n}{\delta t}+\frac{1}{\delta x}\left[f^c_{i+1/2}-f^c_{i-1/2}\right] \\
&\text{(II)} \quad -\frac{1}{\delta x}\left[(\nu-\mu)_{i+1/2}\frac{\rho^n_{i+1}-\rho^n_i}{\delta x}-(\nu-\mu)_{i-1/2}\frac{\rho^n_i-\rho^n_{i-1}}{\delta x}\right] \qquad (1.8)\\
&\text{(III)} \quad -\frac{\delta t}{\delta x^2}\left[\mu_{i+1/2}(f^c_{i+3/2}-2f^c_{i+1/2}+f^c_{i-1/2})\right. \\
&\qquad \left.-\mu_{i-1/2}(f^c_{i+1/2}-2f^c_{i-1/2}+f^c_{i-3/2})\right]=0,
\end{aligned}$$

where antidiffusive fluxes have been split into terms (II) and (III). Term (II), including only the second order derivative of ρ, is dissipative, while the third order derivative of the convective flux gives dispersive properties to term (III). Coefficients ν and μ have not yet been assigned values.

In a first step, each term of the finite difference approximation (1.8) is expanded in a Taylor series around the point $(i\,\delta x, n\,\delta t)$. In this analysis we use the notation $\dot{}$ (dot) for time derivatives, and $'$ (prime) for space derivatives.

Term (I) of Eq. (1.8) is the sum of two terms, (I) $= T + C$, which can be expanded as

$$T=\dot{\rho}+\frac{\delta t}{2}\ddot{\rho}+\frac{\delta t^2}{6}\dddot{\rho}+O(\delta t^3) \qquad (1.9)$$

and

$$C=(\rho\mathbf{u})'+\delta x^2\left[\frac{\rho'''\mathbf{u}}{6}+\frac{\rho''\mathbf{u}'}{4}+\frac{\rho'\mathbf{u}''}{4}+\frac{\rho\mathbf{u}'''}{6}\right]+O(\delta x^4). \qquad (1.10)$$

T and C are composed of a first term, which appears in the original equation, plus a truncation error. We shall denote as ETC the bracketed expression appearing in the truncation error of C:

$$\mathrm{ETC}=\left[\frac{\rho'''\mathbf{u}}{6}+\frac{\rho''\mathbf{u}'}{4}+\frac{\rho'\mathbf{u}''}{4}+\frac{\rho\mathbf{u}'''}{6}\right]. \qquad (1.11)$$

Expanding terms (II) and (III) of (1.8) in series yields

$$(\mathrm{II}) = -[(\nu-\mu)\rho']' + O(\alpha\,\delta x^2)$$
$$(\mathrm{III}) = -\delta t[\mu(\rho\mathbf{u})'']' + \delta t O(\mu\,\delta x^2).$$

Collecting terms of same order in (I), (II) and (III), with notations

$$\sigma = \frac{\delta t}{\delta x}; \qquad \Phi = \frac{\sigma}{\delta x}(\nu-\mu); \qquad \Psi = \frac{\sigma}{\delta x}\mu, \tag{1.12}$$

gives the expression of the modified equation

$$\dot\rho + (\rho\mathbf{u})' + \delta x\left[\frac{\sigma}{2}\ddot\rho - \frac{1}{\sigma}(\Phi\rho')'\right] + \delta x^2\left[\mathrm{ETC} + \frac{\sigma^2}{6}\dddot\rho - [\Psi(\rho\mathbf{u})'']'\right] + O(\delta x^3) = 0. \tag{1.13}$$

Then, in a second step, high order time derivatives are expressed in terms of space derivatives by successive differentiation of the modified equation. For instance, the second order time derivative is computed by multiplying Eq. (1.13) by $-\sigma\,\delta x/2$ and differentiating with respect to time. The resulting expression,

$$-\delta x\left[\frac{\sigma}{2}\ddot\rho + \frac{\sigma}{2}[(\rho\mathbf{u})']^{\cdot}\right] - \delta x^2\left[\frac{\sigma^2}{4}\ddot\rho - \frac{1}{2}(\Phi\rho')'^{\cdot}\right] + O(\delta x^3) = 0,$$

when added to Eq. (1.13) reads

$$\dot\rho + (\rho\mathbf{u})' + \delta x\left[-\frac{\sigma}{2}(\rho\mathbf{u})'^{\cdot} - \frac{1}{\sigma}(\Phi\rho')'\right] + \delta x^2\left[\mathrm{ETC} - \frac{\sigma^2}{12}\dddot\rho - [\Psi(\rho\mathbf{u})'']' + \frac{1}{2}(\Phi\rho')'^{\cdot}\right] + O(\delta x^3) = 0. \tag{1.14}$$

