JOURNAL OF COMPUTATIONAL PHYSICS 57, 327–360 (1985)

# Implicit Total Variation Diminishing (TVD) Schemes for Steady-State Calculations

H. C. YEE AND R. F. WARMING

*NASA Ames Research Center, Moffett Field, California*

AND

A. HARTEN

*Tel Aviv University, Tel Aviv, and New York University, New York*

Received August 25, 1983

The application of a new implicit unconditionally stable high-resolution TVD scheme to steady-state calculations is examined. It is a member of a one-parameter family of explicit and implicit second-order accurate schemes developed by Harten for the computation of weak solutions of one-dimensional hyperbolic conservation laws. This scheme is guaranteed not to generate spurious oscillations for a nonlinear scalar equation and a constant coefficient system. Numerical experiments show that this scheme not only has a fairly rapid convergence rate, but also generates a highly resolved approximation to the steady-state solution. A detailed implementation of the implicit scheme for the one- and two-dimensional compressible inviscid equations of gas dynamics is presented. Some numerical computations of one- and two-dimensional fluid flows containing shocks demonstrate the efficiency and accuracy of this new scheme. 

## 1. INTRODUCTION

Conventional shock capturing schemes for the solution of nonlinear hyperbolic conservation laws are linear and $L_2$-stable (stable in the $L_2$-norm) when considered in the constant coefficient case [1]. There are three major difficulties in using such schemes to compute discontinuous solutions of a nonlinear system, such as the compressible Euler equations:

(i) Schemes that are second (or higher) order accurate may produce oscillations wherever the solution is not smooth.

(ii) Nonlinear instabilities may develop in spite of the $L_2$-stability in the constant coefficient case.

(iii) The scheme may select a nonphysical solution.

It is well known that monotone conservative difference schemes always converge and that their limit is the physical weak solution satisfying an entroy inequality.



0021-9991/85 $3.00

Thus monotone schemes are guaranteed not to have difficulties (ii) and (iii). However, monotone schemes are only first-order accurate. Consequently, they produce rather crude approximations whenever the solution varies strongly in space or time.

When using a second- (or higher) order accurate scheme, some of these difficulties can be overcome by adding a hefty amount of numerical dissipation to the scheme. Unfortunately, this process brings about an irretrievable loss of information that exhibits itself in degraded accuracy and smeared discontinuities. Thus, a typical complaint about conventional schemes which are developed under the guidelines of linear theory is that they are not robust and/or not accurate enough.

To overcome the difficulties, we consider a new class of schemes that is more appropriate for the computation of weak solutions (i.e., solutions with shocks and contact discontinuities) of nonlinear hyperbolic conservation laws. These schemes are required (a) to be total variation diminishing in the nonlinear scalar case and the constant coefficient system case [2, 3] and (b) to be consistent with the conservation law and an entropy inequality [4, 6]. The first property guarantees that the scheme does not generate spurious oscillations. We refer to schemes with this property as total variation diminishing (TVD) schemes (or total variation nonincreasing, TVNI, [2]). The latter property guarantees that the weak solutions are physical ones. Schemes in this class are guaranteed to avoid difficulties (i)–(iii) mentioned above.

The class of TVD schemes contains monotone schemes, but is significantly larger as it includes second-order accurate schemes. Existence of second-order accurate TVD schemes was demonstrated in [2, 3, 7, 8]. Unlike monotone schemes, TVD schemes are not automatically consistent with the entropy inequality. Consequently, some mechanism may have to be explicitly added to a TVD scheme to enforce the selection of the physical solution. In [2, 9], Harten and Harten and Hyman demonstrate a way of modifying a TVD scheme to be consistent with an entropy inequality.

In [10, 11], we have examined the application of an explicit second-order accurate TVD scheme [2] to steady-state calculations. Numerical experiments show that this explicit scheme generates nonoscillatory, highly accurate steady-state solutions.

To retain the characteristic of highly resolved steady-state solutions by explicit second-order accurate TVD schemes without the disadvantage of slow convergence rate of explicit schemes, we considered in [10] the following two possibilities: (1) First, obtain an approximation to the steady state by using a conventional implicit scheme, and then use a second-order accurate TVD scheme as a "post-processor." (2) Use a first-order accurate implicit scheme in delta-formulation and replace the explicit operator by an explicit second-order accurate TVD scheme.

We have found (in one dimension) that both these strategies reduce the overall computational effort needed to obtain the steady-state solution of the explicit second-order accurate TVD scheme. Alternative (1) is a possible way of speeding up the convergence process by providing a better initial condition for the explicit

second-order accurate TVD scheme. Alternative (2) can be viewed as a relaxation procedure to the steady-state solution. Numerical experiments of [10] show that the computational effort is not drastically decreased, although the stability limit is higher than the explicit counterpart.

Recently, Harten [3] has extended the class of explicit TVD schemes to a more general category which includes a one-parameter family of implicit second-order accurate schemes. Included in this class are the commonly used time-differencing schemes such as the backward Euler and the trapezoidal formula.

This paper is a sequel to [10]. Here, we investigate the application to steady-state calculations of this newly developed implicit second-order accurate scheme that is unconditionally TVD. This scheme is guaranteed not to generate spurious oscillations for one-dimensional nonlinear scalar equations and constant coefficient systems. Numerical experiments show that this scheme has a fairly rapid convergence rate, in addition to generating a highly resolved approximation to the steady-state solution. We remark that all of the analysis on the new scheme is for the initial value problem. The numerical boundary conditions are not included.

In the present paper, we stress applications rather than theory, and we refer the interested reader to [2, 3] for more theoretical details. In the next section, we will briefly review the notion of TVD schemes and describe the construction of the second-order accurate TVD scheme from a first-order accurate one for scalar one-dimensional hyperbolic conservation laws. The generalization to one-dimensional hyperbolic systems will be described in Section 3. A description of the algorithm and numerical results for the one- and two-dimensional compressible inviscid equations of gas dynamics will be presented in Sections 4 and 5.

## 2. TVD Schemes for One-Dimensional Scalar Hyperbolic Conservation Laws

Several techniques for the construction of nonlinear, explicit, second-order accurate, high-resolution, entropy satisfying schemes for hyperbolic conservation laws have been developed in recent years. See, for example, van Leer [7], Colella and Woodhard [12], and Harten [2]. From the standpoint of numerical analysis, these schemes are TVD for nonlinear scalar hyperbolic conservation laws and for constant coefficient hyperbolic systems. TVD schemes are usually rather complicated to use compared to the conventional shock-capturing methods such as variants of the Lax–Wendroff scheme.

In [3], Harten introduced the notion of implicit TVD schemes. To keep this paper somewhat self-contained, we will review the construction of the backward Euler TVD schemes for the initial value problem. This is the only unconditionally stable TVD scheme belonging to the one-parameter family of TVD schemes considered in [3]. Before we proceed with the description of the construction, we will first give preliminaries on the definition of explicit and implicit TVD schemes and show a few examples.

## 2.1. *Explicit TVD Schemes*

Consider the scalar hyperbolic conservation law

$$\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x}=0, \tag{2.1}$$

where $a(u)=\partial f/\partial u$ is the characteristic speed. A general three-point explicit difference scheme in conservation form can be written

$$u_j^{n+1}=u_j^n-\lambda(\bar{f}_{j+1/2}^n-\bar{f}_{j-1/2}^n), \tag{2.2}$$

where $\bar{f}_{j+1/2}^n=\bar{f}(u_j^n, u_{j+1}^n)$, $\lambda=\Delta t/\Delta x$, with $\Delta t$ the time step, and $\Delta x$ the mesh size. Here, $u_j^n$ is a numerical solution of (2.1) at $x=j\Delta x$ and $t=n\Delta t$ and $\bar{f}$ is a numerical flux function. We require the numerical flux function $\bar{f}$ to be consistent with the conservation law in the following sense:

$$\bar{f}(u_j, u_j)=f(u_j). \tag{2.3}$$

Consider a numerical scheme with numerical flux functions of the form

$$\bar{f}_{j+1/2}=\tfrac{1}{2}[f_j+f_{j+1}-Q(a_{j+1/2})\,\Delta_{j+1/2}u], \tag{2.4}$$

where $f_j=f(u_j)$, $\Delta_{j+1/2}u=u_{j+1}-u_j$, and

$$a_{j+1/2}=\begin{cases}(f_{j+1}-f_j)/\Delta_{j+1/2}u, & \Delta_{j+1/2}u\neq 0,\\ a(u_j), & \Delta_{j+1/2}u=0.\end{cases} \tag{2.5}$$

Here $Q$ is a function of $a_{j+1/2}$ and $\lambda$. The function $Q$ is sometimes referred to as the coefficient of numerical viscosity. Figure 2.1 shows some examples for the possible choice of $Q$. Three familiar schemes with the numerical fluxes of the form (2.4) are

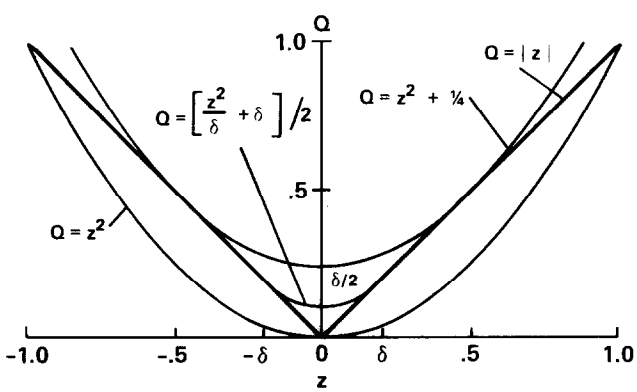


FIG. 2.1. Sample of the $Q(z)$ functions.

(a) A form of the Lax–Wendroff (L–W) scheme with

$$\bar{f}_{j+1/2} = \tfrac{1}{2}[f_j + f_{j+1} - \lambda(a_{j+1/2})^2 \Delta_{j+1/2} u], \tag{2.6}$$

where $Q(a_{j+1/2}) = \lambda(a_{j+1/2})^2$.

(b) Lax–Friedrichs (L–F) scheme with

$$\bar{f}_{j+1/2} = \frac{1}{2}\left[ f_j + f_{j+1} - \frac{1}{\lambda} \Delta_{j+1/2} u \right], \tag{2.7}$$

where $Q(a_{j+1/2}) = 1/\lambda$.

(c) A generalization of the Courant–Isaacson–Rees (GCIR) scheme with

$$\bar{f}_{j+1/2} = \tfrac{1}{2}[f_j + f_{j+1} - |a_{j+1/2}| \Delta_{j+1/2} u], \tag{2.8}$$

where $Q(a_{j+1/2}) = |a_{j+1/2}|$.

We define the total variation of a mesh function $u$ to be

$$\mathrm{TV}(u) = \sum_{j=-\infty}^{\infty} |u_{j+1} - u_j| = \sum_{j=-\infty}^{\infty} |\Delta_{j+1/2} u|. \tag{2.9}$$

We say that the numerical scheme (2.2) is TVD if

$$\mathrm{TV}(u^{n+1}) \leqslant \mathrm{TV}(u^n). \tag{2.10}$$

It can be shown that a sufficient condition for (2.2), together with (2.4), to be a TVD scheme is [2],

$$\lambda C^-_{j+1/2} = \frac{\lambda}{2}[-a_{j+1/2} + Q(a_{j+1/2})] \geqslant 0, \tag{2.11a}$$

$$\lambda C^+_{j+1/2} = \frac{\lambda}{2}[a_{j+1/2} + Q(a_{j+1/2})] \geqslant 0, \tag{2.11b}$$

$$\lambda(C^-_{j+1/2} + C^+_{j+1/2}) = \lambda Q(a_{j+1/2}) \leqslant 1. \tag{2.11c}$$

Applying condition (2.11) and/or (2.10) to the above three examples, it can be easily shown that the L–W scheme is not a TVD scheme, and the latter two schemes are TVD schemes. Note that there is a further distinction between the L–F scheme and GCIR scheme: the L–F scheme is consistent with an entropy inequality whereas the GCIR is not [6].

It should be emphasized that condition (2.11) is only a sufficient condition; i.e., schemes that fail this test might be still TVD. The L–W scheme, besides failing condition (2.11), does not satisfy (2.10).

