A GEOMETRIC APPROACH TO HIGH RESOLUTION TVD SCHEMES*

JONATHAN B. GOODMAN† AND RANDALL J. LeVEQUE‡

Abstract. We use a geometric approach, similar to van Leer's MUSCL schemes, to construct a secondorder accurate generalization of Godunov's method for solving scalar conservation laws. By making suitable approximations we obtain a scheme which is easy to implement and total variation diminishing. We also investigate the entropy condition from the standpoint of the spreading of rarefaction waves. For Godunov's method we obtain quantitative information on the rate of spreading which explains the kinks in rarefaction waves often observed at the sonic point.

Key words. high resolution, TVD methods, conservation laws

AMS(MOS) subject classifications. 65M05, 76N10

1. Introduction. In solving hyperbolic conservation laws of the form

$$(1.1) u_t + f(u)_x = 0,$$

it is desirable to have a method that is at least second-order accurate in smooth regions of the flow and that also gives sharp resolution of discontinuities with no spurious oscillations.

There are two basic approaches which have been used to derive difference schemes for these problems. The first is purely algebraic. A scheme is defined by its coefficients or numerical fluxes, and algebraic relations are derived which guarantee that certain desirable properties hold.

The second approach is more geometrical, in that the structure of certain special solutions to (1.1) is heavily used. The classic method of this type is Godunov's method [3], based on the solution to Riemann problems. Van Leer [10], [11], with his MUSCL schemes, generalized this method to second-order accuracy by using discontinuous piecewise linear approximations at each time step. Higher-order geometrical methods, such as the piecewise parabolic method [1], have also been used.

Recently, however, most theoretical progress toward deriving second-order schemes with desirable properties has been made using the algebraic approach. Harten [4] introduced the concept of a total variation diminishing (TVD) scheme which guarantees that monotone profiles remain monotone. He derived conditions on the coefficients of a scheme that guarantee the TVD property and second-order accuracy. Since then many other second-order accurate TVD schemes have been constructed, e.g., [13], [15], [17], [18], but always using algebraic methods.

Our goal here is to derive a simple second-order accurate TVD scheme using a geometric approach in the spirit of van Leer's work. This is accomplished by choosing the slopes properly in the piecewise linear approximation and then by also approximating the flux function f in (1.1) by a piecewise linear function. The purpose of the latter approximation is that the modified equation with piecewise linear initial data can be efficiently solved analytically.

^{*} Received by the editors March 18, 1985; accepted for publication (in revised form) April 14, 1987.

[†] Courant Institute for Mathematical Sciences, New York University, New York, New York 10012. Present address, Department of Mathematics, University of Washington, Seattle, Washington 98195. The work of this author was supported in part by National Science Foundation grant MCS-82-01599.

[‡] University of California, Los Angeles, California 90024. The work of this author was supported in part by National Science Foundation grant MSC-82-00788 and in part by the National Aeronautics and Space Administration under NASA contract NAS1-17070 while he was in residence at the Institute for Computer Applications in Science and Engineering, NASA Langley Research Center, Hampton, Virginia 23665.

The method we derive is written in conservation form and can be viewed, algebraically, as a "limited" version of the Lax-Wendroff method. For a linear problem it agrees with one of the flux-limiter methods studied by Roe [17] and Sweby [18] but differs for nonlinear problems.

One advantage, we feel, of the geometric aproach is that it gives more insight into the behavior of algorithms. It may make it easier to show, for example, that the resulting numerical solution satisfies the entropy condition. Toward this end we choose a geometric form of the entropy condition, namely that solutions satisfy the spreading estimate

(1.2)
$$\frac{u(x,t) - u(y,t)}{x - y} \le \frac{c}{t} \quad \text{if } x > y$$

for some constant c. Oleinik [12] has shown that weak solutions to (1.1) satisfying (1.2) are unique.

In § 3 we prove (1.2) for the approximations produced by Godunov's method. The analysis shows, moreover, that these grid functions satisfy (1.2) with the correct constant c away from sonic points and points where the CFL condition is binding. Thus, Godunov's method spreads rarefaction waves at the physically correct rate most of the time. At the sonic point, (1.2) is satisfied with a constant that is two to four times larger. This causes a sonic rarefaction to develop a kink or "dog-leg" at the sonic point, as has frequently been observed without explanation in computations.

Unfortunately, the technical arguments used to prove (1.2) for Godunov's method do not carry over directly to the second-order scheme so that we have not obtained the spreading estimate in this case. However, by considering the sonic rarefaction case we will argue that entropy-violating shocks cannot persist. Moreover, numerical results look very good with spreading at the correct rate everywhere, including at the sonic point.

2. Godunov's method and second-order extensions. We consider the scalar version of equation (1.1) and will always assume that the flux function f is convex: f''>0. We will denote the numerical approximation to the solution $u(x_j, t_n)$ by U_j^n . Here $x_j = jh$ and $t_n = nk$ where h and k are the mesh width and time step, respectively. Since we will be discussing formulas for a single step from t_n to t_{n+1} we will generally drop the superscripts and replace U_j^n and U_j^{n+1} by U_j and \overline{U}_j , respectively.

To take a single step with Godunov's method, a piecewise constant function $w(x, t_n)$ is defined which takes the value U_j in the interval $I_j = (x_{j-1/2}, x_{j+1/2})$. The equation (1.1) is then solved exactly up to time t_{n+1} with this initial data to obtain $w(x, t_{n+1})$. This can be done easily if k is sufficiently small by solving a sequence of Riemann problems. The new approximation \bar{U}_j is then obtained by averaging the solution $w(x, t_{n+1})$ over I_j :

(2.1)
$$\bar{U}_j = \frac{1}{h} \int_{x_{i-1/2}}^{x_{j+1/2}} w(x, t_{n+1}) dx.$$

The method can also be rewritten in conservation form as

(2.2)
$$\bar{U}_{j} = U_{j} - \frac{k}{h} [F(U; j) - F(U; j-1)]$$

where

(2.3)
$$F(U; i) = \frac{1}{k} \int_{t_0}^{t_{n+1}} f(w(x_{i+1/2}, t)) dt$$

is the "numerical flux" across $x_{i+1/2}$, and depends only U_i and U_{i+1} . For a scalar conservation law this expression for the flux can be simplified. Assuming for convenience that we are away from the sonic point (so $f'(u) \neq 0$ for u between U_i and U_{i+1}) we have

(2.4)
$$F(U; i) = \begin{cases} f(U_i) & \text{if } f' > 0, \\ f(U_{i+1}) & \text{if } f' < 0. \end{cases}$$