The first order term containing a mixed second derivative is then computed by multiplication of the modified equation (1.13) by the velocity $\mathbf{u}$ and differentiation with respect to space. Finally, the modified equation can be written as

$$\dot\rho + (\rho\mathbf{u})' + \delta x\left[-\frac{\sigma}{2}(\rho\dot{\mathbf{u}})' + \frac{\sigma}{2}[\mathbf{u}(\rho\mathbf{u})']' - \frac{1}{\sigma}(\Phi\rho')'\right] + \delta x^2\left\{\left[\frac{\sigma^2}{3}\left[\frac{\rho\ddot{\mathbf{u}}}{4} - \frac{\dot{\mathbf{u}}}{3}(\rho\mathbf{u})' - \mathbf{u}(\rho\dot{\mathbf{u}})\right] + \frac{\dot\Phi\rho'}{2}\right]' + \mathrm{ETC} + \left[\mathbf{u}\left[\frac{\sigma^2}{3}[\mathbf{u}(\rho\mathbf{u})']' - \frac{1}{2}(\Phi\rho')'\right] - \left[\Psi + \frac{\Phi}{2}\right](\rho\mathbf{u})''\right]'\right\} + O(\delta x^3) = 0. \tag{1.15}$$

Important information can be derived from this expression. Let us show that adjusting the modified equation for the best accuracy is a way to determine coefficients ν and μ.

Let us first examine the simple situation where the velocity $\mathbf{u}$ is independent of space and time. In this case, Eq. (1.15) reduces to

$$\dot{\rho} + (\rho \mathbf{u})' + \delta x \left[\left[\frac{\sigma \mathbf{u}^2}{2} - \frac{\Phi}{\sigma} \right] \rho'' \right] + \delta x^2 \left[\mathrm{ETC} + \mathbf{u} \left[\frac{\sigma^2 \mathbf{u}^2}{3} - \Phi - \Psi \right] \rho''' \right] + O(\delta x^3) = 0. \tag{1.16}$$

Cancelling the first order term of the truncation error leads to

$$\Phi = \frac{\sigma^2 \mathbf{u}^2}{2}, \tag{1.17}$$

hence, with the definitions (1.12) of σ and Φ, to

$$\nu - \mu = \frac{\delta x^2}{\delta t} \frac{\varepsilon^2}{2}. \tag{1.18}$$

Note that this result is easily obtained in a Fourier analysis, adjusting ν and μ to bring the amplification factor $|G|$ as close as possible to unity (cancellation of term proportional to a in Eq. (1.5)).

Using the expression (1.17) of Φ and the expression (1.11) for ETC, and assuming a constant velocity, Eq. (1.16) yields

$$\dot{\rho} + (\rho \mathbf{u})' + \delta x^2 \left\{ \left[\frac{1}{6} - \frac{\sigma^2 \mathbf{u}^2}{6} - \Psi \right] \mathbf{u} \rho''' \right\} + O(\delta x^3) = 0. \tag{1.19}$$

Then, again, cancelling the second order term of the truncation error leads to

$$\Psi = \frac{1}{6}(1 - \sigma^2 \mathbf{u}^2) \Rightarrow \mu = \frac{\delta x^2}{\delta t} \left(\frac{1}{6} - \frac{\varepsilon^2}{6} \right). \tag{1.20}$$

In this analysis we have first adjusted Φ (Eq. (1.17)) to suppress the coefficient of the second order space derivative in the truncation error, thus removing its *dissipative* (or destabilizing if positive) effect [5, 6]. The remaining *dispersive* term (proportional to the third order space derivative) is then cancelled by adjustment of Ψ (Eq. (1.20)). It is not surprising that Eqs. (1.18) and (1.20) yield to the same determination (1.6) of coefficients ν and μ, whose values can be obtained, following Boris and Book [3], through a Fourier analysis where both *amplitude* and *phase* errors are minimized for long wavelengths.

When the velocity is no longer constant, the modified equation method can still be used, as opposed to Fourier analysis. It is there fore a convenient tool with which to study the truncation errors of ETBFCT. The classical use of this scheme

has been described by Boris and Book [2, 10]. They recommend estimating the velocity $\mathbf{u}(t+\delta t/2)$ by integration over a half step, and using it in the full step, for computation of the convective fluxes and coefficients ν and μ. The modified equation of ETBFCT with this time-centered velocity is

$$\dot{\rho}+(\rho\mathbf{u})'+\delta x\left[\frac{\sigma}{2}[\mathbf{u}(\rho\mathbf{u})']'-\frac{1}{\sigma}(\Phi\rho')'\right]$$
$$+\delta x^2\left\{\frac{\sigma^2}{12}\left[\dot{\mathbf{u}}(\rho\mathbf{u})'-\mathbf{u}(\rho\dot{\mathbf{u}})'-\frac{1}{2}\rho\ddot{\mathbf{u}}\right]'\right.$$
$$\left.+\text{ETC}+\left[\mathbf{u}\left[\frac{\sigma^2}{3}[\mathbf{u}(\rho\mathbf{u})']'-\frac{1}{2}(\Phi\rho')'\right]-\left[\Psi+\frac{\Phi}{2}\right](\rho\mathbf{u})''\right]'\right\}+O(\delta x^3)=0. \quad (1.21)$$

Note, with comparison of Eq. (1.15), that the time derivative of the velocity $\mathbf{u}$ has cancelled out in the first order term of the truncation error.