2.2. *Implicit TVD Schemes*

Now we consider a one-parameter family of three-point conservative schemes of the form

$$u_j^{n+1} + \lambda\eta(\bar{f}_{j+1/2}^{n+1} - \bar{f}_{j-1/2}^{n+1}) = u_j^n - \lambda(1-\eta)(\bar{f}_{j+1/2}^n - \bar{f}_{j-1/2}^n), \tag{2.12}$$

where $\eta$ is a parameter, $\lambda = \Delta t/\Delta x$, $\bar{f}_{j+1/2}^n = \bar{f}(u_j^n, u_{j+1}^n)$, $\bar{f}_{j+1/2}^{n+1} = \bar{f}(u_j^{n+1}, u_{j+1}^{n+1})$, and $\bar{f}(u_j, u_{j+1})$ is the numerical flux (2.4). This one-parameter family of schemes contains implicit as well as explicit schemes. When $\eta = 0$, (2.12) reduces to (2.2), the explicit method. When $\eta \neq 0$, (2.12) is an implicit scheme. For example: if $\eta = \frac{1}{2}$, the time differencing is the trapezoidal formula, and if $\eta = 1$, the time differencing is the backward Euler method. To simplify the notation, we will rewrite (2.12) as

$$L \cdot u^{n+1} = R \cdot u^n, \tag{2.13}$$

where $L$ and $R$ are the finite-difference operators

$$(L \cdot u)_j = u_j + \lambda\eta(\bar{f}_{j+1/2} - \bar{f}_{j-1/2}), \tag{2.14a}$$

$$(R \cdot u)_j = u_j - \lambda(1-\eta)(\bar{f}_{j+1/2} - \bar{f}_{j-1/2}). \tag{2.14b}$$

A sufficient condition for (2.12) to be a TVD scheme is that

$$\mathrm{TV}(R \cdot v) \leqslant \mathrm{TV}(v), \tag{2.15a}$$

$$\mathrm{TV}(L \cdot v) \geqslant \mathrm{TV}(v). \tag{2.15b}$$

A sufficient condition for (2.15) is the CFL-like restriction

$$|\lambda a_{j+1/2}| \leqslant \lambda Q(a_{j+1/2}) \leqslant \frac{1}{1-\eta}, \tag{2.16}$$

where $a_{j+1/2}$ is defined in (2.5). For a detailed proof of (2.15) and (2.16), see [3]. Observe that the backward Euler implicit scheme, $\eta = 1$ in (2.12), is unconditionally TVD, while the trapezoidal formula, $\eta = \frac{1}{2}$, is TVD under the CFL-like restriction of 2. The forward Euler explicit scheme, $\eta = 0$ or (2.2), is TVD under the CFL restriction of 1. We remark that three-point conservative TVD schemes of the form (2.12) are generally first-order accurate in space. When $\eta = \frac{1}{2}$, the scheme is second-order accurate in time.

2.3. *First-Order Accurate Backward Euler Implicit TVD Scheme*

In this paper, we are only interested in efficient high-resolution time-dependent methods for steady-state calculations. The backward Euler implicit TVD scheme is the best choice in this one-parameter family of TVD schemes. Therefore, we will only review the proof that the backward Euler scheme is unconditionally TVD. In Section 2.4, we will describe the technique of converting the first-order accurate unconditionally TVD scheme (2.12) with $\eta = 1$ into a second-order accurate one.

The backward Euler three-point scheme in conservative form can be written as

$$u_j^{n+1} + \lambda(\bar{f}_{j+1/2}^{n+1} - \bar{f}_{j-1/2}^{n+1}) = u_j^n. \tag{2.17a}$$

For the purpose of this paper, the function $Q(z)$ in (2.4) is chosen to be

$$Q(z) = \begin{cases} \frac{1}{2}\left(\frac{z^2}{\delta} + \delta\right), & |z| < \delta, \\ |z|, & |z| \geqslant \delta, \end{cases} \tag{2.17b}$$

which is a nonvanishing, continuously differentiable approximation to $|z|$. This is one way of modifying $Q$ in (2.8) so that scheme (2.2) together with (2.8) is an entropy satisfying TVD scheme [2]. For convenience, we introduce the notation

$$C^{\pm}(z) = \tfrac{1}{2}[Q(z) \pm z] \tag{2.18a}$$

and note that

$$C^{\pm}(z) \geqslant 0 \qquad \text{for all} \quad z. \tag{2.18b}$$

Using (2.5) and (2.18a), we can rewrite the numerical fluxes $\bar{f}_{j+1/2}$ in (2.4) as

$$\bar{f}_{j+1/2} = f_j - C^-(a_{j+1/2})\, \Delta_{j+1/2} u, \tag{2.19a}$$

$$\bar{f}_{j-1/2} = f_j - C^+(a_{j-1/2})\, \Delta_{j-1/2} u. \tag{2.19b}$$

It follows from (2.19) that (2.17) can be written in the form

$$u_j^{n+1} - \lambda C^-(a_{j+1/2}^{n+1})\, \Delta_{j+1/2} u^{n+1} + \lambda C^+(a_{j-1/2}^{n+1})\, \Delta_{j-1/2} u^{n+1} = u_j^n. \tag{2.20}$$

Now, if $\delta = 0$ in (2.17b), then $C^{\pm}(z) = (|z| \pm z)/2$, and (2.20) is a first-order accurate, upstream differencing, backward Euler implicit scheme. Equation (2.17) differs from the upstream spatial differencing (with $\delta = 0$) by the addition of a numerical viscosity term with a coefficient $\delta > 0$.

We show now that $C^+(z) \geqslant 0$ implies that the scheme (2.17) is unconditionally TVD (i.e., condition (2.10) is satisfied, independent of the value of $\lambda = \Delta t/\Delta x$ in (2.17a).

To see that, we subtract (2.20) at $j$ from (2.20) at $j+1$ and get after rearranging terms that

$$\begin{aligned} &[1 + \lambda C_{j+1/2}^- + \lambda C_{j+1/2}^+]\, \Delta_{j+1/2} u^{n+1} \\ &\quad = \Delta_{j+1/2} u^n + \lambda C_{j-1/2}^+ \Delta_{j-1/2} u^{n+1} + \lambda C_{j+3/2}^- \Delta_{j+3/2} u^{n+1}. \end{aligned} \tag{2.21a}$$

Here $C_{j+1/2}^{\pm} = C^{\pm}(a_{j+1/2}^{n+1})$. Next we take the absolute value of (2.21a) and use (2.18b) and the triangle inequality to obtain

$$\begin{aligned} &[1 + \lambda C_{j+1/2}^- + \lambda C_{j+1/2}^-]\, |\Delta_{j+1/2} u^{n+1}| \\ &\quad \leqslant |\Delta_{j+1/2} u^n| + \lambda C_{j-1/2}^+ |\Delta_{j-1/2} u^{n+1}| + \lambda C_{j+3/2}^- |\Delta_{j+3/2} u^{n+1}|. \end{aligned} \tag{2.21b}$$

Rearranging terms, we get

$$|\Delta_{j+1/2}u^{n+1}| \leqslant |\Delta_{j+1/2}u^n| + \lambda[C^-_{j+3/2}|\Delta_{j+3/2}u^{n+1}| - C^-_{j+1/2}|\Delta_{j+1/2}u^{n+1}|] - \lambda[C^+_{j+1/2}|\Delta_{j+1/2}u^{n+1}| - C^+_{j-1/2}|\Delta_{j-1/2}u^{n+1}|]. \tag{2.21c}$$

That is,

$$|\Delta_{j+1/2}u^{n+1}| \leqslant |\Delta_{j+1/2}u^n| + \lambda(\Xi_{j+1} - \Xi_j), \tag{2.21d}$$

where

$$\Xi_j = C^-_{j+1/2}|\Delta_{j+1/2}u^{n+1}| - C^+_{j-1/2}|\Delta_{j-1/2}u^{n+1}|. \tag{2.21e}$$

Summing (2.21d) from $j = -\infty$ to $j = +\infty$, we obtain (2.10); thus proving that our backward Euler implicit scheme is unconditionally TVD.

2.4. *Conversion to Second-Order Accurate Scheme*

Next, we want to briefly review the design principle behind the construction of second-order accurate TVD schemes. This is a rather general technique to convert a three-point first-order accurate (in space) TVD scheme (2.12) into a five-point second-order accurate (in both time and space, or just space) TVD scheme of the same generic form. The design of high-resolution TVD schemes rests on the fact that the exact solution to (2.1) is TVD due to the phenomenon of propagation along characteristics, and is independent of the particular form of the flux $f(u)$ in (2.1). Similarly, the first-order accurate scheme is TVD subject only to the CFL-like restriction (2.16), independent of the particular form of the flux. Thus to achieve second-order accuracy while retaining the TVD property, we use the original TVD scheme with an appropriately modified flux $(f + g)$, i.e.,

$$u_j^{n+1} + \lambda(\tilde{f}^{n+1}_{j+1/2} - \tilde{f}^{n+1}_{j-1/2}) = u_j^n, \tag{2.22a}$$

$$\tilde{f}_{j+1/2} = \tfrac{1}{2}[f_j + f_{j+1} + g_j + g_{j+1} - Q(a_{j+1/2} + \gamma_{j+1/2})\,\Delta_{j+1/2}u], \tag{2.22b}$$

where

$$\gamma_{j+1/2} = \begin{cases} (g_{j+1} - g_j)/\Delta_{j+1/2}u, & \Delta_{j+1/2}u \neq 0, \\ 0, & \Delta_{j+1/2}u = 0. \end{cases} \tag{2.22c}$$

The requirements on $g$ are: (1) The function $g$ should have a bounded $\gamma$ in (2.22c) so that (2.22a) is TVD with respect to the modified flux $(f + g)$. (2) The modified scheme should be second-order accurate (except at points of extrema). In [2, 3], Harten devised a recipe for $g$ that satisfies the above two requirements. We will use this particular form of $g$ for the discussion here. It can be written

$$g_j = S \cdot \max[0, \min(\sigma_{j+1/2}|\Delta_{j+1/2}u|, S \cdot \sigma_{j-1/2}\Delta_{j-1/2}u)], \qquad S = \operatorname{sign}(\Delta_{j+1/2}u), \tag{2.22d}$$

with $\sigma_{j+1/2} = \sigma(a_{j+1/2})$ and we choose

$$\sigma(z) = \tfrac{1}{2}Q(z) \geqslant 0 \tag{2.22e}$$

for steady-state applications. It has the property that the steady-state solution is independent of $\Delta t$. Or, we choose

$$\sigma(z) = \tfrac{1}{2}(Q(z) + \lambda z^2) \geqslant 0 \tag{2.22f}$$

for time-accurate calculations. Note that if $\sigma(z) = (Q(z) + \lambda z^2)/2$, then (2.22) is second-order accurate in both time and space [3]. For transient calculations, second-order accurate in time is preferred.

The form of $g$ in (2.22d) satisfies the relations [3]

$$g_j = g(u_{j-1}, u_j, u_{j+1}), \qquad g(u, u, u) = 0, \tag{2.23a}$$

$$|\gamma_{j+1/2}| = |g_{j+1} - g_j|/|u_{j+1} - u_j| \leqslant \sigma(a_{j+1/2}), \tag{2.23b}$$

$$g = \Delta x \sigma(a) \frac{\partial u}{\partial x} + O((\Delta x)^2). \tag{2.23c}$$

Relation (2.23a) shows that the modified numerical flux (2.22b) is consistent with $f(u)$. Relation (2.23b) shows that the mean-value characteristic speed $\gamma_{j+1/2}$ (2.22c) induced by the flux $g$ is uniformly bounded. Relation (2.23c) implies that (2.22b) is second-order accurate in space. The form of $g$ appears more complicated than it really is. The various test functions in (2.22d) can be viewed as an automatic way of controlling the numerical flux function so that (2.22) is TVD.

The scheme (2.22) can be rewritten in the form (2.20) as

$$u_j^{n+1} - \lambda C^-(a+\gamma)_{j+1/2}^{n+1} \Delta_{j+1/2} u^{n+1} + \lambda C^+(a+\gamma)_{j-1/2}^{n+1} \Delta_{j-1/2} u^{n+1} = u_j^n, \tag{2.24}$$

where $C^{\pm}(a+\gamma)_{j\pm 1/2}^{n+1} \equiv C^{\pm}(a_{j\pm 1/2}^{n+1} + \gamma_{j\pm 1/2}^{n+1})$; i.e., $C^{\pm}$ is now a function of $(a+\gamma)$ instead of $a$. The modified scheme (2.22) is of the same generic form as the original first-order scheme (2.17). Therefore (2.22) is an upstream differencing scheme with respect to the characteristic field $(a+\gamma)$. Moreover, we have the relation

$$\operatorname{sign}(a+\gamma) = \operatorname{sign}(a) \tag{2.25}$$

for $|z| \geqslant \delta$, with $z = a$ or $(a+\gamma)$ in (2.17b). Hence (2.24) is also an upstream differencing scheme with respect to the original characteristic field $a(u)$.