The formula for the sonic case is only slightly more complicated. Using this formulation allows Godunov's method to be applied with any size time step for which the Courant number is less than one. For a scalar conservation law the Courant number ν is defined by

(2.5)
$$\nu = \frac{k}{h} \max |f'(u)|.$$

For $\nu < 1$, Godunov's method is TVD, i.e.,

$$TV(\bar{U}) \leq TV(U)$$

where the total variation is defined by

$$TV(U) = \sum_{j} |U_{j+1} - U_{j}|.$$

In modifying Godunov's method to obtain second-order accuracy, we follow the work of van Leer [10], [11] and replace the piecewise constant function $w(x, t_n)$ defined above by a piecewise linear function which we will denote by v. This replacement is conservative provided v is of the form

$$(2.6) v(x, t_n) = U_j + s_j(x - x_j) \text{for } x \in I_j.$$

We wish to pick the slopes s_j so that the total variation of v is the same as with $s_j \equiv 0$. One simple choice is

(2.7)
$$s_{j} = \begin{cases} 0 & \text{if } (U_{j+1} - U_{j})(U_{j} - U_{j-1}) \leq 0, \\ \operatorname{sgn} (U_{j+1} - U_{j}) \min \left\{ \left| \frac{U_{j+1} - U_{j}}{h} \right|, \left| \frac{U_{j} - U_{j-1}}{h} \right| \right\} & \text{otherwise.} \end{cases}$$

An example is shown in Fig. 1.

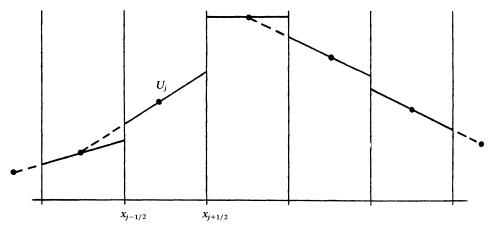


Fig. 1. The piecewise linear function $v(x, t_n)$ with slopes s_j given by (2.7). Dots denote the data values U_j .

Using this choice of slopes we can define the following algorithm for solving (1.1).

Algorithm 2.1.

- (1) Determine $v(x, t_n)$ based on $\{U_i\}$ using (2.7).
- (2) Solve (1.1) exactly with initial data $v(x, t_n)$ to obtain $v(x, t_{n+1})$.
- (3) Average $v(x, t_{n+1})$ as in (2.1) to obtain \overline{U}_j .

Each of these steps is total variation diminishing; so this defines a method which is TVD. Moreover, it can be shown that this method is second-order accurate in smooth regions, at least away from extreme points of u.

Unfortunately, this is not a practical method in most situations since it requires solving the conservation law (1.1) exactly with piecewise linear initial data. This is more difficult than solving Riemann problems. Various modifications can be made to Algorithm 2.1 to give a more readily implemented method.

Here we introduce a variant which remains second-order accurate and TVD and is easily implemented. We solve (1.1) with the piecewise linear initial data (2.6), but only after modifying the flux f in (1.1) to make this tractable. Specifically, we replace f by a piecewise linear function in such a fashion that computing the flux across $x_{j+1/2}$ reduces to solving a *linear* problem with piecewise linear data. The solution to this problem is easy to derive.

To compute the flux across $x_{j+1/2}$, we first compute the slopes (2.7) and consider the function $v(x, t_n)$ in (2.6). Since the sonic point causes difficulties, we delay discussion of this case to the end of this section and begin by assuming that $f'(v(x, t_n)) \neq 0$ for $x \in I_j \cup I_{j+1}$. Recall that we are always assuming f is convex.

Set

$$(2.8) U_i^{\pm} = U_i \pm \frac{1}{2} h s_i.$$

By virtue of our choice of slopes (2.7), the points U_j^- , U_j^+ , U_{j+1}^- , U_{j+1}^+ are monotonically ordered (though two or more may coincide). Let g(u) be a piecewise linear function which interpolates f(u) at these four points and set

(2.9)
$$g_i' = \begin{cases} [f(U_i^+) - f(U_i^-)]/(U_i^+ - U_i^-) & \text{if } s_i \neq 0, \\ f'(U_i) & \text{if } s_i = 0, \end{cases}$$

for i = j, j + 1, so g'_i is the slope of g(u) between U_i^- and U_i^+ . See Fig. 2.

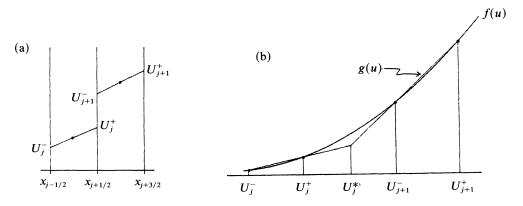


Fig. 2. (a) The piecewise linear function $v(x, t_n)$ over two mesh cells I_j and I_{j+1} . (b) The corresponding piecewise linear approximation g(u) to the flux function f(u). The numerical flux across $x_{j+1/2}$ is computed by solving $v_t + g(v)_x = 0$.

Now consider the problem

$$(2.10) v_t + g(v)_x = 0$$

for $t_n \le t \le t_{n+1}$ with the piecewise linear initial data $v(x, t_n)$. We can easily compute the flux across $x_{j+1/2}$ during $[t_n, t_{n+1}]$ for this problem, which we will denote G(U; j). Since we can rewrite (2.10) as $v_t + g'(v)v_x = 0$ and $g'(v(x_{j+1/2}, t))$ is constant $t_n \le t \le t_{n+1}$, we find that

(2.11)
$$v(x_{j+1/2}, t) = \begin{cases} U_j^+ - (t - t_n) s_j g_j' & \text{if } f' > 0, \\ U_{j+1}^- - (t - t_n) s_{j+1} g_{j+1}' & \text{if } f' < 0. \end{cases}$$

Furthermore,

$$(2.12) g(u) = \begin{cases} f(U_j^+) + (u - U_j^+)g_j' & \text{for } u \in \text{int } [U_j^-, U_j^+], \\ f(U_{j+1}^-) + (u - U_{j+1}^-)g_{j+1}' & \text{for } u \in \text{int } [U_{j+1}^-, U_{j+1}^+] \end{cases}$$

where int [a, b] denotes the interval with endpoints a and b. Using (2.3) and these formulas, we find that the flux is

(2.13a)
$$G(U;j) = \frac{1}{k} \int_{t_n}^{t_{n+1}} g(v(x_{j+1/2},t)) dt$$

$$= \begin{cases} f(U_j^+) - \frac{1}{2} k s_j (g_j')^2 & \text{if } f' > 0, \\ f(U_{j+1}^-) - \frac{1}{2} k s_{j+1} (g_{j+1}')^2 & \text{if } f' < 0. \end{cases}$$