If we now look at the case where $\mathbf{u}$ is a function of space, but still independent of time, the modified equation displays the truncation errors of ETBFCT for a stationary but inhomogeneous velocity field. Dropping the time dependence of $\mathbf{u}$ in Eq. (1.21) and replacing ν and μ (Φ and Ψ) by their classical definition (1.6), it is easy to write the first order term as

$$\frac{\sigma\,\delta x}{2}[\rho(\mathbf{u}\mathbf{u}')]'.$$

Consequently, we can expect a noticeable dispersive effect in regions where both the velocity and its derivative are important. Furthermore, the second order term now contains an expression involving the second order space derivative of ρ, thus having a dissipative effect. It reads

$$\delta x^2\rho''\mathbf{u}'\left[\frac{\varepsilon^2}{3}-\frac{1}{4}\right]. \quad (1.22)$$

Using the second order ZIP form of the convective flux [4] does not help as this term changes only to

$$\delta x^2\rho''\mathbf{u}'\left[\frac{\varepsilon^2}{3}-\frac{1}{2}\right]. \quad (1.23)$$

Then, for both cases, this term is diffusive when $\mathbf{u}'>0$ (rarefaction) and destabilizing when $\mathbf{u}'<0$ (compression or shock), in the range of Courant numbers $\varepsilon<\frac{1}{2}$, required by ETBFCT to maintain positivity [3].

Simulation of the shock tube provides an illustration of this kind of errors. This Riemann problem has been simulated on a 120-point grid, using the limited ETBFCT algorithm. Initialization sets the cells length to unity, a density equal to one and a zero velocity everywhere. Temperature is 1000 in cells 1 to 50, and unity

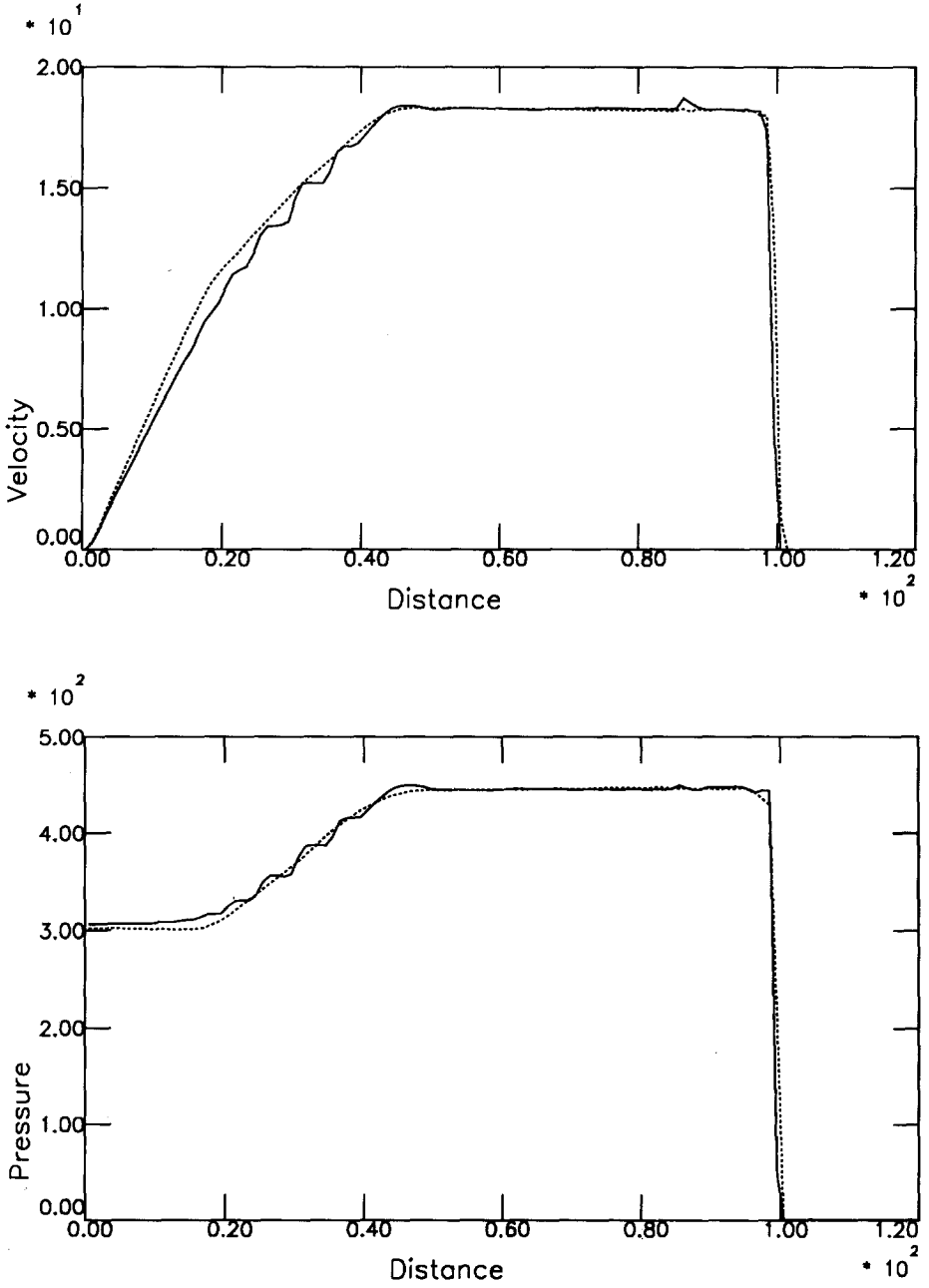

FIG. 2. Velocity and pressure profiles of the shock tube problem (see text). ETBFCT results are solid lines; time-centered FCT results are dotted lines. Note for time-centered FCT the reduced noise level and lowered dissipation in rarefaction regions.