Because of (2.23a), the numerical flux (2.22b) of the second-order accurate TVD scheme depends on four points, i.e., $\tilde{f}_{j+1/2} = \tilde{f}(u_{j-1}, u_j, u_{j+1}, u_{j+2})$, and thus (2.22) is formally a five-point scheme. We note, however, that

$$\tilde{f}(v, u, u, w) = f(u) \tag{2.26}$$

for all $v$ and $w$. Hence, for practical purposes, such as numerical boundary conditions, (2.22) can be regarded as essentially a three-point scheme.

We turn now to examine the behavior of TVD schemes around points of extrema, by considering their application to data, where

$$u_{j-1} \leqslant u_j = u_{j+1} \geqslant u_{j+2}. \tag{2.27}$$

In this case $g_j = g_{j+1} = 0$ in (2.22d), and thus the numerical flux (2.22b) becomes identical to that of the original first-order accurate scheme (2.4); consequently, the truncation error of (2.22) deteriorates to $O((\Delta x)^2)$ at $j$ and $j+1$. This behavior is common to all TVD schemes. Thus, for a second-order accurate scheme to be TVD, it has to have a mechanism that switches itself into a first-order accurate TVD scheme at points of extrema. Because of the above property, second-order accurate TVD schemes are genuinely nonlinear (i.e., they are nonlinear even in the constant coefficient case).

Extension of the one-parameter family of three-point TVD schemes (2.12) to second-order TVD schemes follows the same procedure except (2.22f) becomes

$$\sigma(z) = \tfrac{1}{2}Q(z) + \lambda(\eta - \tfrac{1}{2})\, z^2. \tag{2.28}$$

2.5. *Enhancement of Resolution by Artificial Compression*

The technique to convert the first-order accurate TVD scheme (2.12) into a second-order accurate one is closely related to the concept of artificial compression [13, 14].

Truncation error analysis shows that the first-order accurate scheme (2.12) is a second-order accurate approximation to solutions of the modified equation

$$\frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} = \Delta x \frac{\partial}{\partial x}\left(\sigma(a)\frac{\partial u}{\partial x}\right), \tag{2.29}$$

where $\sigma(a)$ is defined in (2.22e) or (2.28). We note that the CFL-like restriction (2.16) implies that $\sigma(a) \geqslant 0$; thus, the right-hand side of (2.29) is a viscosity term. Hence the first-order accurate TVD scheme (2.12) is a better approximation to the viscous equation (2.29) than it is to the original conservation law.

We obtain a second-order approximation to $\partial u/\partial t + \partial f/\partial x = 0$ by applying the first-order scheme (2.12) to the modified flux $(f+g)$, where $g$ is an approximation to the right-hand side of (2.29); i.e.,

$$g = \Delta x \sigma(a)\frac{\partial u}{\partial x} + O((\Delta x)^2). \tag{2.30}$$

The application of the first-order scheme to $(f+g)$ has the effect of canceling the error due to the numerical viscosity to $O((\Delta x)^2)$; thus $g$ is an "anti-diffusion" flux.

If we apply the first-order TVD scheme to $(f + (1+\bar{\omega})\, g)$, $\bar{\omega} > 0$, rather than to $(f+g)$, we find that the resolution of discontinuities improves with increasing $\bar{\omega}$. This observation allows us to use the notion of artificial compression to enhance the resolution of discontinuities computed by the second-order accurate TVD

scheme (2.22). This is done by increasing the size of $g$ in (2.22d) by adding a term that is $O((\Delta x)^2)$ in regions of smoothness, e.g.,

$$\tilde{g}_j = (1 + \omega\theta_j)\, g_j, \qquad \omega > 0 \tag{2.31a}$$

with

$$\theta_j = \frac{|\Delta_{j+1/2} u - \Delta_{j-1/2} u|}{|\Delta_{j+1/2} u| + |\Delta_{j-1/2} u|}. \tag{2.31b}$$

Using $\tilde{g}_j$ (2.31) instead of $g_j$ makes the numerical characteristic speed more convergent, and therefore improves the resolution of computed shocks. Since $\theta = O(\Delta x)$, this change does not adversely affect the order of accuracy of the scheme. See [2] for more details. From numerical experiments, $\omega = 2$ seems to be a good choice.

We remark that applying too much artificial compression in a region of expansion (i.e., divergence of the characteristic field $a = \partial f/\partial u$) may result in violation of the entropy condition. Hence when applying artificial compression, we have to either turn it off in regions of expansion or limit the size of $\omega$ in (2.31a), say, by the value that makes (2.22) with (2.31a) third-order accurate (in regions of monotonicity).

## 2.6. *Linearized Version of the Implicit TVD Scheme*

To solve for $u^{n+1}$ for the first- or second-order implicit scheme, we have to solve a set of nonlinear algebraic equations. To overcome this obstacle, we will present a way of linearizing the implicit TVD scheme. The method will destroy the conservative property but preserve its unconditionally TVD property. We will refer to this method as the linearized nonconservative implicit (LNI) form. The LNI form is mainly useful for steady-state calculations, since the scheme is only conservative after the solution reaches steady state. On the other hand, we have the advantage of stability and TVD of an unlimited CFL number. Note that the procedure of obtaining the LNI form is applicable to both the first- and second-order accurate implicit TVD schemes. We will discuss the LNI for the second-order accurate one. To get the LNI for the first-order accurate TVD scheme, we simply set $g = \gamma = 0$ in the second-order form.

The LNI form is obtained simply by replacing the coefficients $(C^{\pm})^{n+1}$ in (2.24) by $(C^{\pm})^n$, i.e.,

$$u_j^{n+1} - \lambda C^-(a+\gamma)^n_{j+1/2}\, \Delta_{j+1/2} u^{n+1} + \lambda C^+(a+\gamma)^n_{j-1/2}\, \Delta_{j-1/2} u^{n+1} = u_j^n. \tag{2.32}$$

Since $C^{\pm} \geqslant 0$, it follows from (2.21) that (2.32) is unconditionally TVD.

In delta form notation, (2.32) can be rewritten as

$$\begin{aligned}[1 - \lambda C^-(a+\gamma)^n_{j+1/2}\, \Delta_{j+1/2} + \lambda C^+(a+\gamma)^n_{j-1/2}\, \Delta_{j-1/2}](u^{n+1} - u^n) \\ = -\lambda[\tilde{f}^n_{j+1/2} - \tilde{f}^n_{j-1/2}],\end{aligned} \tag{2.33a}$$

where the left-hand side equals

$$d_j - \lambda C^-(a+\gamma)^n_{j+1/2}\,\Delta_{j+1/2}d + \lambda C^+(a+\gamma)^n_{j-1/2}\,\Delta_{j-1/2}d, \tag{2.33b}$$

with $d_j = u_j^{n+1} - u_j^n$ and $\Delta_{j+1/2}d = d_{j+1} - d_j$. Rearranging terms, we get

$$E_1 d_{j-1} + E_2 d_j + E_3 d_{j+1} = -\lambda[\tilde{f}^n_{j+1/2} - \tilde{f}^n_{j-1/2}], \tag{2.34a}$$

with

$$E_1 = -\lambda C^+(a+\gamma)^n_{j-1/2}, \tag{2.34b}$$

$$E_2 = 1 + \lambda[C^-(a+\gamma)^n_{j+1/2} + C^+(a+\gamma)^n_{j-1/2}], \tag{2.34c}$$

$$E_3 = -\lambda C^-(a+\gamma)^n_{j+1/2}. \tag{2.34d}$$

Here, $\tilde{f}^n_{j+1/2}$ is (2.22b)–(2.22e) calculated at the time level $n$. It follows from (2.22b) and (2.33a) that the steady-state solution of (2.33) is

(i) consistent with the conservation form, and

(ii) a spatially second-order accurate approximation to the steady state of the partial differential equation

(iii) Independent of the time-step $\Delta t$ used in the iterations.

Moreover, the iteration matrix associated with (2.34) is a *diagonally dominant*, tridiagonal matrix. Note that this linearized construction is not trivial, since the second-order method is a five-point scheme. Normally the matrix associated with (2.34) could have been a pentadiagonal matrix. As mentioned before, (2.32) or (2.34) is not in conservation form and therefore should not be used to approximate time-dependent solutions (transient solutions). However, it is a suitable scheme for the calculation of steady-state solutions.

We can also obtain another TVD linearized form by setting $\gamma = 0$ in (2.34), i.e.,

$$\bar{E}_1 d_{j-1} + \bar{E}_2 d_j + \bar{E}_3 d_{j+1} = -\lambda[\tilde{f}^n_{j+1/2} - \tilde{f}^n_{j-1/2}], \tag{2.35a}$$

with

$$\bar{E}_1 = -\lambda C^+(a^n_{j-1/2}), \tag{2.35b}$$

$$\bar{E}_2 = 1 + \lambda[C^-(a^n_{j+1/2}) + C^+(a^n_{j-1/2})], \tag{2.35c}$$

$$\bar{E}_3 = -\lambda C^-(a^n_{j+1/2}). \tag{2.35d}$$

Scheme (2.35) is spatially first-order accurate for the implicit operator and spatially second-order accurate for the explicit operator. It can be shown that (2.35) is still TVD.

## 3. Generalization to One-Dimensional Hyperbolic System of Conservation Laws

In the present state of development, the concept of TVD schemes, like monotone schemes, is only defined for nonlinear scalar conservation laws or constant coefficient hyperbolic systems. The main difficulty stems from the fact that, unlike the scalar case, the total variation in $x$ of the solution to the system of nonlinear conservation laws is not necessarily a monotonic decreasing function of time. The total variation of the solution may actually increase at moments of interaction between waves. Not knowing a diminishing functional that bounds the total variation in $x$ in the system case, makes it impossible to fully extend the theory of the scalar case to the system case. What we can do at the moment is to extend the new scalar TVD scheme to system cases so that the resulting scheme is TVD for the "locally frozen" constant coefficient system. To accomplish this, we define at each point a "local" system of characteristic fields. This extension technique is a somewhat generalized version of the procedure suggested by Roe [15].

Now, we briefly describe the above approach of extending the second-order accurate TVD schemes to hyperbolic systems of conservation laws

$$\frac{\partial U}{\partial t}+\frac{\partial F(U)}{\partial x}=0, \qquad A(U)=\frac{\partial F}{\partial U}, \tag{3.1}$$

Here $U$ and $F(U)$ are column vectors of $m$ components and $A(U)$ is the Jacobian matrix. The assumption that (3.1) is hyperbolic implies that $A(U)$ has real eigenvalues $a^l(U)$ and a complete set of right eigenvectors $R^l(U)$, $l=1,\ldots,m$. Hence the matrix

$$R(U)=(R^1(U),\ldots,R^m(U)) \tag{3.2a}$$

is invertible. The rows $L^1(U),\ldots,L^m(U)$ of $R(U)^{-1}$ constitute an orthonormal set of left eigenvectors of $A(U)$; thus

$$R^{-1}AR=\operatorname{diag}(a^l). \tag{3.2b}$$

Here $\operatorname{diag}(a^l)$ denotes a diagonal matrix with diagonal elements $a^l$.

We define characteristic variables $W$ with respect to the state $U$ by

$$W=R^{-1}U. \tag{3.3}$$

In the constant coefficient case, (3.1) decouples into $m$ scalar equations for the characteristic variables

$$\frac{\partial w^l}{\partial t}+a^l\frac{\partial w^l}{\partial x}=0, \qquad a^l=\text{constant}. \tag{3.4}$$

This offers a natural way of extending a scalar scheme to a constant coefficient

system by applying it "scalarly" to each of the $m$ scalar characteristic equations (3.4).