Notice that if $s_j \equiv 0$ for all j, i.e., if we use piecewise constant initial data, we recover the flux (2.4) of Godunov's method. Also note the similarity of both (2.13a) and (2.13b) to the flux of the Lax-Wendroff method which can be written in the form (2.2) with flux

$$(2.14) F_{LW}(U;j) = \frac{1}{2} (f(U_j) + f(U_{j+1})) - \frac{1}{2} k \left(\frac{U_{j+1} - U_j}{h} \right) \left(\frac{f(U_{j+1}) - f(U_j)}{U_{j+1} - U_j} \right)^2.$$

In fact, for smooth solutions,

(2.15)
$$G(U; j) = F_{\text{LW}}(U; j) + O(h^2)$$

and the $O(h^2)$ term is a smooth function of j (except where $s_j = 0$, i.e., at extreme points of U). It follows that in computing

(2.16)
$$\bar{U}_{j} = U_{j} - \frac{k}{h} [G(U; j) - G(U; j-1)],$$

the $O(h^2)$ terms cancel to $O(h^3)$ showing that our scheme agrees with the Lax-Wendroff scheme to $O(h^3)$ locally and hence is second-order accurate in smooth regions, except near extrema of U (where it seems that all known second-order TVD schemes reduce to first-order accuracy [15]).

The use of the limited values s_j and g_j in (2.13) rather than the corresponding expressions in (2.14) gives us a method that, unlike that of Lax-Wendroff, is TVD. To see that this is so, note that it suffices to check the following conditions:

- (A) If U_j is a local maximum, i.e., $U_j \ge U_{j-1}$ and $U_j \ge U_{j+1}$ (resp. local minimum), then $\bar{U}_i \le U_i$ (resp. $\bar{U}_i \ge U_j$).
- (B) If $U_{j-1} \leq U_{j} \leq U_{j+1}$ (resp. $U_{j-1} \geq U_{j} \geq U_{j+1}$), then $U_{j-1}^{-} \leq \bar{U}_{j} \leq U_{j+1}^{+}$ (resp. $U_{j-1}^{-} \geq \bar{U}_{j} \geq U_{j+1}^{+}$).

We are still assuming that we are away from the sonic point, specifically that f'(u) has one sign on $I_{j-1} \cup I_j \cup I_{j+1}$. The sonic case is discussed below. So suppose, for example, that f' > 0 and that $U_i \ge U_{j-1}$ (all other cases are completely analogous).

If $U_j \ge U_{j+1}$, then we must check (A). In this case $s_j = 0$ and $G(U; j) = f(U_j)$ while $s_{j-1} \ge 0$ and

$$G(U; j-1) = f(U_{j-1}^+) - \frac{1}{2}ks_{j-1}(g_{j-1})^2 \le f(U_{j-1}^+) \le f(U_j).$$

So,

$$\bar{U}_j = U_j - \frac{k}{h} [G(U; j) - G(U; j-1)]$$

$$\leq U_i$$

and (A) is satisfied.

If $U_j \le U_{j+1}$ then we must check (B). We require, of course, that the Courant number be less than one; hence \bar{U}_j depends only on values of $v(x, t_n)$ in $I_{j-1} \cup I_j \cup I_{j+1}$. Since $U_{j-1} \le U_j \le U_{j+1}$ we have $s_i \ge 0$ for i = j-1, j, j+1, and

$$(2.17) U_{j-1}^- \le U_{j-1}^+ \le U_j^- \le U_j^+ \le U_{j+1}^- \le U_{j+1}^+.$$

The new value \bar{U}_j is determined by averaging the exact solution to (2.10). In our derivation we defined the piecewise linear flux g(u) locally; it had one definition in computing G(U;j) and a different definition in computing G(U;j-1). However, by the condition (2.17) these definitions are consistent in the region where they overlap, and so we can define a single function g(u) to compute both fluxes. Specifically, we can take

$$g(u) = \begin{cases} f(U_{j-1}^+) + (u - U_{j-1}^+) g_{j-1}' & \text{for } u \leq U_{j-1}^*, \\ f(U_j^+) + (u - U_j^+) g_j' & \text{for } U_{j-1}^* \leq u \leq U_j^*, \\ f(U_{j+1}^+) + (u - U_{j+1}^+) g_{j+1}' & \text{for } U_j^* \leq u, \end{cases}$$

where

$$U_i^* = (f(U_i^+) - f(U_{i+1}^+) + g_{i+1}'U_{i+1}^+ - g_i'U_i^+) / (g_{i+1}' - g_i')$$

for i = j - 1, j are the points of intersection of the piecewise linear segments. See Fig. 2. By the convexity of f and (2.17), we find that

$$U_{j-1}^+ \le U_{j-1}^* \le U_j^- \le U_j^+ \le U_j^* \le U_{j+1}^-$$

so that this definition is consistent with the previous definitions and gives the correct fluxes. Moreover, since $g'_{j-1} \le g'_j \le g'_{j+1}$, this piecewise linear approximation is also convex. It follows that \bar{U}_j is obtained from U_i , i = j - 1, j, j + 1, by solving a conservation law with a convex flux function and, hence, by the TVD property of the exact solution,

$$U_{j-1}^- \leqq \bar{U}_j \leqq U_{j+1}^+$$

so condition (B) is satisfied and the method is TVD.

We now turn to the sonic case and derive formulas for the flux across $x_{j+1/2}$ when the sonic point u_0 lies in $\inf[U_j^-, U_{j+1}^+]$. In some cases the previous formulas (2.13) are still valid and, to avoid repeating these expressions, we define

(2.18)
$$G_{j} = f(U_{j}^{+}) - \frac{1}{2}ks_{j}(g_{j}^{\prime})^{2},$$

$$G_{j+1} = f(U_{j+1}^{-}) - \frac{1}{2}ks_{j+1}(g_{j+1}^{\prime})^{2}.$$

We first note that if g'_j and g'_{j+1} have the same sign, then the linearized problem (2.10) can be solved just as before and the flux agrees with (2.13)

$$G(U;j) = \begin{cases} G_j & \text{if } g_j' \ge 0, \quad g_{j+1}' \ge 0, \\ G_{i+1} & \text{if } g_i' \le 0, \quad g_{j+1}' \le 0. \end{cases}$$

If g'_{j} and g'_{j+1} have different signs, then we must be more careful.