in cells 51 to 120. A perfect gas law with $\gamma = \frac{5}{3}$ and reflective boundary conditions are used. The time step is limited by the condition: $|\mathbf{u} + |c_s| \leqslant 0.4\delta x/\delta t$, where c_s is the sound velocity, and a flux limiter given by the prescription of Zalesak, without its peak restitution [13], controls the antidiffusion. Figure 2 shows the velocity and pressure profiles when the right moving shock has reached cell number 100, and the left moving rarefaction fan has begun its reflection from the left boundary. ETBFCT results are drawn with solid lines; dotted lines show solutions obtained by an improved time-centered algorithm, which will be discussed in Section 2. ETBFCT solutions are noisy in the rarefaction region where the velocity and its derivative are important. In this region, the smooth slope inhibits the flux limiter until some steps are created. The second order diffusion (see Eq. (1.22)), can be seen, near cell number 18, by comparison with the results of the time-centered scheme (dotted lines), where this diffusion has been cancelled out (see Section 2). Surprisingly, velocity and pressure in ETBFCT solutions have a small bump at the contact discontinuity (near cell number 86) where they should be constant.

In Ref. [10], Boris described how to use ETBFCT for Eulerian hydrodynamics. The three equations of conservation (mass, momentum, and energy) are first solved by the complete algorithm, with time step $\delta t/2$, to compute mid-step solutions. Mid-step velocity and pressure are then estimated. These two time-centered quantities are then used in the full step algorithm to calculate the solution at $t + \delta t$. With such a procedure, the complete algorithm, including its flux limitation, is called twice for each equation. This double pass and the favorable effect of time centering the velocity have focused our interest toward a fully time-centered FCT scheme, which we describe in the following section.

2. Time-Centered FCT

The scheme is made of two time-centered FCT steps. Each of them includes a strong explicit diffusion that limits the time steps to Courant numbers $\varepsilon < \frac{1}{2}$ to ensure positivity. Therefore, to estimate the solutions at $t + \delta t/2$, we have chosen to begin with an interpolation between times t and $t - \delta t$, and to advance these values from $t - \delta t/2$ to $t + \delta t/2$ with a first time-centered FCT step. The alternative predictor technique that would compute solutions at $t + \delta t$ in a double length time step, and then interpolate back at $t + \delta t/2$, would have a more severe time step constraint.

Solutions are supposed to be known at times t and $t - \delta t$ and we denote by * (respectively $\tilde{}$) quantities evaluated at time $t - \delta t/2$ (respectively $t + \delta t/2$). ρ^* is computed at almost no cost, by linearly interpolating between ρ^n and ρ^{n-1}. Then, the solution $\tilde{\rho}$ at $t + \delta t/2$ is computed by a first FCT step with time-centered convective fluxes and time-centered diffusion–antidiffusion coefficients,

$$\frac{\tilde{\rho}_i - \overset{*}{\rho}_i}{\delta t} = -\frac{1}{\delta x}\left[f^c_{i+1/2} - f^c_{i-1/2} - \overset{*}{f}{}^d_{i+1/2} + \overset{*}{f}{}^d_{i-1/2} + \overset{*}{f}{}^a_{i+1/2} - \overset{*}{f}{}^a_{i-1/2}\right], \qquad (2.1)$$

where the fluxes are

$$f^{c}_{i+1/2} = \mathbf{u}_{i+1/2}\left[\frac{\rho^{n}_{i+1}+\rho^{n}_{i}}{2}\right]; \qquad \overset{*}{f}{}^{d}_{i+1/2} = v_{+1/2}\left[\frac{\overset{*}{\rho}_{i+1}-\overset{*}{\rho}_{i}}{\delta x}\right];$$
$$\overset{*}{f}{}^{a}_{i+1/2} = \mu_{i+1/2}\left[\frac{\overset{*}{\rho}{}^{t}_{i+1}-\overset{*}{\rho}{}^{t}_{i}}{\delta x}\right]. \tag{2.2}$$