Let $U_{j+1/2}$ denote some symmetric average of $U_j$ and $U_{j+1}$ (to be discussed later); i.e.,

$$U_{j+1/2} = \Psi(U_j, U_{j+1}). \tag{3.5}$$

Let $a^l_{j+1/2}, R^l_{j+1/2}, L^l_{j+1/2}$ denote the respective quantities of $a^l, R^l, L^l$ related to $A(U_{j+1/2})$. Let $w^l$ be the vector elements of $W$, and let $\alpha^l_{j+1/2} = w^l_{j+1} - w^l_j$ be the component of $\Delta_{j+1/2}U = U_{j+1} - U_j$ in the $l$th characteristic direction; i.e.,

$$\Delta_{j+1/2}U = R_{j+1/2}\alpha_{j+1/2}, \qquad \alpha_{j+1/2} = R^{-1}_{j+1/2}\Delta_{j+1/2}U. \tag{3.6}$$

With the above notation, we can apply scheme (2.22) scalarly to each of the locally defined (*frozen coefficient*) characteristic variables of (3.1) as

$$U^{n+1}_j + \lambda(\tilde{F}^{n+1}_{j+1/2} - \tilde{F}^{n+1}_{j-1/2}) = U^n_j, \tag{3.7a}$$

$$\tilde{F}_{j+1/2} = \frac{1}{2}(F_j + F_{j+1}) + \frac{1}{2}\sum_{l=1}^{m}[g^l_j + g^l_{j+1} - Q(a^l_{j+1/2} + \gamma^l_{j+1/2})\,\alpha^l_{j+1/2}]\,R^l_{j+1/2}, \tag{3.7b}$$

where

$$\begin{aligned} &g^l_j = S\cdot\max[0, \min(\sigma^l_{j+1/2}|\alpha^l_{j+1/2}|, S\cdot\sigma^l_{j-1/2}\alpha^l_{j-1/2})],\\ &S = \operatorname{sign}(\alpha^l_{j+1/2}), \end{aligned} \tag{3.7c}$$

and

$$\gamma^l_{j+1/2} = \begin{cases} (g^l_{j+1} - g^l_j)/\alpha^l_{j+1/2}, & \alpha^l_{j+1/2} \neq 0,\\ 0, & \alpha^l_{j+1/2} = 0. \end{cases} \tag{3.7d}$$

Here $\sigma^l_{j+1/2} = \sigma(a^l_{j+1/2})$, where $\sigma(z)$ is (2.22e) and $\alpha^l_{j+1/2}$ is (3.6). The corresponding $\tilde{g}_j$ in (2.31) for the added artificial compression term is

$$\tilde{g}^l_j = (1 + \omega^l\theta^l_j)\,g^l_j, \qquad \omega^l > 0, \tag{3.7e}$$

with

$$\theta^l_j = \frac{|\alpha^l_{j+1/2} - \alpha^l_{j-1/2}|}{|\alpha^l_{j+1/2}| + |\alpha^l_{j-1/2}|}. \tag{3.7f}$$

The $\omega^l$ can be different from one characteristic field to another.

Similarly, we generalize the LNI form (2.33) to the system case by

$$[I - \lambda J^-_{j+1/2}\Delta_{j+1/2} + \lambda J^+_{j-1/2}\Delta_{j-1/2}](U^{n+1} - U^n) = -\lambda[\tilde{F}^n_{j+1/2} - \tilde{F}^n_{j-1/2}] \tag{3.8a}$$

or

$$\tilde{E}_1 D_{j-1} + \tilde{E}_2 D_j + \tilde{E}_3 D_{j+1} = -\lambda[\tilde{F}^n_{j+1/2} - \tilde{F}^n_{j-1/2}], \tag{3.8b}$$

with

$$\tilde{E}_1 = -\lambda J^+_{j-1/2}, \tag{3.8c}$$

$$\tilde{E}_2 = I + \lambda[J^-_{j+1/2} + J^+_{j-1/2}], \tag{3.8d}$$

$$\tilde{E}_3 = -\lambda J^-_{j+1/2} \tag{3.8e}$$

and

$$J^{\pm}_{j+1/2} = R^n_{j+1/2}\,\mathrm{diag}(C^{\pm}(a^l + \gamma^l)^n_{j+1/2})(R^{-1})^n_{j+1/2}, \tag{3.8f}$$

$$D_j = U^{n+1}_j - U^n_j, \tag{3.8g}$$

where the left-hand side of (3.8a) is equal to

$$D_j - \lambda J^-_{j+1/2}\Delta_{j+1/2}D + \lambda J^+_{j-1/2}\Delta_{j-1/2}D, \tag{3.8h}$$

with $\Delta_{j+1/2}D = D_{j+1} - D_j$.

In the constant coefficient case where $A(U) = \text{constant}$, both (3.7) and (3.8) are TVD by construction. However, they are not identical; Eq. (3.7) is fully nonlinear while (3.8) is a version with a linearized left-hand side.

Note that the total variation for the vector mesh function $U$ of the constant coefficient case is defined as

$$TV(U) = \sum_{j=-\infty}^{m} \sum_{l=1}^{m} |\alpha^l_{j+1/2}|. \tag{3.9}$$

A particular form of averaging in (3.5) is essential if we require the scheme (3.7) for $m = 1$ to be identical to the scalar scheme of Section 2, since we have to choose (3.5) so that $a^l_{j+1/2}$ is the same as the mean value in Eq. (2.5). This can be accomplished by taking the eigenvalues $a^l_{j+1/2}$ and the eigenvectors $R^l_{j+1/2}$ in (3.2) to be those of $A(U_j, U_{j+1})$, where $A(U_j, U_{j+1})$ is the mean value Jacobian. This matrix should satisfy

(i) $F(U) - F(V) = A(U, V)(U - V)$,

(ii) $A(U, U) = A(U)$,

(iii) $A(U, V)$ has real eigenvalues and a complete set of eigenvectors.

Roe [15] constructs a mean value Jacobian for the Euler equations of gas dynamics of the form $A(U, V) = A(\Psi(U, V))$, where $\Psi(U, V)$ is some particular average. We will discuss Roe's mean value Jacobian in the next two sections.

## 4. Applications to One-Dimensional Compressible Inviscid Equations of Gas Dynamics

In this section we describe how to apply the implicit TVD scheme (3.8) to the compressible inviscid equations of gas dynamics (Euler equations). Included in this

section are: (1) a detailed description of each of the terms of Eqs. (3.5) and (3.8) for the Euler equations, (2) Roe's form for $A(U_j, U_{j+1})$, (3) an algorithm to compute $U_j^{n+1}$, (4) a description of the numerical example, and (5) a discussion of the numerical results.

4.1. *Description of Algorithm*

In one spatial dimension, the Euler equations of gas dynamics can be written in the conservative form as

$$\frac{\partial U}{\partial t}+\frac{\partial F(U)}{\partial x}=0, \tag{4.1a}$$

where

$$U=\begin{bmatrix}\rho\\ m\\ e\end{bmatrix}, \qquad F=\begin{bmatrix}m\\ m^2/\rho+p\\ (e+p)\,m/\rho\end{bmatrix}. \tag{4.1b}$$

Here $U$ is the vector of conservative variables, $F$ is the flux vector, and $m=\rho u$. The primitive variables are the density $\rho$, the velocity $u$, and the pressure $p$. The total energy per unit volume $e$, is defined as

$$e=\rho\varepsilon+\rho u^2/2 \tag{4.1c}$$

with $\varepsilon$ as the internal energy per unit mass. The pressure $p$ for a perfect gas is defined as

$$p=(\gamma-1)[e-m^2/2\rho], \tag{4.1d}$$

where $\gamma$ is the ratio of specific heats and should not be confused with the $\gamma_{j+1/2}$ in (2.22c) or $\gamma'_{j+1/2}$ in (3.7d).

Let $A$ denote the Jacobian matrix $\partial F(U)/\partial U$ whose eigenvalues are

$$(a^1, a^2, a^3)=(u-c, u, u+c), \tag{4.2}$$

where $c$ is the local speed of sound. The right eigenvectors of $A$ form the matrix $R=(R^1, R^2, R^3)$ given by

$$R=\begin{bmatrix}1 & 1 & 1\\ u-c & u & u+c\\ H-uc & \frac{1}{2}u^2 & H+uc\end{bmatrix}, \tag{4.3a}$$

$$H=\frac{c^2}{\gamma-1}+\frac{u^2}{2}, \tag{4.3b}$$

and

$$R^{-1} = \begin{bmatrix} \frac{1}{2}\left(b_1 + \frac{u}{c}\right) & \frac{1}{2}\left(-b_2 u - \frac{1}{c}\right) & \frac{1}{2}b_2 \\ 1 - b_1 & b_2 u & -b_2 \\ \frac{1}{2}\left(b_1 - \frac{u}{c}\right) & \frac{1}{2}\left(-b_2 u + \frac{1}{c}\right) & \frac{1}{2}b_2 \end{bmatrix}, \tag{4.3c}$$

with

$$b_1 = b_2 \frac{u^2}{2}, \tag{4.3d}$$

$$b_2 = \frac{\gamma - 1}{c^2}. \tag{4.3e}$$

Using the same notation as in Section 3, the vector $\alpha$ of Eq. (3.6) is

$$\begin{bmatrix} \alpha^1_{j+1/2} \\ \alpha^2_{j+1/2} \\ \alpha^3_{j+1/2} \end{bmatrix} = \begin{bmatrix} (aa - bb)/2 \\ \Delta_{j+1/2}\rho - aa \\ (aa + bb)/2 \end{bmatrix}, \tag{4.4a}$$

where

$$aa = \frac{\gamma - 1}{c^2_{j+1/2}}\left[\Delta_{j+1/2}e + \frac{u^2_{j+1/2}}{2}\Delta_{j+1/2}\rho - u_{j+1/2}\Delta_{j+1/2}m\right], \tag{4.4b}$$

$$bb = \frac{\Delta_{j+1/2}m - u_{j+1/2}\Delta_{j+1/2}\rho}{c_{j+1/2}}, \tag{4.4c}$$

The simplest form of $U_{j+1/2}$ is

$$U_{j+1/2} = (U_{j+1} + U_j)/2. \tag{4.5}$$

Roe [15] uses a special form of averaging that has the computational advantage of perfectly resolving stationary discontinuities. Roe's averaging takes the form

$$u_{j+1/2} = \frac{\bar{D}u_{j+1} + u_j}{\bar{D} + 1}, \tag{4.6a}$$

$$H_{j+1/2} = \frac{\bar{D}H_{j+1} + H_j}{\bar{D} + 1}, \tag{4.6b}$$

$$c^2_{j+1/2} = (\gamma - 1)\left(H_{j+1/2} - \frac{1}{2}u^2_{j+1/2}\right), \tag{4.6c}$$

$$\bar{D} = \sqrt{\rho_{j+1}/\rho_j}, \tag{4.6d}$$

$$H = \frac{\gamma p}{(\gamma - 1)\rho} + \frac{1}{2}u^2. \tag{4.6e}$$

To use Roe's averaging, we compute $u_{j+1/2}, c_{j+1/2}$ in (4.2)–(4.4) by (4.6). In the numerical experiments for the one-dimensional test problem, we use Roe's

[redacted]

ing a fixed CFL number as input):

(i) Compute $u_j = m_j/\rho_j$ and $p_j$.

(ii) Compute $u_{j+1/2}$ and $c_{j+1/2}$ from (4.5) or (4.6), calculate

$$M = \max_j (|u_{j+1/2}| + c_{j+1/2}),$$

evaluate $\alpha^l_{j+1/2}$, $l = 1, 2, 3$ by (4.4a), and define

$$\lambda = \Delta t/\Delta x = \mu/M,$$

where $\mu$ is the prescribed CFL number as input.

(iii) Compute $a^l_{j+1/2}$ by (4.2), $g^l_{j+1/2}$ by (3.7c).

(iv) Compute $\gamma^l_{j+1/2}$ by (3.7d), and $\tilde{F}_{j+1/2}$ by (3.7b) and relation (4.3a).

(v) Compute $C^{\pm}(a^l_{j+1/2} + \gamma^l_{j+1/2})$ by (2.18a), with $z = a^l_{j+1/2} + \gamma^l_{j+1/2}$.

(vi) Compute $J^{\pm}_{j+1/2}$ by (3.8f), and $\tilde{E}_1$, $\tilde{E}_2$, and $\tilde{E}_3$, by (3.8c)–(3.8e).

(vii) Solve the tridiagonal system (3.8b) for $D_j$ and then compute $U^{n+1}_j$ from (3.8g).