If $g_j' < 0$ and $g_{j+1}' > 0$, then the discontinuity at $x_{j+1/2}$ is a sonic rarefaction. By solving (2.10) in this case we find that the flux is $f(U_j^+)$ if $U_j^- \le u_0 \le U_j^+$ and $f(U_{j+1}^-)$ if $U_{j+1}^- \le u_0 \le U_{j+1}^+$. In the remaining case, $U_j^+ < u_0 < U_{j+1}^-$, the sonic point lies within the discontinuity in $v(x, t_n)$. In this case we must take special precautions to ensure that the rarefaction wave spreads properly and an entropy violating shock does not persist (see § 3). Rather than using the usual piecewise linear flux g(u) we include another interpolating point $(u_0, f(u_0))$ in g(u). The flux is then simply $f(u_0)$.

These last three expressions for the flux can be conveniently combined to give

$$G(U; j) = f(v_0)$$
 if $g'_i < 0, g'_{i+1} > 0$

where

(2.19)
$$v_0 = \min \left(\max \left(U_i^+, u_0 \right) m \ U_{i+1}^- \right).$$

Now suppose that $g'_j < 0$ and $g'_{j+1} > 0$. In order to solve the linearized problem (2.10) we must also specify g(u) for $U_j^+ \le u \le U_{j+1}^-$ in this case. We take another linear segment interpolating f at these points with slope

(2.20)
$$g'_{j+1/2} = \begin{cases} (f(U_{j+1}^-) - f(U_j^+) / (U_{j+1}^- - U_j^+) & \text{if } U_{j+1}^- \neq U_j^+, \\ f'(U_j^+) & \text{if } U_{j+1}^- = U_j^+. \end{cases}$$

In the solution to (2.10) the discontinuity at $x_{j+1/2}$ is now a shock which may propagate to the left or right or be stationary. If the shock moves to the left (for all $t \in (t_n, t_{n+1})$), then the flux across $x_{j+1/2}$ is G_{j+1} ; if it moves to the right, then the flux is G_j . Unfortunately, since the initial data is not constant on the two sides of the discontinuity, the shock may switch direction and cross $x_{j+1/2}$ at some time $t \in (t_n, t_{n+1})$ at which point the flux is discontinuous. This is most easily visualized by first considering the multi-valued solution obtained in solving the linearized problem and then inserting a shock according to the equal area rule. By taking the slopes s_j and s_{j+1} quite different one can construct examples where the shock is first on one side of $x_{j+1/2}$ and later on the other.

For simplicity we ignore this possibility and always use G_j or G_{j+1} depending on the *initial* motion of the shock. This is one situation in which we do not use the exact solution to the linearized problem, but experimentally this approximation seems to work well.

If $v(x, t_n)$ is discontinuous at $x_{j+1/2}$, i.e., if $U_j^+ \neq U_{j+1}^-$, then the initial motion of the shock is determined by the motion of the discontinuity in the multi-valued solution and hence by the sign of $g'_{j+1/2}$. We use

$$G(U; j) = \begin{cases} G_j & \text{if } g'_{j+1/2} > 0, \\ G_{j+1} & \text{if } g'_{j+1/2} < 0. \end{cases}$$

If $g'_{j+1/2} = 0$, then the discontinuity is stationary and the initial motion of the shock is determined by the relative sizes of $s_j(g'_j)^2$ and $s_{j+1}(g'_{j+1})^2$. Specifically,

$$G(U; j) = \begin{cases} G_j & \text{if } s_j(g_j')^2 \ge s_{j+1}(g_{j+1}')^2, \\ G_{j+1} & \text{otherwise.} \end{cases}$$

It turns out that this is also the appropriate formula if $U_j^+ = U_{j+1}^-$. In this case $s_j = s_{j+1}$ and we are simply comparing the magnitudes of g_j' and g_{j+1}' .

We have now derived formulas for every possible case. Luckily, all of these formulas can be summarized quite neatly as follows.

Algorithm 2.2.

$$\bar{U}_j = U_j - \frac{k}{h} [G(U;j) - G(U;j-1)]$$

where:

(1) If
$$g_j' > 0$$
, $g_{j+1/2}'(U_{j+1}^- - U_j^+) = 0$, and $g_{j+1}' < 0$:
$$G(U; j) = \begin{cases} G_j & \text{if } s_j(g_j')^2 \ge s_{j+1}(g_{j+1}')^2, \\ G_{j+1} & \text{otherwise.} \end{cases}$$

(2) Otherwise:

$$G(U;j) = \begin{cases} G_j & \text{if } g_j' \ge 0, & g_{j+1/2}' \ge 0, \\ G_{j+1} & \text{if } g_{j+1/2}' \le 0, & g_{j+1}' \le 0, \\ f(v_0) & \text{if } g_j' < 0, & g_{j+1}' > 0. \end{cases}$$

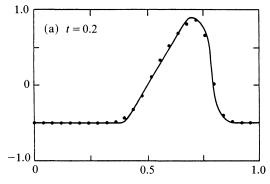
Here we have used the expressions (2.7), (2.8), (2.9), (2.18), (2.19) and (2.20). These formulas cover all cases: the sonic shock, sonic rarefaction and also the usual nonsonic case (2.13). However, in implementing this method it is of course best to use (2.13) whenever $g'_j g'_{j+1} > 0$. We should compute $g'_{j+1/2}$ and perform the various tests above only in the relatively rare sonic case.

It is possible to show that the method remains TVD even near sonic points when these formulas are used. This is done in precisely the same way as before but is slightly more complicated since several cases must be considered. We omit the details.

Numerical experiments confirm that the method is second-order accurate and TVD. To check the second-order accuracy we applied the method to Burgers' equation $u_t + uu_x = 0$ with smooth (sine wave) initial data and periodic boundary conditions. Both the L_1 and L_{∞} norm of the errors decrease at the correct rate as the mesh is refined.

Figures 3.1(a) and 3.1(b) show the results of a typical calculation in which a shock forms. Again the method is applied to Burgers' equation with initial data

$$u(x, 0) = \begin{cases} -0.5, & x < 0.5, \\ 0.2 + 0.7 \cos(2\pi x), & x > 0.5 \end{cases}$$



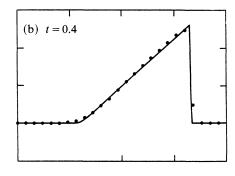


Fig. 3.1. The solution u(x, t) to Burgers' equation at two different times computed by Algorithm 2.2.

and periodic boundary conditions. The discontinuity at x = 0.5 spreads into a rarefaction fan, and the smooth decreasing profile sharpens into a shock. Figures 3.1(a) and 3.1(b) show the results at time t = 0.2 and t = 0.4, respectively. For comparison, Fig. 3.2 shows the results of Godunov's method. Notice the improved accuracy in the smooth portion of the solution with the second-order method and the lack of oscillations near the shock. Godunov's method suffers in particular from a lack of smoothness in the rarefaction wave at the sonic point $u_0 = 0$ which does not occur with the second-order scheme. This is discussed in § 3.