These mid-step values are used in the convective fluxes and coefficients of the whole step as follows,

$$\frac{\rho^{n+1}_{i}-\rho^{n}_{i}}{\delta t} = -\frac{1}{\delta x}\left[\tilde{f}^{c}_{i+1/2}-\tilde{f}^{c}_{i-1/2}-f^{d}_{i+1/2}+f^{d}_{i-1/2}+f^{a}_{i+1/2}-f^{a}_{i-1/2}\right], \tag{2.3}$$

where

$$\tilde{f}^{c}_{i+1/2} = \tilde{\mathbf{u}}_{i+1/2}\left[\frac{\tilde{\rho}_{i+1}+\tilde{\rho}_{i}}{2}\right]; \qquad f^{d}_{i+1/2} = \tilde{v}_{i+1/2}\left[\frac{\rho^{n}_{i+1}-\rho^{n}_{i}}{\delta x}\right];$$
$$f^{a}_{i+1/2} = \tilde{\mu}_{i+1/2}\left[\frac{\rho^{t}_{i+1}-\rho^{t}_{i}}{\delta x}\right]. \tag{2.4}$$

Note that antidiffusions are calculated using "transported" values that are estimations of end-of-step solutions,

$$\overset{*}{\rho}{}^{t}_{i} = \overset{*}{\rho}_{i} - \frac{\delta t}{\delta x}\left[f_{i+1/2}-f_{i-1/2}\right], \qquad \rho^{t}_{i} = \rho^{n}_{i} - \frac{\delta t}{\delta x}\left[\tilde{f}_{i+1/2}-\tilde{f}_{i-1/2}\right]. \tag{2.5}$$

On the other hand, diffusions are computed from solutions at the beginning of the step and convections with mid-step values.

Let us Fourier analyse this scheme. Using definitions (1.4) of a and b and notations, we obtain

$$Z = \frac{\tilde{\rho}_{i+1}-\tilde{\rho}_{i-1}}{\rho^{n}_{i}}, \qquad U = \frac{\tilde{\rho}_{i+2}-\tilde{\rho}_{i-2}}{\rho^{n}_{i}}.$$

The amplification factor $G = \rho^{n+1}_{0}/\rho^{n}_{0}$ reads

$$G = \left[1-(v-\mu)ab\right] - \frac{\varepsilon Z}{2} + \mu\varepsilon\frac{b}{4}(U-2Z). \tag{2.6}$$

If we denote by H the ratio $\tilde{\rho}_{0}/\rho^{n}_{0}$, then

$$Z = 2jH\sin\beta, \qquad U = 2jH\sin 2\beta$$

and G changes to

$$G = \left[1-(v-\mu)ab\right] - jH\varepsilon\sin\beta\left[1+\mu ab\right]. \tag{2.7}$$

Let us suppose now, for a while, that the two steps have the same amplification factor G. Then

$$\tilde{\rho}_0 = G\rho^* = G\left[\frac{\rho_i^n + \rho_i^{n-1}}{2}\right] = \rho_0^n\left[\frac{G+1}{2}\right] \quad \text{and} \quad H = \frac{G+1}{2}. \tag{2.8}$$

The squared module of the amplification factor G is then

$$|G|^2 = \frac{[1-(v-\mu)ab]^2 + [\varepsilon/2 \sin \beta(1+\mu ab)]^2}{1 + [\varepsilon/2 \sin \beta(1+\mu ab)]^2}.$$

If one chooses equal diffusion/antidiffusion coefficients, the module of G is strictly unity. The numerical phase velocity is then

$$\begin{aligned} v_\varphi &= \frac{1}{\beta\varepsilon} \arctan[-\mathrm{Im}(G)/\mathrm{Re}(G)] \\ &= 1 + \beta^2\left[\frac{\mu b}{2} - \frac{1}{6} - \frac{\varepsilon^2}{12}\right] + \beta^4\left[\frac{1}{120} + \varepsilon^2\left(\frac{1}{24} - \frac{\mu b}{8}\right) + \frac{\varepsilon^4}{80} - \frac{\mu b}{8}\right]. \end{aligned}$$

Phase errors can be reduced by cancelling the second order term in v_φ, which completely determines v and μ:

$$v = \mu = \frac{\delta x^2}{\delta t}\left[\frac{1}{6} + \frac{\varepsilon^2}{12}\right]. \tag{2.9}$$

Without the simplifying hypothesis (2.8), the exact value of H is

$$H = \left[\frac{1}{2} + \frac{1}{2G}\right][1-(v-\mu)ab] - j\varepsilon \sin \beta[1+\mu ab] \tag{2.10}$$

and letting $X = 1-(v-\mu)ab$ and $Y = \varepsilon \sin \beta[1+\mu ab]$, G is now the root of

$$G^2 + G\left[Y^2 - X + j\frac{XY}{2}\right] + j\frac{XY}{2} = 0, \tag{2.11}$$

which has a positive propagation velocity.

We have plotted $|G|$ given by (2.11) in Fig. 1b, with the choice (2.9) for v and μ. It can be seen that $|G|$ is everywhere strictly less than unity; this scheme is linearly stable for all modes.

As with ETBFCT, and for the same reasons of positivity preservation, the time step is limited to Courant numbers $\varepsilon < \frac{1}{2}$. Within this range, we have run the classical test of the passively convected square wave [1]. Results obtained with ETBFCT and the present scheme exhibit no significant difference. For passive convection of more complicated profiles including short modes, it can be shown that our time-centered FCT is less noisy than ETBFCT. These conclusions for a constant velocity

are not unexpected: the two schemes have comparable fourth order phase errors, but, in contrast to ETBFCT, our time-centered FCT has an amplification factor strictly less than unity for all modes.