4.2. *Numerical Example*

For the numerical experiments, a quasi-one-dimensional nozzle problem was selected. The governing equations for the nozzle problem can be written

$$\frac{\partial \bar{U}}{\partial t} + \frac{\partial F(\bar{U})}{\partial x} + H(\bar{U}) = 0, \tag{4.7a}$$

where

$$\bar{U} = \begin{bmatrix} \rho\kappa \\ m\kappa \\ e\kappa \end{bmatrix}, \quad F = \begin{bmatrix} m\kappa \\ (m^2/\rho + p)\,\kappa \\ (e+p)\,m\kappa/\rho \end{bmatrix}, \quad H(\bar{U}) = \begin{bmatrix} 0 \\ p\dfrac{\partial \kappa}{\partial x} \\ 0 \end{bmatrix}, \tag{4.7b}$$

with $\kappa$, the area of the nozzle, a function of $x$. The nozzle we consider (see Fig. 4.1) is a divergent nozzle [16] with

$$\kappa(x) = 1.398 + 0.347 \tanh(0.8x - 4). \tag{4.7c}$$

The steady flow conditions were supersonic inflow, subsonic outflow with a shock. In all of the calculations the computational domain was $0 \leqslant x \leqslant 10$. We used a very

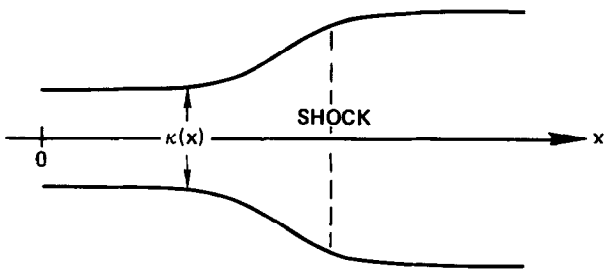


FIG. 4.1. Divergent nozzle.

coarse equal mesh spacing of $\Delta x = 0.5$ (i.e., 20 spatial intervals), to evaluate the resolution of the scheme.

*Initial conditions.* We use linear interpolation between the exact steady-state boundary values as initial conditions.

*Analytical boundary conditions.* We specified all three conservative variables $\rho$, $u$, and $e$ for the supersonic inflow, and the variable $e$ for the subsonic outflow.

*Numerical boundary conditions.* We used zeroth- or first-order space extrapolation to obtain the numerical boundary conditions for the unknown flow variables ($\rho$ and $m$) at the outflow boundary. Since the spatially second-order accurate TVD scheme is a five-point scheme, we also need the values of $g_j^l$ and $\theta_j^l$ on both boundaries. For convenience, we will use zeroth-order space extrapolation for these values.

### 4.3. *Discussion of Numerical Results*

All of the computations for the quasi-one-dimensional nozzle problem were done in single precision on the VAX 11/780 computer (a 6 digit machine). To illustrate the stability and/or accuracy of the LNI form of the implicit TVD scheme (3.8b), we compare in Fig. 4.2 the computed results with the explicit TVD scheme (forward Euler in time, obtained by setting $\tilde{E}_1 = 0$, $\tilde{E}_2 = I$, and $\tilde{E}_3 = 0$ in Eq. (3.8b)), a first-order flux-vector splitting scheme [17], and a conventional implicit method using backward Euler in time and central spatial differencing with an added fourth-order explicit numerical dissipation [18]. Note that not all 20 points of the solutions are plotted in Fig. 4.2. The points missing from both ends of the $x$ axis are equal to the exact solution. Figure 4.2a shows the numerical result of the explicit TVD scheme. It took approximately 700 steps to converge to the steady-state with a fixed CFL = 0.8. Figure 4.2b shows the computation of the LNI form of the second-order accurate implicit TVD scheme with CFL = $10^6$ after 25 steps. Both the explicit and implicit TVD schemes produce similar high resolution solutions (with

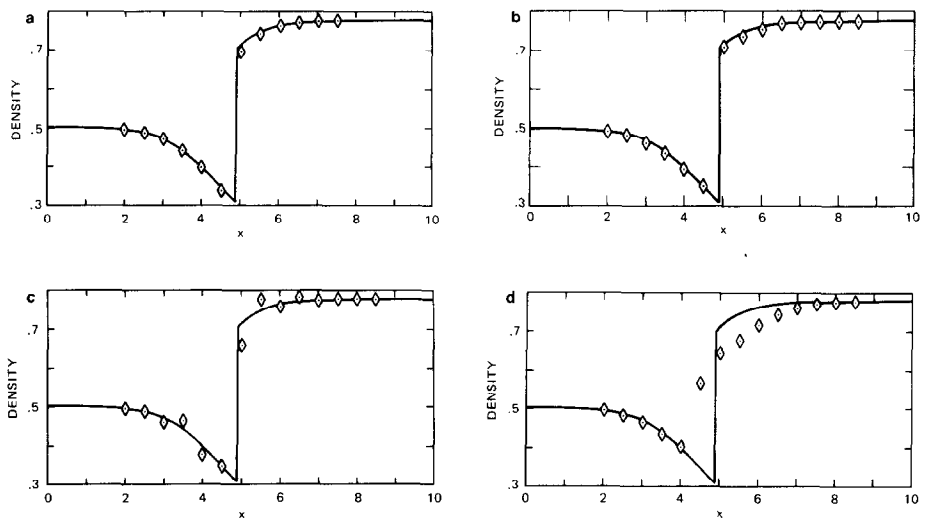


FIG. 4.2. Density distribution: supersonic inflow, subsonic outflow; 20 spatial intervals; (———), exact; (◇), numerical. (a) Explicit TVD method, CFL = 0.8, 700 steps; (b) implicit TVD method, CFL = $10^6$, 25 steps; (c) conventional implicit method, CFL = 10, 50 steps; (d) first-order implicit flux-vector splitting method, CFL = $10^6$, 25 steps.

$\delta = 0.125$ in Eq. (2.17b)). From numerical experiments, we found that artificial compression is not necessary for this one-dimensional problem.

The convergence criterion is based on the condition that all the components of the right-hand side of Eq. (3.8b) are less than or equal to $10^{-4}$ for all grid points. The implicit TVD method requires approximately triple the CPU time per time-step of the explicit TVD method but results in an enhanced convergence rate.

A steady-state solution can be reached in 25–30 steps with CFL ranges from $10^6$ to $10^7$. The steady-state solution profiles are independent of the CFL number. The number of steps for convergence monotonically decreases as the CFL number increases. However, the reduction in the number of steps is less pronounced for CFL number in the range from $10^3$ to $10^6$. There are four primary factors affecting the convergence rate for CFL numbers higher than $10^3$: (1) the initial condition (or the initial guess), (2) the numerical boundary conditions, (3) the interaction of nonlinear waves, and (4) the machine accuracy of the VAX 11/780.

Figure 4.2c shows the converged solution by a conventional implicit method with CFL = 10. The oscillation near the shock is typical of a three-point central spatial difference scheme. The experimentally determined maximum CFL number is around 10 with the above initial and numerical boundary conditions. (We can improve the stability by adding an implicit second-order numerical dissipation term.) Figure 4.2d shows the converged solution by the first-order flux-vector splitting method [20] with CFL = $10^6$. Again, the steady state is reached after 25–30 time-steps. The solution is very smeared but is independent of the CFL number. From the results, we can see a definite improvement in shock resolution by the

TVD schemes over the conventional methods. The implicit TVD scheme requires approximately 80% more CPU time per time-step than the conventional implicit method. Moreover, there is a dramatic increae in the convergence rate of the implicit TVD scheme over the explicit TVD scheme.

## 5. Applications to Two-Dimensional Compressible Inviscid Equations of Gas Dynamics

In this section we describe how to formally extend the one-dimensional implicit TVD scheme to an *alternating direction implicit* (ADI) version for the two-dimensional compressible inviscid equations of gas dynamics (Euler equations). This is a *formal* extension of the TVD scheme from one dimension to two dimensions. At the present state of development, there is not yet a similar theoretical analysis of the TVD properties for the approximation of two-dimensional hyperbolic equations. In here, we explore the use of an ADI approach as a possible relaxation procedure to obtain a steady-state numerical solution. Included in this section are (1) a detailed description of the analogue of each of the terms in Eqs. (4.2)–(4.6) for the two-dimensional Euler equations, (2) a discussion on the extension of the explicit and implicit TVD schemes to two dimensions, (3) Roe's form for the Jacobian matrices $A(U_{j,k}, U_{j+1,k})$ and $B(U_{j,k}, U_{j,k+1})$, (4) an algorithm to compute $U_{j,k}^{n+1}$ by the ADI approach, (5) a description of a numerical example, and (6) a discussion of the numerical results.

### 5.1. *Numerical Fluxes in Two Dimensions*

In two spatial dimensions, the Euler equations of gas dynamics can be written in the conservative form

$$\frac{\partial U}{\partial t}+\frac{\partial F(U)}{\partial x}+\frac{\partial G(U)}{\partial y}=0, \tag{5.1a}$$

where

$$U=\begin{bmatrix}\rho\\ m\\ n\\ e\end{bmatrix}, \qquad F=\begin{bmatrix}m\\ m^2/\rho+p\\ mv\\ (e+p)\,m/\rho\end{bmatrix}, \qquad G=\begin{bmatrix}n\\ nu\\ n^2/\rho+p\\ (e+p)\,n/\rho\end{bmatrix}, \tag{5.1b}$$

with $m=\rho u$ and $n=\rho v$. The primitive variables are the density $\rho$, the velocity components $u$ and $v$, and the pressure $p$. The total energy per unit volume $e$ is related to $p$ by the equation of state for a perfect gas

$$p=(\gamma-1)\left[e-\frac{(m^2+n^2)}{2\rho}\right], \tag{5.1c}$$

where $\gamma$ is the ratio of specific heats.

Let $A$ denote the Jacobian matrix $\partial F(U)/\partial U$ whose eigenvalues are

$$(a_x^1, a_x^2, a_x^3, a_x^4) = (u-c, u, u+c, u), \tag{5.2}$$

where $c$ is the local speed of sound. The right eigenvectors of $A$ form the matrix $R_x = (R_x^1, R_x^2, R_x^3, R_x^4)$ given by

$$R_x = \begin{bmatrix} 1 & 1 & 1 & 0 \\ u-c & u & u+c & 0 \\ v & v & v & 1 \\ H-uc & (u^2+v^2)/2 & H+uc & v \end{bmatrix}, \tag{5.3a}$$

where

$$H = \frac{c^2}{\gamma-1} + \frac{u^2+v^2}{2} \tag{5.3b}$$

and

$$R_x^{-1} = \begin{bmatrix} \frac{1}{2}\left(b_1 + \frac{u}{c}\right) & \frac{1}{2}\left(-b_2 u - \frac{1}{c}\right) & \frac{1}{2}\left(-b_2 v\right) & \frac{1}{2}b_2 \\ 1-b_1 & b_2 u & b_2 v & -b_2 \\ \frac{1}{2}\left(b_1 - \frac{u}{c}\right) & \frac{1}{2}\left(-b_2 u + \frac{1}{c}\right) & \frac{1}{2}\left(-b_2 v\right) & \frac{1}{2}b_2 \\ -v & 0 & 1 & 0 \end{bmatrix} \tag{5.3c}$$

with

$$b_1 = b_2 \frac{(u^2+v^2)}{2}, \tag{5.3d}$$

$$b_2 = \frac{\gamma-1}{c^2}. \tag{5.3e}$$

Let the grid spacing be denoted by $\Delta x$ and $\Delta y$ such that $x = j\Delta x$ and $y = k\Delta y$. Using the same notation as in Section 4, the vector $\alpha$ of Eq. (3.6) for the $x$ direction (omitting the $k$ index) is

$$\begin{bmatrix} \alpha_{j+1/2}^1 \\ \alpha_{j+1/2}^2 \\ \alpha_{j+1/2}^3 \\ \alpha_{j+1/2}^4 \end{bmatrix} = \begin{bmatrix} (aa-bb)/2 \\ \Delta_{j+1/2}\rho - aa \\ (aa+bb)/2 \\ \Delta_{j+1/2} n - v_{j+1/2}\Delta_{j+1/2}\rho \end{bmatrix}, \tag{5.4a}$$

where

$$aa = \frac{\gamma - 1}{c_{j+1/2}^2}\left[\Delta_{j+1/2}e + \frac{u_{j+1/2}^2 + v_{j+1/2}^2}{2}\Delta_{j+1/2}\rho - u_{j+1/2}\Delta_{j+1/2}m - v_{j+1/2}\Delta_{j+1/2}n\right], \tag{5.4b}$$