It is interesting to compare this method to the flux-limiter methods. We find that for a linear problem it is the same as one of the flux-limiter methods of [18] but that it differs for nonlinear problems.

First consider the linear problem

$$u_t + au_x = 0$$

with constant a > 0. Then according to (2.13),

$$G(U;j) = aU_j^+ - \frac{1}{2}ks_ja^2 = aU_j + \frac{1}{2}ha\left(1 - a\frac{k}{h}\right)s_j$$

so that

(2.21)
$$\bar{U}_{j} = U_{j} - \frac{k}{h} [G(U; j+1) - G(U; j)]$$

$$= U_{j} - \nu (U_{j+1} - U_{j}) - \frac{1}{2} h \nu (1 - \nu) (s_{j} - s_{j-1})$$

where $\nu = ak/h$. If we define r_i to be the ratio

$$(2.22) r_j = \frac{U_j - U_{j-1}}{U_{j+1} - U_j},$$

then, by virtue of (2.7),

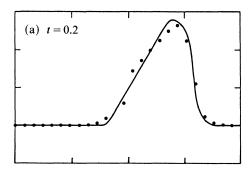
$$s_j = \phi(r_j)(U_{j+1} - U_j)$$

where

(2.23)
$$\phi(r) = \begin{cases} 0 & \text{if } r \leq 0, \\ r & \text{if } 0 \leq r \leq 1, \\ 1 & \text{if } r \geq 1. \end{cases}$$

Then (2.21) becomes

(2.24)
$$\bar{U}_j = U_j - \nu (U_{j+1} - U_j) - \frac{1}{2} (1 - \nu) \nu \Delta_- [\phi(r_j) (U_{j+1} - U_j)]$$



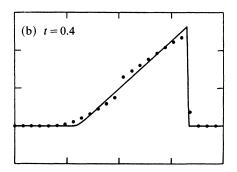


Fig. 3.2. The solution u(x, t) to Burgers' equation at two different times computed by Godunov's method.

where $\Delta_{-}\omega_{j} = \omega_{j} - \omega_{j-1}$. This is precisely the flux-limiter method of [18] with the so-called "minmod" limiter given by (2.23). We note in passing that other flux-limiter methods are given by different choices of the limiter ϕ and that taking $\phi(r) = 1$ for all r gives the Lax-Wendroff method.

The flux-limiter method is extended to the nonlinear problem (1.1) by generalizing (2.24) to

$$(2.25) \quad \bar{U}_j = U_j - \frac{k}{h} [f(U_j) - f(U_{j-1})] - \frac{1}{2} \frac{k}{h} \Delta_{-} [\phi(r_j)(1 - \nu_{j+1/2})(f(U_{j+1}) - f(U_j))]$$

in the case f' > 0, where

$$\nu_{j+1/2} = \frac{k}{h} \left[\frac{f(U_{j+1}) - f(U_j)}{U_{j+1} - U_j} \right]$$

and

$$r_{j} = \frac{(1 - \nu_{j-1/2})(f(U_{j}) - f(U_{j-1}))}{(1 - \nu_{j+1/2})(f(U_{j+1}) - f(U_{j}))}.$$

By contrast, our method in the same situation gives

$$\bar{U}_j = U_j - \frac{k}{h} \Delta_{-} \left[f(U_j^+) - \frac{1}{2} k s_j(\mathbf{g}_j')^2 \right].$$

Using the definitions of the various quantities appearing here, we can rearrange this to obtain a form similar to (2.25):

$$(2.26) \bar{U}_{j} = U_{j} - \frac{k}{h} (f(U_{j}) - f(U_{j-1})) - \frac{1}{2} \frac{k}{h} \Delta_{-} \left[\left(1 - \frac{k}{h} g_{j}' \right) (f(U_{j}^{+}) - f(U_{j})) \right].$$

This is very similar to (2.25) but the limiting is done in a different manner.

3. Spreading of rarefaction waves. Weak solutions to conservation laws are not necessarily unique. In general there is some additional condition, such as an entropy condition, required to identify the unique physically relevant solution [5], [9]. For the scalar conservation law (1.1) with a convex flux f, such conditions are well known in several equivalent forms. One form considered by Oleinik [12] requires that the solution satisfy the spreading estimate

(3.1)
$$u(x,t) - u(y,t) \le \frac{x-y}{at}$$

for all x > y and t > 0 where a > 0 is some constant. In fact, one can take $a = \alpha$ where

$$\alpha = \inf f''(u)$$

and $\alpha > 0$ by convexity. We can define α locally to obtain more precise information on the rate of spreading of rarefaction waves in different regions of the solution. Note that at points where $u_x(x, t)$ exists we obtain

$$(3.3) u_x(x,t) \leq \frac{1}{\alpha t}.$$

We would like to prove an estimate analogous to (3.1) for the numerical solutions generated by a particular scheme as the mesh is refined with k/h held fixed. If there exists a constant c > 0 such that

$$(3.4) U_{j+1}^n - U_j^n \leq \frac{1}{cn}$$

for each point on every grid, then the limit solution satisfies (3.1) with a = ch/k and hence satisfies the entropy condition. The possibility of proving estimates of this form was independently noticed by Tadmor [19]. He proved the estimate (3.4) for the Lax-Friedrichs scheme using essentially the same technique.

This form of the entropy condition seems easiest to deal with when studying second-order schemes of the type considered here. Moreover, by obtaining an estimate of the form (3.4) we can compare the rate of spreading in the numerical solution with the correct rate. Ideally we would like $c = \alpha k/h$ in (3.4).

Our interest in obtaining such quantitative information stems from the observation that rarefaction waves computed with some numerical methods, including Godunov's method, do not always spread at the proper rate in spite of the fact that the entropy condition is satisfied. This difficulty is most frequently observed at the sonic point. As we will show with Godunov's method, rarefaction waves spread with at best one-half the correct rate in this region. This leads to a kink in the rarefaction wave at the sonic point. This is frequently observed in practice and has been termed a "dog-leg" by Sweby [18]. For an example see Fig. 3.2 where we have applied Godunov's method to the Burgers equation $u_t + uu_x = 0$.