Let us proceed to the modified equation of this time-centered algorithm. First, we shall confirm the determination (2.9) of ν and μ when $\mathbf{u}$ is a constant. Then, we shall underline differences between the two schemes for inhomogeneous velocity fields.

It is easy to see that values at $t+\delta t/2$, obtained by the time-centered FCT step (2.1), from interpolated solutions at $t-\delta t/2$, can be written as

$$\tilde{\rho}_i = \rho + \frac{\delta t}{2}\dot{\rho} + \frac{\delta t^2}{4}\ddot{\rho} + O(\delta t^3)$$

$$\tilde{\mathbf{u}}_i = \mathbf{u} + \frac{\delta t}{2}\dot{\mathbf{u}} + \frac{\delta t^2}{4}\ddot{\mathbf{u}} + O(\delta t^3).$$

Using these expressions in flux definitions (2.4) and expanding each term of Eq. (2.3) in a Taylor series around the point $(n\,\delta t, i\,\delta x)$ lead to the modified equation. Colleting terms of same order yields

$$\dot{\rho} + (\rho\mathbf{u})' + \delta x\left[\frac{\sigma}{2}\ddot{\rho} - \frac{1}{\sigma}(\Phi\rho')' + \frac{\sigma}{2}(\rho\mathbf{u})^{\cdot}\right]$$
$$+ \delta x^2\left[\mathrm{ETC} + \frac{\sigma^2}{6}\ddot{\rho} - [\Psi(\rho\mathbf{u})'']' - \frac{1}{2}[\Phi^{\cdot}\rho']' + \frac{\sigma^2}{4}[\ddot{\rho}\mathbf{u} + \dot{\rho}\dot{\mathbf{u}} + \rho\ddot{\mathbf{u}}]'\right] + O(\delta x^3) = 0,$$

where ETC is still the value of Eq. (1.11). Replacing high order time derivatives by space derivatives provides some simplifications in the modified equation, which can be written as

$$\dot{\rho} + (\rho\mathbf{u})' - \frac{\delta x}{\sigma}(\Phi\rho')' + \delta x^2\left\{\mathrm{ETC} - \left[\left(\Psi + \frac{\Phi}{2}\right)(\rho\mathbf{u})'' - \frac{\sigma^2}{12}\mathbf{u}[\mathbf{u}(\rho\mathbf{u})']'\right]'\right.$$
$$\left. + \frac{\sigma^2}{12}[\rho(\ddot{\mathbf{u}} - \mathbf{u}\dot{\mathbf{u}}' + \dot{\mathbf{u}}\mathbf{u}')]'\right\} + O(\delta x^3) = 0. \tag{2.12}$$

As in the analysis of ETBFCT, restricting this expression to the case $\mathbf{u} = \text{const}$ is an alternative method of determining the diffusion and antififfusion coefficients. This modified equation then reads

$$\dot{\rho} + (\rho\mathbf{u})' - \frac{\delta x}{\sigma}(\Phi\rho')' + \delta x^2\left[\mathbf{u}\rho'''\left[\frac{1}{6} - \left(\Psi + \frac{\Phi}{2}\right) + \frac{\sigma^2\mathbf{u}^2}{12}\right]\right] + O(\delta x^3) = 0. \tag{2.13}$$

Cancelling the first and second order terms leads to $\Phi = 0$ $(\nu = \mu)$ and $\Psi = (1+\varepsilon^2/2)/6$ and confirms the determination (2.9) for coefficients ν and μ obtained from Fourier analysis.

Furthermore, extended use of the modified equation to inhomogeneous velocity cases is a fruitful approach to identifying and cancelling out nonlinear dissipative terms. Using the values (2.9) of ν and μ and definition (1.11) of ETC, and assuming a stationary but space dependent velocity, Eq. (2.12) yields

$$\dot{\rho} + (\rho \mathbf{u})' + \delta x^2 \left\{ \rho \left[\frac{\sigma^2 \mathbf{u}\mathbf{u}'\mathbf{u}''}{6} + \frac{\sigma^2 \mathbf{u}'^3}{12} \right] + \rho' \left[-\frac{\mathbf{u}''}{4} + \frac{\sigma^2 \mathbf{u}\mathbf{u}'^2}{4} + \frac{\sigma^2 \mathbf{u}^2 \mathbf{u}''}{12} \right] + \rho'' \left[-\frac{\mathbf{u}'}{4} + \frac{\sigma^2 \mathbf{u}^2 \mathbf{u}'}{12} \right] \right\} + O(\delta x^3) = 0. \tag{2.14}$$