$$bb = [\Delta_{j+1/2}m - u_{j+1/2}\Delta_{j+1/2}\rho]/c_{j+1/2}. \tag{5.4c}$$

Similarly, let $B$ denote the Jacobian matrix $\partial G(U)/\partial U$ whose eigenvalues are

$$(a_y^1, a_y^2, a_y^3, a_y^4) = (v - c, v, v + c, v), \tag{5.5}$$

where $c$ is the local speed of sound. The right eigenvectors of $B$ form the matrix $R_y = (R_y^1, R_y^2, R_y^3, R_y^4)$ given by

$$R_y = \begin{bmatrix} 1 & 1 & 1 & 0 \\ u & u & u & 1 \\ v - c & v & v + c & 0 \\ H - vc & (u^2 + v^2)/2 & H + vc & u \end{bmatrix} \tag{5.6a}$$

and

$$R_y^{-1} = \begin{bmatrix} \frac{1}{2}\left(b_1 + \frac{v}{c}\right) & \frac{1}{2}(-b_2 u) & \frac{1}{2}\left(-b_2 v - \frac{1}{c}\right) & \frac{1}{2}b_2 \\ 1 - b_1 & b_2 u & b_2 v & -b_2 \\ \frac{1}{2}\left(b_1 - \frac{v}{c}\right) & \frac{1}{2}(-b_2 u) & \frac{1}{2}\left(-b_2 v + \frac{1}{c}\right) & \frac{1}{2}b_2 \\ -u & 1 & 0 & 0 \end{bmatrix}, \tag{5.6b}$$

with $b_1$ and $b_2$ defined in Eqs. (5.3d) and (5.3e). The vector $\alpha$ for the $y$ direction (omitting the $j$ index) is

$$\begin{bmatrix} \alpha_{k+1/2}^1 \\ \alpha_{k+1/2}^2 \\ \alpha_{k+1/2}^3 \\ \alpha_{k+1/2}^4 \end{bmatrix} = \begin{bmatrix} (cc - dd)/2 \\ \Delta_{k+1/2}\rho - cc \\ (cc + dd)/2 \\ \Delta_{k+1/2}m - u_{k+1/2}\Delta_{k+1/2}\rho \end{bmatrix}, \tag{5.7a}$$

where

$$cc = \frac{\gamma - 1}{c_{k+1/2}^2}\left[\Delta_{k+1/2}e + \frac{u_{k+1/2}^2 + v_{k+1/2}^2}{2}\Delta_{k+1/2}\rho - u_{k+1/2}\Delta_{k+1/2}m - v_{k+1/2}\Delta_{k+1/2}n\right], \tag{5.7b}$$

$$dd = [\Delta_{k+1/2}n - v_{k+1/2}\Delta_{k+1/2}\rho]/c_{k+1/2}. \tag{5.7c}$$

As mentioned previously, the simplest form for $U_{j+1/2,k}$ is

$$U_{j+1/2,k} = (U_{j+1,k} + U_{j,k})/2. \tag{5.8}$$

Roe's special form of the averaging in the $x$ direction is

$$u_{j+1/2,k} = \frac{\bar{D}u_{j+1,k} + u_{j,k}}{\bar{D}+1}, \tag{5.9a}$$

$$v_{j+1/2,k} = \frac{\bar{D}v_{j+1,k} + v_{j,k}}{\bar{D}+1}, \tag{5.9b}$$

$$H_{j+1/2,k} = \frac{\bar{D}H_{j+1,k} + H_{j,k}}{\bar{D}+1}, \tag{5.9c}$$

$$c^2_{j+1/2,k} = (\gamma-1)\left[H_{j+1/2,k} - \frac{1}{2}(u^2_{j+1/2,k} + v^2_{j+1/2,k})\right], \tag{5.9d}$$

$$\bar{D} = \sqrt{\rho_{j+1,k}/\rho_{j,k}}, \tag{5.9e}$$

$$H = \frac{\gamma p}{(\gamma-1)\rho} + \frac{1}{2}(u^2+v^2). \tag{5.9f}$$

Therefore to use Roe's averaging for the $x$ differencing, all we have to do is compute $u_{j+1/2,k}$, $v_{j+1/2,k}$, and $c_{j+1/2,k}$, in (5.2)–(5.4) by (5.9). Similarly, we can obtain Roe's averaging for $u_{j,k+1/2}$, $v_{j,k+1/2}$, and $c_{j,k+1/2}$. In the numerical experiments for the two-dimensional test problem, we use Roe's averaging.

The two-dimensional form of numerical fluxes (3.7b) for the Euler equations of gas dynamics (5.1) can be written as

$$\tilde{F}^n_{j+1/2,k} = \frac{1}{2}[F(U_{j,k}) + F(U_{j+1,k})] + \frac{1}{2}\sum_{l=1}^{4}[g^l_j + g^l_{j+1} - Q(a^l_{j+1/2} + \gamma^l_{j+1/2})\alpha^l_{j+1/2}]R^l_{j+1/2} \tag{5.10a}$$

$$\tilde{G}^n_{j,k+1/2} = \frac{1}{2}[G(U_{j,k}) + G(U_{j,k+1})] + \frac{1}{2}\sum_{l=1}^{4}[g^l_k + g^l_{k+1} - Q(a^l_{k+1/2} + \gamma^l_{k+1/2})\alpha^l_{k+1/2}]R^l_{k+1/2} \tag{5.10b}$$

with $g^l_{j+1}$ and $g^l_{k+1}$ corresponding to (3.7c) evaluated at the appropriate "locally frozen" $l$th characteristic speed and $l$th characteristic variable in the $x$ and $y$ direction, and $Q(z)$ is defined in (2.17b). The $\alpha^l_{j+1/2}$ and $\alpha^l_{k+1/2}$ are defined in (5.4) and (5.7). Also, the following notations have been used:

$$a^l_{j+1/2} \equiv (a^l_x)_{j+1/2,k}, \tag{5.10c}$$

$$a^l_{k+1/2} \equiv (a^l_y)_{j,k+1/2}, \tag{5.10d}$$

$$\gamma^l_{j+1/2} \equiv (\gamma^l_x)_{j+1/2,k}, \tag{5.10e}$$

$$\gamma^l_{k+1/2} \equiv (\gamma^l_y)_{j,k+1/2}, \tag{5.10f}$$

where $a^l_x$ is defined in (5.2), $a^l_y$ is defined in (5.5), and

$$\gamma^l_{j+1/2} = \begin{cases} (g^l_{j+1} - g^l_j)/\alpha^l_{j+1/2}, & \alpha^l_{j+1/2} \neq 0, \\ 0, & \alpha^l_{j+1/2} = 0 \end{cases} \tag{5.10g}$$

and

$$\gamma^l_{k+1/2} = \begin{cases} (g^l_{k+1} - g^l_k)/\alpha^l_{k+1/2}, & \alpha^l_{k+1/2} \neq 0, \\ 0, & \alpha^l_{k+1/2} = 0. \end{cases} \tag{5.10h}$$

Here, it is understood that the scalar values and the vector $R^l$ in the summation in Eqs. (5.10a) and (5.10b) are values of (5.2)–(5.9) evaluated at the corresponding $x$ and $y$ coordinates. For simplicity, we omitted the $k$ index inside the summation sign of Eq. (5.10a), and omitted the $j$ index inside the summation sign of Eq. (5.10b).

5.2. *Extension of the Explicit TVD Scheme by the Fractional Step Method*

In this subsection, we are going to review the extension of the explicit TVD scheme to two dimensions by the fractional step (time-splitting) method. We will also give a discussion on the use of the artificial compression term. Later, we will show a comparison between the explicit and implicit TVD schemes.

The explicit TVD scheme can be implemented in two space dimensions by the method of fractional steps as

$$U^*_{j,k} = U^n_{j,k} - \frac{\Delta t}{\Delta x}(\tilde{F}^n_{j+1/2,k} - \tilde{F}^n_{j-1/2,k}) = \mathscr{L}_x U^n_{j,k}, \tag{5.11a}$$

$$U^{n+1}_{j,k} = U^*_{j,k} - \frac{\Delta t}{\Delta y}(\tilde{G}^*_{j,k+1/2} - \tilde{G}^*_{j,k-1/2}) = \mathscr{L}_y U^*_{j,k}, \tag{5.11b}$$

that is,

$$U^{n+1}_{j,k} = \mathscr{L}_y \mathscr{L}_x U^n_{j,k}, \tag{5.12}$$

where $\tilde{F}^n_{j+1/2,k}$ and $\tilde{G}^n_{j,k+1/2}$ are defined in (5.10).

To retain the original time accuracy of the method, we use a Strang type of fractional step operators, namely,

$$U^{n+1}_{j,k} = \mathscr{L}^*_x \mathscr{L}^*_y \mathscr{L}^*_y \mathscr{L}^*_x U^n_{j,k} \tag{5.13}$$

or for steady-state calculations, we use

$$U_{j,k}^{n+2} = \mathscr{L}_x^* \mathscr{L}_y \mathscr{L}_x \mathscr{L}_y \cdots \mathscr{L}_y \mathscr{L}_x^* U_{j,k}^0, \tag{5.14}$$

where $\mathscr{L}_x^*$ denotes the operator with the time-step equal to $\Delta t/2$ and $U_{j,k}^0$ is the initial condition.

We may enhance resolution the same way as discussed in Section 2.5 by increasing the value of $\gamma_{j+1/2}^l$ (and $\gamma_{k+1/2}^l$) in Eqs. (5.10g) and (5.10h). To accomplish this, we increase the size of the corresponding $g_j^l$ in (3.7c), for example, by multiplying the right-hand side of (3.7c) by $[1+\omega^l\theta_j^l]$; i.e.,

$$\tilde{g}_j^l = [1+\omega^l\theta_j^l]\, g_j^l, \qquad \omega^l > 0, \tag{5.15a}$$

with

$$\theta_j^l = \frac{|\alpha_{j+1/2}^l - \alpha_{j-1/2}^l|}{|\alpha_{j+1/2}^l| + |\alpha_{j-1/2}^l|}, \tag{5.15b}$$

where $\alpha_{j+1/2}^l$ is defined by (5.4a). Similarly, we can obtain $\tilde{g}_k^l = [1+\omega^l\theta_k^l]\, g_k^l$ for the $y$ direction.

A preliminary experiment in two-dimensional calculations using an explicit TVD scheme by the fractional step approach indicated a need for such an enhancement mechanism [2]. The artificial compression term is especialy needed for the linearly degenerate characteristic field, i.e., $a_x = u$ and $a_y = v$. We will use this form of the $\tilde{g}$ function for our two-dimensional numerical experiments.

### 5.3. *Extension of the Implicit Scheme by the Alternating Direction Implicit Method*

The two-dimensional LNI form of (3.8) for the Euler equations of gas dynamics (5.1) can be written as

$$\begin{aligned}[I - \lambda^x J_{j+1/2,k}^- \Delta_{j+1/2,k} + \lambda^x J_{j+1/2,k}^+ \Delta_{j-1/2,k} \\ - \lambda^y K_{j,k+1/2}^- \Delta_{j,k+1/2} + \lambda^y K_{j,k-1/2}^+ \Delta_{j,k-1/2}](U^{n+1} - U^n) \\ = -\lambda^x[\tilde{F}_{j+1/2,k}^n - \tilde{F}_{j-1/2,k}^n] - \lambda^y[\tilde{G}_{j,k+1/2}^n - \tilde{G}_{j,k-1/2}^n]\end{aligned} \tag{5.16a}$$

with $\lambda^x = \Delta t/\Delta x$, $\lambda^y = \Delta t/\Delta y$, where

$$J_{j+1/2,k}^{\pm} = (R_x \operatorname{diag}(C_x^{\pm})\, R_x^{-1})_{j+1/2,k}^n, \tag{5.16b}$$

$$K_{j,k+1/2}^{\pm} = (R_y \operatorname{diag}(C_y^{\pm})\, R_y^{-1})_{j,k+1/2}^n, \tag{5.16c}$$

and

$$(C_x^{\pm})_{j+1/2,k}^n = \tfrac{1}{2}[Q(a_x^l + \gamma_x^l) \pm (a_x^l + \gamma_x^l)]_{j+1/2,k}^n, \tag{5.16d}$$

$$(C_y^{\pm})_{j,k+1/2}^n = \tfrac{1}{2}[Q(a_y^l + \gamma_y^l) \pm (a_y^l + \gamma_y^l)]_{j,k+1/2}^n.$$

$$l = 1, 2, 3, 4 \tag{5.16e}$$

It is well known that solving the two-dimensional implicit difference equation (5.16) is very costly. This leads to the popularity of using the alternating direction implicit (ADI) method to solve gas dynamics problems. Formally, we can write an ADI form of (5.16),

$$[I-\lambda^x J^-_{j+1/2,k}\Delta_{j+1/2,k}+\lambda^x J^+_{j-1/2,k}\Delta_{j-1/2,k}]\,D^* = -\lambda^x[\tilde{F}^n_{j+1/2,k}-\tilde{F}^n_{j-1/2,k}]-\lambda^y[\tilde{G}^n_{j,k+1/2}-\tilde{G}^n_{j,k-1/2}], \tag{5.17a}$$

$$[I-\lambda^y K^-_{j,k+1/2}\Delta_{j,k+1/2}+\lambda^y K^+_{j,k-1/2}\Delta_{j,k-1/2}]\,D=D^*, \tag{5.17b}$$

$$U^{n+1}=U^n+D. \tag{5.17c}$$

Note that there is not yet a theoretical analysis of the TVD properties for two-dimensional scalar conservation laws or constant coefficient hyperbolic systems. For reference purposes, we refer to (5.17) as the ADI form of the implicit TVD scheme for the "frozen" coefficient two-dimensional Euler equations.