Although our main interest is in the second-order methods of § 2, we will begin by analyzing Godunov's method in some detail. This will provide a basis of comparison and also provides some insight and quantitative information on the dog-leg phenomenon.

Let

$$D^n = \max_j (U_{j+1}^n - U_j^n).$$

Then our goal is to find a constant c > 0 so that

$$(3.5) D^n \leq \frac{1}{cn}$$

for all n. In fact we will consider only a single time step and, as in § 2, replace D^n and D^{n+1} by D and \bar{D} , respectively. We will determine a constant c > 0 for which

$$(3.6) \bar{D} \leq D - cD^2$$

from which (3.5) follows by induction.

We also let

$$(3.7) D_i = U_{i+1} - U_i$$

and note that

$$(3.8) D_j \leq \frac{2h}{k\alpha}$$

for all time steps by virtue of the Courant number restriction $|(k/h)f'(u)| \le 1$. Here α is as in (3.2) and (3.8) follows from

$$\frac{k}{h}D_{j}\alpha \leq \frac{k}{h}\int_{U_{j}}^{U_{j+1}} f''(\xi) d\xi$$

$$= \frac{k}{h} [f'(U_{j+1}) - f'(U_{j})]$$

$$\leq 2.$$

Now consider the meshpoint x_j and suppose, to begin with, that f'(u) > 0 on $I_{j-1} \cup I_j \cup I_{j+1}$. Then we will show that

$$(3.9) \bar{D}_i \leq D - cD^2$$

with $c \ge \alpha k/2h$ and that, in fact, we can generally use $c = \alpha k/h$. Since f' > 0, applying Godunov's method (2.2) and (2.4),

(3.10)
$$\bar{U}_{j+1} = U_{j+1} - \frac{k}{h} [f(U_{j+1}) - f(U_j)],$$

$$\bar{U}_j = U_j - \frac{k}{h} [f(U_j) - f(U_{j-1})].$$

Subtracting these and using (3.7) gives

$$\bar{D}_{j} = D_{j} - \frac{k}{h} [f(U_{j+1}) - 2f(U_{j}) + f(U_{j-1})]$$

$$= D_{j} - \frac{k}{h} [(D_{j} - D_{j-1})f'(U_{j}) + \frac{1}{2}D_{j}^{2}f''(\xi_{j}) + \frac{1}{2}D_{j-1}^{2}f''(\xi_{j-1})]$$

where we have expanded both $f(U_{j+1})$ and $f(U_{j-1})$ in a Taylor series about U_j and $U_{j-1} \le \xi_{j-1} \le U_j \le \xi_j \le U_{j+1}$. Rearranging gives

$$(3.11) \quad \bar{D}_{j} = \left(1 - \frac{k}{h}f'(U_{j})\right)D_{j} + \frac{k}{h}f'(U_{j})D_{j-1} - \frac{1}{2}\frac{k}{h}[D_{j}^{2}f''(\xi_{j}) + D_{j-1}^{2}f''(\xi_{j-1})].$$

Since $0 < (k/h)f'(U_j) \le 1$, the first two terms are a convex combination of D_j and D_{j-1} and consequently are bounded by

$$(3.12) D_{\psi} = \max (D_{i-1}, D_i).$$

Dropping the term corresponding to the smaller of D_j and D_{j-1} from the sum in brackets in (3.11) and using (3.2) gives

(3.13)
$$\bar{D}_{j} \leq D_{*} - \frac{1}{2} \frac{k}{h} \alpha D_{*}^{2}.$$

The term on the right of this inequality is an increasing function of D_* for D_* satisfying (3.8), and so we can replace D_* by $D = \max D_j$ and conclude that (3.9) holds with $c = \alpha k/2h$.

Moreover, except near the extreme points of U (i.e., the edge of the rarefaction wave), we can do better than this. Typically in the interior of the rarefaction wave we have $D_{j-1} \approx D_j$.

If in fact

$$(3.14) D_{j-1} = D_j,$$

then by retaining both terms in the sum in brackets in (3.11) we remove the factor $\frac{1}{2}$ in (3.13) and obtain

$$(3.15) \bar{D}_j \leq D - \frac{k}{h} \alpha D^2.$$

Even if (3.14) does not hold exactly we can consider modifying the data by increasing U_{j+1} or decreasing U_{j-1} so that (3.14) does hold. Then (3.15) holds for the modified \overline{D}_j , but it is easy to verify (using the fact that Godunov's method is monotone) that

the original \bar{D}_j is bounded above by the modified \bar{D}_j so (3.15) holds for the original data. This argument works provided modification of the data does not violate the Courant number restriction, which it might near extreme points of U.

Exactly the same arguments can be applied to the case where f'(u) < 0 on $I_j \cup I_{j+1} \cup I_{j+2}$. We conclude that, away from the sonic point, rarefaction waves spread at the correct rate except perhaps near the edge of the rarefaction wave if the Courant number restriction is binding there.

Now consider the sonic case, where $U_{j-1} \le u_0 \le U_{j+1}$. Applying Godunov's method in this case gives

$$\bar{U}_{j+1} = U_{j+1} - \frac{k}{h} [f(U_{j+1}) - f(u_0)],$$

$$\bar{U}_j = U_j - \frac{k}{h} [f(u_0) - f(U_j)].$$

Subtracting gives

$$(3.16) \bar{D}_j = D_j - \frac{k}{h} F_2$$

where

$$F_2 = f(U_{i+1}) - 2f(u_0) + f(U_i).$$

Since $f'(u_0) = 0$, expanding $f(U_{i+1})$ and $f(U_i)$ about u_0 gives us

(3.17)
$$F_2 = \frac{1}{2} (\Delta_{i+1}^2 f''(\xi_{i+1}) + \Delta_i^2 f''(\xi_i))$$

where $\Delta_{j+1} = U_{j+1} - u_0$, $\Delta_j = u_0 - U_j$, and $U_j \le \xi_j \le u_0 \le \xi_{j+1} \le U_{j+1}$. Since $\Delta_j + \Delta_{j+1} = D_j$, we have

$$(3.18) \qquad \qquad \frac{1}{4}\alpha D_i^2 \leq F_2 \leq \frac{1}{2}\alpha D_i^2.$$

So from (3.16) we obtain (3.6) with a value of c which is at worst one-quarter the correct value and at best one-half the correct value. The worst case occurs when $\Delta_j = \Delta_{j+1} = \frac{1}{2}D_j$, i.e., if the sonic point falls halfway between U_j and U_{j+1} . The best case occurs when $\Delta_j \Delta_{j+1} = 0$ and either U_j or U_{j+1} is equal to u_0 .