The third term in the truncation error is proportional to the second derivative of ρ. It is now the main dissipative term, and one can derive corrections to coefficients ν and μ that will exactly cancel it out. If we set

$$\begin{aligned} \nu &= \frac{\delta x^2}{dt} \left\{ \left[\frac{1}{6} + \frac{\varepsilon^2}{12} \right] + \frac{\delta t \mathbf{u}'}{2} \left[-\frac{1}{4} + \frac{\varepsilon^2}{12} \right] \right\} \\ \mu &= \frac{\delta x^2}{\delta t} \left\{ \left[\frac{1}{6} + \frac{\varepsilon^2}{12} \right] - \frac{\delta t \mathbf{u}'}{2} \left[-\frac{1}{4} + \frac{\varepsilon^2}{12} \right] \right\} \end{aligned} \tag{2.15}$$

the first order term in (2.12) is no longer zero but has a second order value that exactly cancels out the dissipative term. The modified equation simply reduces to

$$\dot{\rho} + (\rho \mathbf{u})' + \frac{\sigma^2 \delta x^2}{12} [\rho \mathbf{u}\mathbf{u}'^2]' + O(\delta x^3) = 0. \tag{2.16}$$

The truncation error is considerably decreased and, at this order, no longer includes any dissipative term. Note, by comparison with analysis of ETBFCT, that the truncation error remains of second order.

The modified equation provides also a straightforward way of studying the role of ZIP convective fluxes [4]. If they are defined as

$$\mathcal{F}^{\text{zip}}_{i+1/2} = \tfrac{1}{2} [\tilde{\rho}_{i+1} \tilde{\mathbf{u}}_i + \tilde{\rho}_i \tilde{\mathbf{u}}_{i+1}]$$

the truncation error ETC does not include any first and second order space derivatives of the velocity. With these fluxes, the coefficient $-\frac{1}{4}$ in corrections (2.15) simply changes to $-\frac{1}{2}$, and the resulting modified equation is exactly the same as Eq. (2.16).

In Fig. 2, we have plotted, with dotted lines, results of a simulation of the same shock tube problem used in Section 1. They have been obtained using our limited time-centered FCT algorithm, with the convective fluxes in ZIP form, and the antidiffusion controlled by the same flux limiter used in Section 1, for ETBFCT simulations. Because solutions in the shock region (cell number 100) are strongly dependent on the properties of the flux limiter, they do not differ appreciably from ETBFCT calculations. Pressure and velocity profiles are now constant at the contact

discontinuity (around cell number 86). The main differences between the two algorithms can be seen in the rarefaction region (cell numbers 15 to 45). Time-centered FCT, which has been shown to be free of first order dispersion term, has smooth solutions (dotted lines), whereas ETBFCT results (solid lines) have a staircase appearance. Let us emphasize that it is not due to an increased diffusion. Further, the reflected rarefaction wave can be seen more clearly in time-centered FCT profiles (around cell 18). The small slope variation of the velocity gradient is partially smoothed out in ETBFCT results. This is due to the remaining second order diffusion and cumulative effects of the flux limiter controlling the first order dispersion errors.

Our test case was chosen to enter into the frame of the review paper by Woodward and Colella [11], which compares solutions of a shock tube problem computed by up-to-date numerical methods, run at a Courant number less than or equal to 0.4. The initial temperature jump of 1000 to 1 makes it a more servere test than the original benchmark of Ref. [10], which has a ratio of 10 to 1. Furthermore the Courant number $\varepsilon < 0.4$ is an unfavourable situation for ETBFCT, whose small linear instability amplifies any numerical noise. For completeness, we should have denoted the original scheme as "theoretical" ETBFCT and mentioned that it can be damped slightly [15] to eliminate this difficulty for small Courant numbers. This "optimized" ETBFCT is only described by its recently published Fortran listing (Appendix A of Ref. [12]), which is not the original ETBFCT listing [10]. Coefficients v and μ differ by a small amount from those given by Eq. (1.6). They are

$$v = \frac{\delta x^2}{\delta t}\,[0.167 + 0.333\,\varepsilon^2] \quad \text{or} \quad v = \frac{\delta x^2}{\delta t}\left[\frac{1}{6}(1 + 2.10^{-3}) + \frac{1}{3}(1 - 10^{-3})\varepsilon^2\right]$$

$$\mu = \frac{\delta x^2}{\delta t}\left[0.25 - 0.5\,\frac{\delta t}{\delta x^2}\,v\right] \quad \text{or} \quad \mu = \frac{\delta x^2}{\delta t}\left[\frac{1}{6}(1 - 10^{-3}) - \frac{1}{6}(1 - 10^{-3})\varepsilon^2\right].$$

The linear amplification factor G of this "optimized" algorithm is represented in Fig. 3. It shows that the Courant number $\varepsilon = 0.2$ is an optimal value for this version of ETBFCT, which has, for long wavelengths, an amplification factor equal to unity, the ideal value. The solutions of our shock tube problem, run with this version of ETBFCT at a Courant number less than 0.2, are almost identical to the results of our time-centered scheme, run at a Courant number less than 0.4. For Courant numbers greater than 0.2, the "optimized" ETBFCT cannot completely damp the nonlinear dispersive errors. Our time-centered FCT, having a reduced noise generation and being linearly stable for all $\varepsilon < 0.5$, can be run at Courant numbers approaching this limit, therefore reducing the computing cost.

