MATHEMATICS OF COMPUTATION
VOLUME 51, NUMBER 184
OCTOBER 1988, PAGES 535-558

A Third-Order Accurate Variation Nonexpansive Difference Scheme for Single Nonlinear Conservation Laws

By Richard Sanders*

Abstract. It was widely believed that all variation nonexpansive finite difference schemes for single conservation laws must reduce to first-order at extreme points of the approximation. It is shown here that this belief is in fact false. A third-order scheme, which at worst may reduce to second order at extreme points, is developed and analyzed. Moreover, extensive numerical experiments indicate that the third-order scheme introduced here yields superior approximations when compared with other variation nonexpansive difference schemes.

1. Introduction. In this paper we introduce and analyze a formally third-order accurate finite difference scheme used to approximate solutions to the single hyperbolic conservation law:

$$\frac{\partial}{\partial t}u + \frac{\partial}{\partial x}f(u) = 0, \qquad u(x,0) = u_0(x). \tag{1.1}$$

Future work will be devoted to extending the techniques presented here to hyperbolic systems in one and many space dimensions.

In recent years it has been found that incorporating the property that the variation of the exact solution to (1.1) does not increase in time into the design of approximating schemes often leads to superior numerical results when compared to approximations coming from methods where this property is ignored. However, it was widely believed that total variation nonexpansive schemes (or loosely speaking, total variation diminishing, or TVD schemes) must automatically reduce to first-order accuracy near extreme points of the solution; see [8], [10]. In this paper we show that TVD schemes need not be incompatible with high-order accuracy. We introduce a high-order technique that yields approximate solutions to (1.1) which have nonexpanding variation in time and which also satisfy the same maximum principle given by the exact solution to (1.1). Moreover, away from extreme points of the solution, we show that our method is formally third-order accurate, and around extreme points it can reduce to no less than second-order accuracy.

This paper is divided into four sections. In Section 2 we develop an approximation procedure for functions of bounded variation. Our approximation procedure cannot increase the variation of the function being approximated, and it is third-order accurate in regions where the approximated function is smooth. In Section 3 we show how to evolve this approximation in time in a way that is variation nonexpansive, consistent with the weak form of (1.1), and we prove that the resulting

Received March 2, 1987.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 65M10; Secondary 65M05.
*Research supported by NSF Grant No. DMS85-05422.

scheme is formally third-order accurate in space and time. Finally, in Section 4 we present some numerical examples.

2. Variation Nonexpansive Third-Order Accurate Approximation. The goal of this section is to develop an approximation procedure that has the following desirable properties. First, we require that our procedure yield an approximation that retains the cell average of the function being approximated. Second, we require that our procedure contract the total variation as well as local extrema of the function being approximated. Finally, when our procedure is applied to a sufficiently smooth function, we require that the resulting approximation have third-order accuracy with respect to the mesh size. As is seen in the next section, these factors are fundamental to the development of our stable and high-order accurate finite difference method.

To begin, we give a few preliminary definitions. Partition the real line into nonoverlapping intervals, $\mathbf{R} = \bigcup_j I_j$, where I_j denotes the semiclosed interval $I_j = [x_j, x_{j+1})$. Let $\bar{I}_j$ denote the closure of I_j and let $I_j^0 = (x_j, x_{j+1})$. Let Δx represent the length of I_j, $\Delta x = x_{j+1} - x_j$, and denote the midpoint of I_j by $x_{j+1/2}$. Finally, let $\chi_j(x)$ represent the characteristic function of the *open* interval I_j^0, i.e.,

$$\chi_j(x) = \begin{cases} 1, & x \in I_j^0, \\ 0, & \text{otherwise}, \end{cases}$$

and let $\delta(x)$ be given by

$$\delta(x) = \begin{cases} 1, & x = 0, \\ 0, & \text{otherwise}. \end{cases}$$

Now consider an interpolation taking the form

$$\mathscr{R}^{\Delta}(u)(x) = \sum_j [P_j(x - x_{j+1/2})\chi_j(x) + u(x_j)\delta(x - x_j)], \tag{2.1}$$

where each P_j is a polynomial. To obtain the desired third-order accurate approximation, the degree of P_j must be at least two. For organizational reasons, we list here all of the properties that $\mathscr{R}^{\Delta}(u)$ is required to satisfy. For any $u \in BV$ we require that

$$\int_{I_j} \mathscr{R}^{\Delta}(u)(x)\,dx = \int_{I_j} u(x)\,dx, \tag{2.2a}$$

$$\operatorname{Var}(\mathscr{R}^{\Delta}(u)) \le \operatorname{Var}(u), \tag{2.2b}$$

$$\sup(\mathscr{R}^{\Delta}(u)) \le \sup(u), \qquad \inf(\mathscr{R}^{\Delta}(u)) \ge \inf(u), \tag{2.2c}$$

and in the event that $u \in C^3$ we require that

$$\mathscr{R}^{\Delta}(u)(x) = u(x) + O(\Delta x^3). \tag{2.2d}$$

To produce a first candidate for P_j, which below is denoted by P_j^1, we for the moment forego considering properties (2.2b) and (2.2c) and concentrate on first satisfying (2.2a) and (2.2d). Once this is accomplished, the complete problem is treated by suitably modifying P_j^1 in such a way that the resulting scheme yields an approximation that satisfies all of the properties above.