Summarizing these results, we can obtain a global bound of the form (3.5) with $c = \alpha k/4h$ which shows that the entropy condition is satisfied. Moreover, we generally have spreading at the correct rate except near the sonic point, where the rate is at best one-half the correct rate.

These results can be seen more geometrically by considering Fig. 4, which for simplicity we have shown for Burgers' equation where rarefaction waves are linear in x. Moreover, we have taken $D_{j-1} = D_j$. Figure 4(a) shows the initial conditions U as the solid piecewise constant function. After solving Burgers' equation with this data, we obtain the dashed line. It is this function which is averaged to give \overline{U}_j (indicated by dots). Clearly U_j and U_{j+1} both decrease, but U_{j+1} decreases by a greater amount so that the difference D_j also decreases. A little reflection shows that D_j decreases by 1/h times the area of the shaded rectangle drawn in Fig. 4(b). The height of this rectangle is D_j , and the length is easily computed to be $kD_j\alpha$, where $\alpha = f'' = 1$ for Burgers' equation. So,

$$\bar{D}_j = D_j - \frac{k}{h} \alpha (D_j)^2$$

which agrees with (3.15).

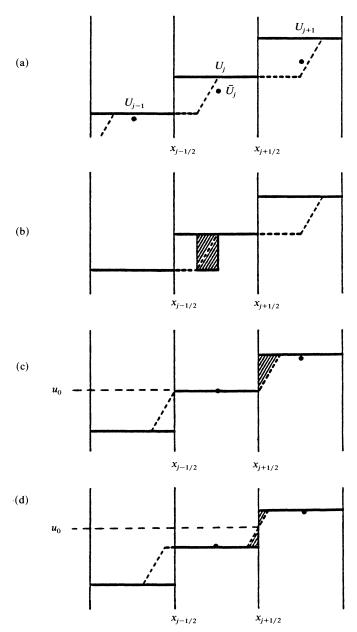


Fig. 4. (a) In Godunov's method, the exact solution (dashed line) is averaged to obtain \bar{U}_j (the dots). (b) The difference $D_j - \bar{D}_j$ is represented by the shaded area in the typical nonsonic case. (c) The sonic case with $u_0 = U_j$. (d) The sonic case with $u_0 = \frac{1}{2}(U_j + U_{j+1})$.

Now consider the sonic case and suppose first that $U_j = 0$. Then we have the situation in Fig. 4(c) and D_j decreases by 1/h times the shaded triangle, which is one-half the area of the rectangle in Fig. 4(b). This accounts for the spreading rate being one-half the correct rate in this case.

Finally, considering the worst possible sonic case, where $u_0 = \frac{1}{2}(U_j + U_{j+1})$, we obtain Fig. 4(d). Again, D_j decreases by 1/h times the shaded area, which is now one-quarter the original area.

We now turn to the second-order accurate scheme given by Algorithm 2.2. Unfortunately, we have not yet been able to prove the desired spreading estimates in general. The approach used above for Godunov's method does not apply directly since it is possible to construct initial data for which the difference D_n actually increases in a single time step. Example: An example of this is given below. This undesirable behavior does not persist in later time steps, and experimental evidence indicates that rarefaction waves do spread at the correct rate asymptotically, but clearly we can no longer use a bound of the form (3.9) to prove this.

Figure 5 gives an example where $D_{n+1} = \bar{D} \ge D_n = D$ for $u_t + u_x = 0$. Take $U_j = 0$, $j \le -1$, $U_0 = \frac{1}{2}$, $U_j = 1$, $j \ge 1$, and $k/h = \frac{1}{2}$. This gives $s_j = 0$, $j \ne 0$, $s_0 = \frac{1}{2}$, and $D = \frac{1}{2}$. Since the flux function f(u) = u is linear, all the linear interpolations used in defining the scheme are exact. Thus, the scheme constructs the exact solution to the piecewise linear Riemann problem by shifting the solution one half cell to the right. Upon averaging we find that $\bar{D} = \bar{U}_1 - \bar{U}_0 = \frac{13}{16} - \frac{3}{16} = \frac{5}{8} > \frac{1}{2} = D$. If f(u) were slightly (but strictly) convex we would still have $\bar{D} > D$. The values U_0 and U_1 decrease by the shaded areas in Fig. 5. Since areas A_1 and A_2 cancel, the value of D will increase by the area A_3 . Note that this area would be nonzero for any slope $s_0 > 0$.

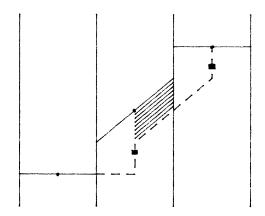


Fig. 5. Violation of the spreading estimate (3.6) in the second order scheme.

Nonetheless, we can gain some theoretical confidence in the method by considering the sonic rarefaction case. In practice this is the most worrisome case; violation of the entropy condition is usually manifested in the form of a sonic shock which fails to spread. Intuitively, we expect that the second-order method, being an extension of Godunov's method, will not permit such behavior. In fact, if we consider data for which

$$U_i^+ < u_0 < U_{i+1}^-$$

so that the sonic point lies within the discontinuity between U_i and U_{i+1} , we find that

(3.19)
$$G(U; j+1) = f(U_{j+1}^+) - \frac{1}{2}ks_{j+1}(g'_{j+1})^2 \ge f(U_{j+1}),$$

$$G(U; j) = f(u_0),$$

$$G(U; j-1) = f(U_i^-) - \frac{1}{2}ks_i(g'_i)^2 \ge f(U_i).$$

The inequalities here follow from the Courant number restriction and the convexity of f. For example, $0 > kg'_i \ge -h$ and by the definition of g'_i ,

$$f(U_j^+) = f(U_j^-) + hs_j g_j'$$

so that

$$f(U_j) \leq \frac{1}{2} (f(U_j^+) + f(U_j^-))$$

= $f(U_j^-) + \frac{1}{2} h s_j g_j'$
 $\leq f(U_j^-) - \frac{1}{2} k s_j (g_j')^2$

and similarly for the other inequality.

From (3.19), we obtain

$$\begin{split} \bar{U}_{j+1} - \bar{U}_j &= U_{j+1} - U_j - \frac{k}{h} [G(U; j+1) - 2G(U; j) + G(U; j-1)] \\ &\leq U_{j+1} - U_j - \frac{k}{h} [f(U_{j+1}) - 2f(u_0) + f(U_j)]. \end{split}$$

Comparing this to (3.16) shows that the second-order method has spreading with at least the same rate as Godunov's method in this situation.