Given $U^n_{j,k}$, for each $(j,k)$, we now list the operations needed to calculate $U^{n+1}_{j,k}$ by the ADI method (assuming a given CFL number):

(i) Compute $u_{j,k}=m_{j,k}/\rho_{j,k}$, $v_{j,k}=n_{j,k}/\rho_{j,k}$ and $p_{j,k}$.

(ii) Compute $u_{j+1/2,k}$, $v_{j+1/2,k}$, and $c_{j+1/2,k}$ from (5.8) or (5.9), calculate

$$M_x=\max_{j,k}(|u_{j+1/2,k}|+c_{j+1/2,k}), \qquad M_y=\max_{j,k}(|v_{j+1/2,k}|+c_{j+1/2,k})$$

evaluate $\alpha^l_{j+1/2}$, $l=1,2,3,4$, by (5.4a) (remember that the $k$ index is omitted from the equation), and define

$$\lambda^x=\Delta t/\Delta x=\mu/\max_{j,k}(M_x,M_y),$$

where $\mu$ is the prescribed CFL as input

(iii) Compute $(a^l_x)_{j+1/2,k}$ by (5.2), $g^l_{j+1/2,k}$ by (3.7c), with $(a^l_x)_{j+1/2,k}$ as $a^l_{j+1/2}$ in (3.7c).

(iv) Compute $(\gamma^l_x)_{j+1/2,k}$ by (5.10g), and $\tilde{F}_{j+1/2,k}$ by (5.10a) and relation (5.3a).

(v) Compute $C^{\pm}_x$ by (5.16d), with $z=(a^l_x)_{j+1/2,k}+(\gamma^l_x)_{j+1/2,k}$ in Eq. (2.18a).

(vi) Compute $J^{\pm}_{j+1/2,k}$ by (5.16b).

(vii) Compute $u_{j,k+1/2}$, $v_{j,k+1/2}$, and $c_{j,k+1/2}$, calculate

$$M_x=\max_{j,k}(|u_{j,k+1/2}|+c_{j,k+1/2}), \qquad M_y=\max_{j,k}(|v_{j,k+1/2}|+c_{j,k+1/2}),$$

evaluate $\alpha^l_{k+1/2}$, $l=1,2,3,4$, by (5.7a) (remember that the $j$ index is omitted from the equation), and define

$$\lambda^y=\Delta t/\Delta y=\mu/\max_{j,k}(M_x,M_y).$$

(viii) Compute $(a_y^l)_{j,k+1/2}$ by (5.5), $g^l_{j,k+1/2}$ by (3.7c), with $(a_y^l)_{j,k+1/2}$ as $a^l_{j+1/2}$ in (3.7c).

(ix) Compute $(\gamma_y^l)_{j,k+1/2}$ by (5.10h), and $\tilde{G}_{j,k+1/2}$ by (5.10b) and relation (5.6a).

(x) Solve Eq. (5.17a) for $D^*$.

(xi) Compute $C_y^{\pm}$ by (5.16e), with $z=(a_y^l)_{j,k+1/2}+(\gamma_y^l)_{j,k+1/2}$ in Eq. (2.18a).

(xii) Compute $K^{\pm}_{j,k+1/2}$ by (5.16c).

(xiii) Solve Eq. (5.17c) for $U^{n+1}_{j,k}$.

### 5.4. *Numerical Example*

To examine the applicability of the new method for two-dimensional shock calculations, we consider a simple inviscid flow field developed by a shock wave reflecting from a rigid surface (Fig. 5.1). The steady-state solution can be calculated exactly and thus can aid us in evaluating the quality of the numerical method. Figure 5.1 shows the indexing of the computational mesh.

*Initial conditions.* Initially, the entire flow field is set equal to the free stream supersonic inflow values plus the analytical boundary conditions as described below.

*Analytical boundary conditions.* The boundary conditions are given as follows: (a) supersonic inflow at $j=1$, $k=1,\ldots,K$, which allows the values $U_{1,k}$ to be fixed at free stream conditions; (b) prescribed fixed values of $U_{j,k}$ at $k=K$, $j=1,\ldots,J$, which produce the desired shock strength and shock angle; (c) supersonic outflow at $j=J$, $k=1,\ldots,K$; (d) a rigid flat surface at $k=1$, $j=1,\ldots,J$ which can be shown to be properly represented by the condition $v=0$, with the additional condition $\partial p/\partial y=0$ at $k=1$ from the normal $y$-momentum equation.

*Numerical boundary conditions.* The supersonic outflow values $U_{J,k}$, $k=1,\ldots,K-1$ are obtained by zeroth-order extrapolation, i.e.,

$$U_{J,k}=U_{J-1,k}, \qquad k=1,\ldots,K-1. \tag{5.18}$$

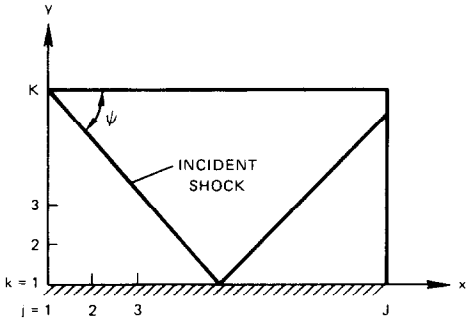


FIG. 5.1. Indexing of computational mesh for shock reflection problem.

The values of $\rho_{j,1}, m_{j,1}$ on the rigid surface with $j=1,..., J$ are also obtained by zeroth-order extrapolation, i.e.,

$$\rho_{j,1}=\rho_{j,2}, \qquad m_{j,1}=m_{j,2}, \quad j=1,..., J. \tag{5.19}$$

Three different methods were used in approximating $\partial p/\partial y=0$ to get $e_{j,1}$, $j=1,..., J$:

(a) Normal derivative of $e$ equal to zero (first order)

$$e_{j,1}=e_{j,2}. \tag{5.20a}$$

(b) Normal derivative of $e$ equal to zero (second order)

$$e_{j,1}=(4e_{j,2}-e_{j,3})/3. \tag{5.20b}$$

(c) Normal derivative of $p$ equal to zero (second order)

$$p_{j,1}=(4p_{j,2}-p_{j,3})/3, \tag{5.20c}$$

$$e_{j,1}=\frac{p_{j,1}}{(\gamma-1)}+\frac{m_{j,1}^2}{2\rho_{j,1}}. \tag{5.20d}$$

Equation (5.20a) together with Eq. (5.19) is an approximation to $p_{j,1}=p_{j,2}$, i.e., a first-order approximation for the normal derivative of $p$ equal to zero. Equation (5.20b) together with Eq. (5.19) is an approximation to Eq. (5.20c). We use Eq. (5.20a) or (5.20b) for the implicit method mainly because of their ease of application with implicit numerical boundary conditions. From the numerical experiments, we found that Eq. (5.20b) and (5.20c) produce better numerical solutions near the wall than Eq. (5.20a).

Since this implicit TVD scheme is a 5-point scheme (in each spatial direction), we also need the values of $g$ and $\theta$ at the boundaries. For convenience, we will use zeroth-order extrapolation.

### 5.5. *Discussion of Numerical Results*

The purpose of these numerical experiments is threefold:

(i) To test the performance of the ADI form of the implicit TVD scheme (5.17) on the shock reflection problem. For reference purposes, we denote scheme (5.17) as STVD.

(ii) To test the performance of scheme (5.17) with $\gamma_x^l=\gamma_y^l=0$ in (5.16d) and (5.16e) (i.e., the linearized form (2.35) with first-order spatial difference for the implicit operator) on the shock reflection problem. For reference purposes, we

(iii) To test the effect of the choice of $\theta$ and $Q$ function in (2.22c) and (2.22f), and the artificial compression term in (5.15) on the convergence rate and resolution of the above two-dimensional problem.

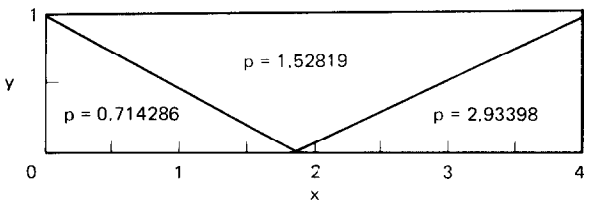


FIG. 5.2. The exact pressure solution for the shock reflection problem.

In all of the numerical experiments, the incident shock angle $\psi$ was 29° and the free stream Mach number $M_\infty$ was 2.9. The computational domain was $0 \leqslant x \leqslant 4.1$, and $0 \leqslant y \leqslant 1$, with a uniform grid size of $61 \times 21$. The appropriate analytical boundary conditions were applied along the boundaries of the domain. Only pressure contours and pressure coefficients will be illustrated. Here, the pressure coefficient is defined as

$$c_p = \frac{2}{\gamma M_\infty^2}\left(\frac{p}{p_\infty} - 1\right),$$

with $\gamma$ the ratio of specific heats, and $p_\infty$ the free stream pressure. The exact minimum pressure corresponding to $\psi = 29°$ and $M_\infty = 2.9$ is 0.714286 and the exact maximum pressure is 2.93398. The exact pressure solution and the computation domain is shown in Fig. 5.2. Forty-one pressure contour levels between the values of 0 and 4 with uniform increment 0.1 were used for the contour plots. The pressure coefficient was evaluated at $y = 0.5$ for $0 \leqslant x \leqslant 4.1$. All of the computations for the two-dimensional shock reflection problem were done in single precision on a CDC 7600 computer.