Concluding Remarks

In deriving our time-centered scheme, we have not considered FCT as the blending of two convective fluxes, a first order one and a higher order one. We have

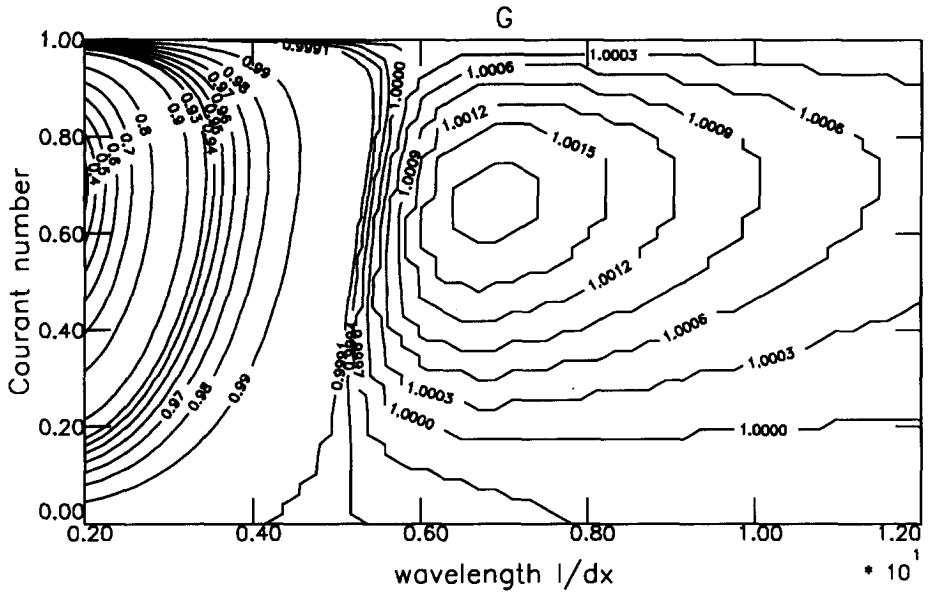

FIG. 3. Amplification factor for "optimized" ETBFCT (see text) as in Fig. 1. $G = 1$ for $\varepsilon = 0.2$ and large wavelength makes this Courant number an optimum for this version of the scheme.

kept the original three-flux formulation: convection, diffusion, and antidiffusion. Computing diffusion and antidiffusion at different times introduces third order space derivatives in the modified equation and gives an opportunity to cancel out both time and space dispersion errors through velocity dependent coefficients. It has been shown [14] that two-flux schemes based on very high order convective fluxes can only deal with the dispersion due to space discretization.

Finally, we want to point out that our time-centered FCT algorithm is independent of any specific system of partial differential equations. This property is obtained in our derivation of coefficients ν and μ (Eq. (2.15)), neglecting the time derivatives of the velocity in the modified equation (2.12). Estimating those would require using the momentum equation and would introduce corrections dependent on the specific system of fluid equations. This time-centered scheme is therefore a general purpose tool and we plan to use it to simulate MHD convective effects, where the effective velocities are strongly inhomogeneous.

ACKNOWLEDGMENTS

The author acknowledges helpful discussions with A. Hemon, S. Huberson, A. Lerat, and J. Ovadia. It is a pleasure to thank J. Virmont for his patient support and C. Chenais-Popovics for carefully reading the manuscript.

REFERENCES

1. J. P. BORIS AND D. L. BOOK, *J. Comput. Phys.* **11**, 38 (1973).
2. D. L. BOOK, J. P. BORIS, AND K. HAIN, *J. Comput. Phys.* **18**, 248 (1975).
3. J. P. BORIS AND D. L. BOOK, *J. Comput. Phys.* **20**, 397 (1976).
4. S. T. ZALESAK, *J. Comput. Phys.* **40**, 497 (1981).
5. C. W. HIRT, *J. Comput. Phys.* **2**, 339 (1968).
6. R. F. WARMING AND B. J. HYETT, *J. Comput. Phys.* **14**, 159 (1974).
7. A. LERAT AND R. PEYRET, *C. R. Acad. Sci. Paris Sér. A* **276**, 759 (1973).
8. R. PEYRET, Office National d'Etudes et de Recherche Aérospatiales, Publication No. 1977-5 (unpublished).
9. A. LERAT, Thésis, Université Pierre et Marie Curie, Paris VI, Fév. 1981 (unpublished).
10. J. P. BORIS, Naval Research Laboratory, Memo. Report 3237 (March 1976) (unpublished).
11. P. R. WOODWARD AND P. COLELLA, *J. Comput. Phys.* **54**, 115 (1984).
12. D. L. BOOK, J. P. BORIS, AND S. T. ZALESAK, "Flux-Corrected Transport," in *Finite Difference Techniques for Vectorized Fluid Dynamics Calculations*, edited by D. L. Book (Springer-Verlag, New York, 1981).
13. S. T. ZALESAK, Naval Research Laboratory, Memo. Report 3716 (May 1978) (unpublished).
14. S. T. ZALESAK, *Advances in Computer Methods for Partial Differential Equations, V*, edited by R. Vichnevetsky and R. S. Stepleman (IMACS, Rutgers University, 1981).
15. J. P. BORIS, private communication (1989).