In practice, it seems to be much better than Godunov's method as in Fig. 3.1. The prominent kinks in the rarefaction wave computed with Godunov's method are entirely missing in the second-order calculation.

As a final comment about entropy, we note that by placing additional constraints on the slopes s_i we could obtain the standard entropy inequality for all entropies of the form $\eta(u) = |u - c|$ with c any real constant. Kruzkov [7] has shown that this is sufficient to guarantee uniqueness. We will sketch this only briefly since we do not feel that such a modification is necessary in practice.

For Godunov's method it follows from Jensen's inequality for convex functions that the entropy condition is satisfied. If (η, q) is any convex entropy pair (see, e.g., [9]), then

$$\eta(\bar{U}_j) = \frac{1}{h} \int_{x_{j-1/2}}^{x_{j+1/2}} \eta(\bar{U}_j) \ dx \le \frac{1}{h} \int_{x_{j-1/2}}^{x_{j+1/2}} \eta(w(x, t_{n+1})) \ dx$$

where, as in § 1, w(x, t) is the exact solution to (1.1) on (t_n, t_{n+1}) with piecewise constant initial data U_i . It follows that

(3.20)
$$\eta(\bar{U}_j) - \eta(U_j) < \frac{k}{h} [q(w(x_{j+1/2}, t_{n+1})) - q(w(x_{j-1/2}, t_{n+1}))]$$

which is the discrete form of the entropy inequality.

For the second-order method we must also consider the step going from \bar{U}_j to the new piecewise linear function $\bar{v}(x,t_{n+1})=\bar{U}_j+\bar{s}_j(x-x_j)$ on $(x_{j-1/2},x_{j+1/2})$. For the Kruzkov entropies we would like to show that

(3.21)
$$\int_{x_{i-1/2}}^{x_{j+1/2}} |\bar{v}(x, t_{n+1}) - c| dx \le \int_{x_{i-1/2}}^{x_{j+1/2}} |v(x, t_{n+1}) - c| dx,$$

since we can then show the discrete entropy inequality

$$\bar{E}_j - E_j \le \frac{k}{h} [Q(U; j+1) - Q(U; j)]$$

where

$$E_{j} = \frac{1}{h} \int_{x_{j-1/2}}^{x_{j+1/2}} |v(x, t_{n}) - c| dx,$$

$$Q(U; j) = \frac{1}{k} \int_{t_{n}}^{t_{n+1}} q(v(x_{j-1/2}, t)) dt.$$

Here v(x, t) is the solution to the linearized form of (1.1) with piecewise linear initial data.

However, (3.21) may fail to hold for some values of c if the slopes \bar{s}_j defining \bar{v} are chosen according to (2.7). Since (3.21) does hold if $\bar{s}_j = 0$, it should also hold for \bar{s}_i sufficiently small. In particular, we can show that a sufficient condition for (3.21) is

(3.22)
$$|\bar{s}_j| < \min_{x \in I_j} |v_x(x, t_{n+1})|.$$

Since in smooth regions of the flow we expect $s_j \approx u_x$, this is not much of a restriction and the modified method should still be second-order accurate. On the other hand, the restriction (3.22) would be difficult to impose in practice and does not seem necessary since we have not encountered any difficulties with the unmodified method.

REFERENCES

- [1] P. COLLELA AND P. WOODWARD, The piecewise-parabolic method (PPM) for gas dynamical simulations, J. Comput. Phys., 54 (1984), pp. 174-201.
- [2] C. M. DAFERMOS, Polygonal approximation of solutions of the initial value problem for a conservation law, J. Math. Anal. Appl., 38 (1972), pp. 33-41.
- [3] S. K. GODUNOV, Finite difference methods for numerical computations of discontinuous solutions of equations of fluid dynamics, Mat. Sb., 47 (1959), pp. 271-295. (In Russian.)
- [4] A. HARTEN, High resolution schemes for hyperbolic conservation laws, J. Comput. Phys., 49 (1983), pp. 357-393.
- [5] A. HARTEN, J. M. HYMAN AND P. D. LAX, On finite difference approximation and entropy conditions for shocks, Comm. Pure Appl. Math., 29 (1976), pp. 297-322.
- [6] A. HARTEN, P. D. LAX AND B. VAN LEER, On upstream differencing and Godunov-type schemes for hyperbolic conservation laws, SIAM Rev., 25 (1983), pp. 35-61.
- [7] S. N. KRUZKOV, First order quasi-linear equations with several space variables, Math. USSR-Sb., 10 (1970), pp. 217-243.
- [8] P. D. LAX, The formation and decay of shock waves, Amer. Math. Monthly, 79 (1972), pp. 227-241.
- [9] ——, Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves, CBMS-NSF Regional Conference Series in Applied Mathematics 11, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1973.
- [10] B. VAN LEER, Towards the ultimate conservative difference scheme, III, J. Comput. Phys., 23 (1977), pp. 263-275.
- [11] ——, Towards the ultimate conservative difference scheme: V. A second-order sequel to Godunov's method, J. Comput. Phys., 32 (1979), pp. 101-136.
- [12] O. A. OLEINIK, Construction of a generalized solution of the Cauchy problem for a quasi-linear equation of first order by the introduction of "vanishing viscosity," Amer. Math. Soc. Transl., Series 2, 33 (1963), pp. 277-283.
- [13] S. OSHER, Convergence of generalized MUSCL schemes, NASA Contractor Report 172306, ICASE, NASA Langley Research Center, Hampton, VA, 1984.
- [14] ——, Riemann solvers, the entropy condition and difference approximations, this Journal, 21 (1984), pp. 217-235.
- [15] S. OSHER AND S. R. CHAKRAVARTHY, High resolution schemes and the entropy condition, this Journal, 21 (1984), pp. 955-984.
- [16] P. L. Roe, Approximate Riemann solvers, parameter vectors, and difference schemes, J. Comput. Phys., 43 (1981), pp. 357-372.
- [17] ——, Some contributions to the modeling of discontinuous flows, Proc. AMS-SIAM Summer Seminar on Computational Fluid Dynamics, La Jolla, CA, 1983.
- [18] P. K. Sweby, High resolution schemes using flux limiters for hyperbolic conservation laws, this Journal, 21 (1984), pp. 995-1011.
- [19] E. TADMOR, The large-time behavior of the scalar, genuinely nonlinear Lax-Friedrichs scheme, NASA Contractor Report 172138, ICASE, NASA Langley Research Center, Hampton, VA, 1983.
- [20] ——, Numerical viscosity and the entropy condition for conservative difference schemes, NASA Contractor Report 172141, ICASE, NASA Langley Research Center, Hampton, VA, 1983.