*Comparison of method.* Pressure contours and the pressure coefficients evaluated at $y = 0.5$ are shown in Fig. 5.3 for four different methods. Figure 5.3a shows the numerical result of the explicit TVD scheme by the fractional step method (5.14) with

$$Q(z) = z^2 + \tfrac{1}{4}, \tag{5.21a}$$

$$\sigma(z) = \tfrac{1}{8}. \tag{5.21b}$$

It took approximately 350 steps to converge with a fixed CFL = 0.8. The average smearing of the shocks is two points. We found that with Eq. (5.21), a slightly better shock resolution was obtained than $Q(z)$ in (2.17b) for the explicit method. Note that we can get (5.21b) by simply substituting (5.21a) into (2.28) with $\eta = 0$. Figure 5.3b shows the steady-state solution of a conventional implicit (ADI) method [18] with CFL = 1.0 after approximately 600 steps. The oscillations are spread over 8 grid points. The experimental maximum CFL number for this conventional implicit method is around 1.5. Figure 5.3c shows the numerical result of the ADI form of the implicit TVD scheme (5.17) (STVD method with $\delta = 0.125$ and $\sigma$ in Eq. (2.22e)) after 300 steps with CFL = 3. The experimental maximum CFL

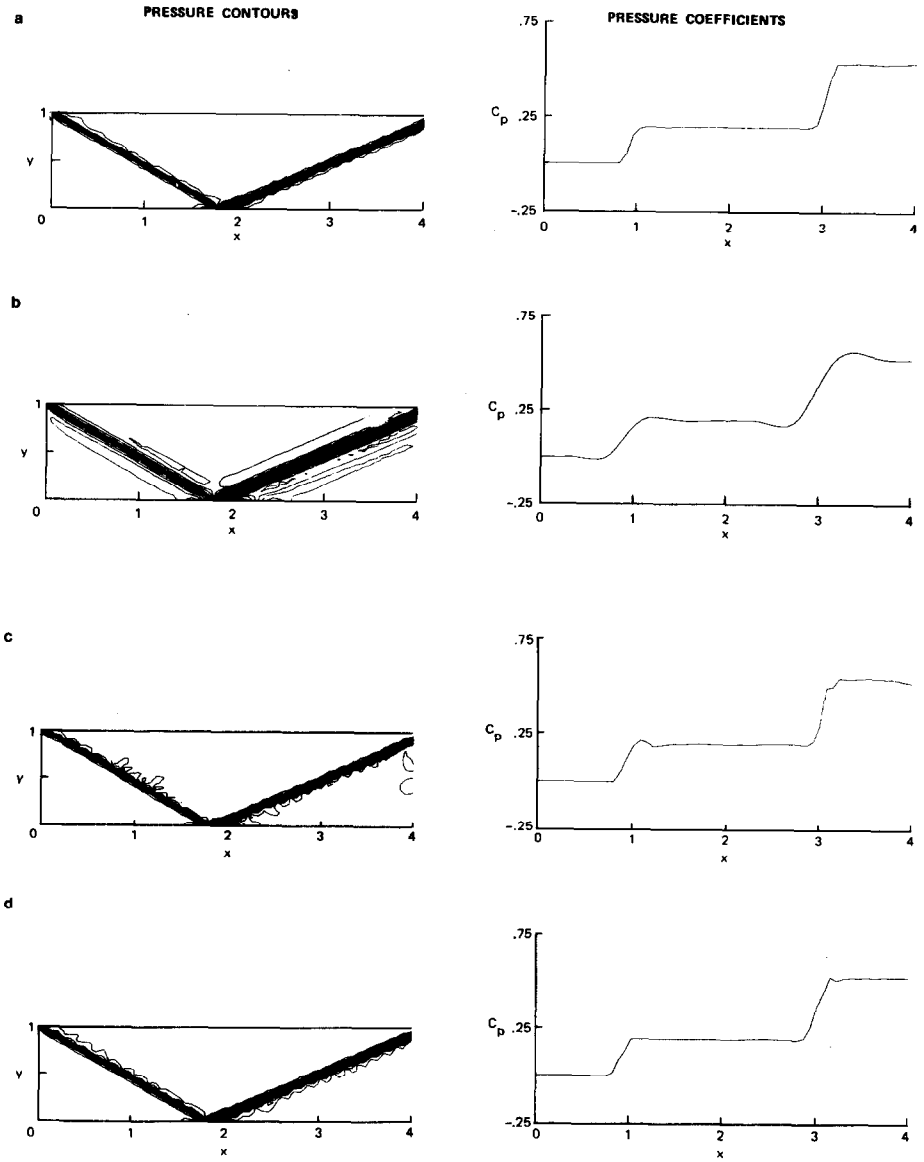


FIG. 5.3. Pressure contours and pressure coefficients for the shock reflection problem. (a) Explicit TVD method, 350 steps, CFL = 0.8; (b) conventional implicit method, 600 steps, CFL = 1; (c) implicit TVD method (STVD), 300 steps, CFL = 3; (d) implicit TVD method (FTVD), 60 steps, CFL = 6.

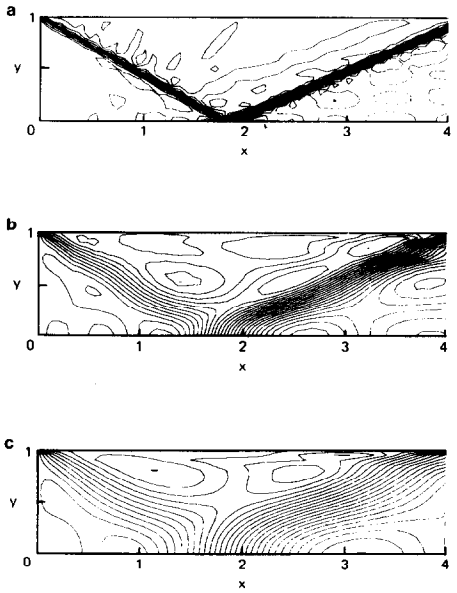


FIG. 5.4. Pressure contours for three different CFL numbers. (a) 250 steps, CFL = 50, (b) 250 steps, CFL = 500, (c) 250 steps, CFL = 1000.

number for this method is 5 under the CFL sampling sequence of (1, 2, 3, 5, 10, 20). Figure 5.3d shows the numerical result of the FTVD method (with $\delta = 0.125$ and $\sigma$ in Eq. (2.22e)) after 60 steps with CFL = 6. The average smearing of the shocks is 2 points. The convergence rate of the FTVD method is far better than the STVD method. It was found that the optimal CFL (with $\Delta t$ fixed) for fastest convergence rate with the FTVD method is between 5 and 10.

To show that the convergence rate is not a monotone decreasing function of CFL number for the FTVD method, Fig. 5.4a illustrates the same method as Fig. 5.3d after 250 steps with CFL = 50. The solution is not quite convergent yet. Figure 5.4b shows the same method after 250 steps with CFL = 500 (we ran for another 400 steps, without reaching steady state). Figure 5.4c shows the same method after 250 steps with CFL = 1000.

A primary factor affecting the stability and convergence rate of the ADI form of the implicit TVD scheme is the approximate factorization error.

*The choice of $\sigma$ and $Q$ functions, and the artificial compression term.* The primary difference in shock resolution between the explicit and implicit TVD methods of Figs. 5.3a and c is the result of using a different $\sigma$ and $Q$. Equations (2.22e) and (2.17b) are used for the implicit scheme and Eq. (5.21) is used for the explicit scheme. We have found that the shock resolution is somewhat degraded by the explicit TVD scheme if $\sigma$ in (2.22e) is used instead of (5.21). The particular choice of $\sigma$ in (2.22e) is especially appropriate for the ADI method since the steady-state solutions are independent of $\Delta t$.

All of the TVD schemes were operated with an artificial compression term (5.15a) with $\omega^l = 2$, $l = 1, 2, 3, 4$, except for the FTVD method. For the FTVD method, we used $\omega^1 = \omega^3 = 1$ and $\omega^2 = \omega^4 = 2$. We found that with this set of $\omega^l$, we get a faster convergence rate.

*Approximate CPU time and actual implementation of the numerical flux functions.* The conventional implicit method requires approximately twice the CPU time per time-step as the explicit TVD scheme. The implicit TVD method requires approximately 2.5 times more CPU time per step than the conventional implicit method.

In actual implementation of the explicit and implicit TVD methods into a computer code, the following form of the numerical flux $\tilde{F}^n_{j+1/2,k}$ (similarly for $\tilde{G}^n_{j,k+1/2}$) was used instead of (5.10),

$$\tilde{F}^n_{j+1/2,k} = \tfrac{1}{2}[F(U_{j,k}) + F(U_{j+1,k})] + \tfrac{1}{2}\sum_{l=1}^{4}[\xi^l_{j+1/2}\sigma^l_{j+1/2}(\bar{g}^l_j + \bar{g}^l_{j+1}) - Q(a^l_{j+1/2} + \gamma^l_{j+1/2})\alpha^l_{j+1/2}]R^l_{j+1/2}, \tag{5.22a}$$

with

$$\xi^l_{j+1/2} = 1 + \omega^l \max(\theta^l_j, \theta^l_{j+1}), \tag{5.22b}$$

where $\theta^l_j$ is defined in Eq. (5.15b), and

$$\begin{aligned} \bar{g}^l_j &= S \cdot \max[0, \min(|\alpha^l_{j+1/2}|, S \cdot \alpha^l_{j-1/2})], \\ S &= \operatorname{sign}(\alpha^l_{j+1/2}), \end{aligned} \tag{5.22c}$$

and

$$\gamma^l_{j+1/2} = \xi^l_{j+1/2}\sigma^l_{j+1/2}\begin{cases} (\bar{g}^l_{j+1} - \bar{g}^l_j)/\alpha^l_{j+1/2}, & \alpha^l_{j+1/2} \neq 0, \\ 0, & \alpha^l_{j+1/2} = 0. \end{cases} \tag{5.22d}$$

That is, the $\sigma^l_{j+1/2}$ has been taken out from the definition of $g^l_j$ in Eq. (3.7c) for simplicity and better resolution. Furthermore, the artificial compression is incorporated into the definition of the numerical flux function.

## 6. Concluding Remarks

The nonlinear, spatially second-order accurate, unconditionally stable implicit TVD scheme in a linearized nonconservative form has been applied to obtain steady-state solutions for the one-dimensional compressible inviscid equations of gas dynamics. This linearized form of the implicit TVD scheme is only conservative after the solution reaches steady state. Numerical experiments for a quasi-one-dimensional nozzle problem show that the experimentally determined stability limit correlates exactly with the theoretical stability limit for the nonlinear scalar hyperbolic conservation laws. Steady-state solution can be reached in approximately 25 steps.

We have formally extended the second-order accurate implicit TVD scheme by an ADI method for two-dimensional calculations. Numerical experiments with the ADI form (5.17) for a shock reflection problem show the gain in efficiency is not as pronounced as the one-dimensional counterpart.

A spatially first-order accurate left-hand side (i.e., by setting $\gamma_x' = \gamma_y' = 0$ in (5.11b) and (5.11c) of the ADI form provides better stability and a faster convergence rate. A steady-state solution can be reached in approximately 60 steps for a two-dimensional shock reflection problem. Numerical experiments also show that the rate of convergence is very sensitive to the CFL number. The iteration count grows rapidly when the calculation is carried out away from an optimal time-step.

More rigorous analyses of the influence of approximate factorization error and the influence of various ways of linearizing the implicit operator on the stability and efficiency of the method are needed. This will be the subject of a forthcoming paper [19].

## REFERENCES


1. R. D. RICHTMEYER AND K. W. MORTON, "Difference Methods for Initial-Value Problems," Interscience–Wiley, New York, 1967.
2. A. HARTEN, *J. Comput. Phys.* **49** (1983), 357.
3. A. HARTEN, "On a Class of High Resolution Total-Variation-Stable Finite-Difference Schemes," NYU Report, New York University, New York, October 1982.
4. A. HARTEN, J. M. HYMAN, AND P. D. LAX, *Commun. Pure Appl. Math.* **29** (1976), 297.
5. A. Y. LEROUX, *Math. Comput.* **31** (140) (1977), 848.
6. M. G. CRANDALL AND A. MAJDA, *Math. Comput.* **34** (149) (1980), 1.
7. B. VAN LEER, *J. Comput. Phys.* **32** (1979), 101.
8. P. K. SWEBY AND M. J. BAINES, "Convergence of Roe's Scheme for the General Non-linear Scalar Wave Equation," Numerical Analysis Report 8/81, University of Reading, 1981.
9. A. HARTEN AND J. M. HYMAN, "A Self-Adjusting Grid for the Computation of Weak Solutions of Hyperbolic Conservation Laws," Los Alamos Nat. Lab. Report LA9105, 1981.
10. H. C. YEE, R. F. WARMING, AND A. HARTEN, "On the Application and Extension of Harten's High-Resolution Scheme," NASA TM-84256, June 1982.

dynamic problems with weak solutions, *in* "Proceedings, Eighth International Conference on Numerical Methods in Fluid Dynamics" (E. Krause, Ed.), Lecture Notes in Physics, Vol. 170, Springer-Verlag, Berlin/New York, 1982.

12. P. COLELLA AND P. R. WOODWARD, "The Piecewise-Parabolic Method (PPM) for Gas-Dynamical Simulations," LBL Report, No. 14661, July 1982.
13. A. HARTEN, *Commun. Pure Appl. Math.* **30** (1977), 611.
14. A. HARTEN, *Math. Comput.* **32** (142) (1978), 363.
15. P. L. ROE, *J. Comput. Phys.* **43** (1981), 357.
16. G. R. SHUBIN, A. B. STEPHENS, AND H. M. GLAZ, *J. Comput. Phys.* **39** (1981), 364.
17. J. STEGER AND R. F. WARMING, *J. Comput. Phys.* **40** (1981), 263.
18. R. F. WARMING AND R. M. BEAM, On the construction and application of implicit factored schemes for conservation laws, *in* "SIAM–AMS Proceedings, Symposium on Comp. Fluid Dynamics," Vol. 11, New York, April 1978.
19. H. C. YEE AND A. HARTEN, Implicit TVD Schemes for Hyperbolic Conservation Laws in Curvillinear Coordinates, to appear.