Let $U(x)$ denote the antiderivative of $u(x)$; $U(x) = \int_0^x u(s)\,ds$. It is well known (see [4], for example) that the cubic Hermite interpolation to U on the interval $\bar{I}_j$ is given by

$$H(I_j, U)(x) = \sum_{k=0}^{3} \alpha_k^j(U) h_k^j(x),$$

where

$$h_0^j(x) = 1, \qquad h_1^j(x) = (x - x_j),$$
$$h_2^j(x) = (x - x_j)^2, \qquad h_3^j(x) = (x - x_j)^2(x - x_{j+1}),$$

and where

$$\alpha_0^j(U) = U(x_j), \qquad \alpha_1^j(U) = U'(x_j),$$
$$\alpha_2^j(U) = [\Delta^+ U(x_j) - \Delta x U'(x_j)]/\Delta x^2,$$
$$\alpha_3^j(U) = [\Delta x(U'(x_j) + U'(x_{j+1})) - 2\Delta^+ U(x_j)]/\Delta x^3.$$

Note that Δ^+ represents the forward difference operator, $\Delta^+ U(x_j) = U(x_{j+1}) - U(x_j)$. It is furthermore well known that when $U \in C^4$ the formula with remainder,

$$U(x) = H(I_j, U)(x) + \frac{1}{4!} U^{IV}(\xi(x))(x - x_j)^2(x - x_{j+1})^2,$$

is valid for some $\xi(x) \in I_j^0$. Differentiating this formula, we arrive at

$$\begin{aligned} u(x) = {} & \frac{d}{dx} H(I_j, U)(x) \\ & + \frac{1}{3!} u'''(\xi(x))(x - x_{j+1/2}) \left((x - x_{j+1/2})^2 - \left(\frac{\Delta x}{2} \right)^2 \right) \\ & + \kappa(x)(x - x_j)^2(x - x_{j+1})^2. \end{aligned} \tag{2.3}$$

Recalling that $u(x) = U'(x)$, we easily conclude that

$$\frac{d}{dx} H(I_j, U)(x) = [\mathscr{A} - (\widehat{u_R} + \widehat{u_L})/4] + [\widehat{u_R} - \widehat{u_L}]\theta + 3[\widehat{u_R} + \widehat{u_L}]\theta^2, \tag{2.4}$$

where above and throughout we let $\mathscr{A}$, $\widehat{u_L}$, and $\widehat{u_R}$ and θ denote the normalized variables

$$\begin{aligned} \mathscr{A} &= \frac{1}{\Delta x} \int_{I_j} u(s)\,ds, \\ \widehat{u_L} &= u(x_j) - \mathscr{A}, \\ \widehat{u_R} &= u(x_{j+1}) - \mathscr{A}, \\ \theta &= (x - x_{j+1/2})/\Delta x. \end{aligned} \tag{2.5}$$

The subscripts on the left-hand side above have been omitted for ease of presentation. At this point we are led to take as our first candidate for P_j the quadratic polynomial

$$P_j^1(x - x_{j+1/2}) = \frac{d}{dx} H(I_j, U)(x), \tag{2.6}$$

which by construction satisfies both (2.2a) and (2.2d) when restricted to the interval I_j.

Next, we give an example to demonstrate that the interpolation formula given by (2.4) does not in general yield an approximation that satisfies either (2.2b) or

(2.2c). Consider interpolating the function $u(x) = x^3$ on the cell $[0, 1]$ using formula (2.4) above. Doing the calculations we arrive at

$$P_0^1 = \frac{3}{2}\left(x - \frac{1}{2}\right)^2 + \left(x - \frac{1}{2}\right) + \frac{1}{8}.$$

However, the value of P_0^1 at $x = 1/6$ is $-1/24$. This demonstrates that P_0^1 violates both (2.2b) and (2.2c) locally. Therefore, one can modify u outside $[0, 1]$ to obtain an example where the interpolation procedure implied by (2.4) violates (2.2b) and (2.2c) globally. This "overshoot" phenomenon can however be rectified. Overcoming this problem, which is inherent to cell average preserving piecewise quadratic interpolation, is the subject of the remainder of this section.

Before continuing, we introduce some further notations. Let a quadratic $\hat{P}(a, b)(\theta)$ be defined by

$$\hat{P}(a, b)(\theta) = 3(a + b)\theta^2 + (b - a)\theta - (a + b)/4, \tag{2.7}$$

and note that this quadratic satisfies

$$\hat{P}(a, b)\left(-\tfrac{1}{2}\right) = a, \quad \hat{P}(a, b)\left(\tfrac{1}{2}\right) = b, \quad \int_{-1/2}^{1/2} \hat{P}(a, b)(\theta)\, d\theta = 0.$$

Recalling the definitions of (2.5), set

$$M = \max\mathrm{mod}(\widehat{u_L}, \widehat{u_R}), \qquad m = \min\mathrm{mod}(\widehat{u_L}, \widehat{u_R}),$$

where here the max mod and min mod functions are defined by

$$\max\mathrm{mod}(a, b) = \begin{cases} a & \text{if } |a| \geq |b|, \\ b & \text{if } |a| < |b|, \end{cases}$$

$$\min\mathrm{mod}(a, b) = \begin{cases} b & \text{if } |a| \geq |b|, \\ a & \text{if } |a| < |b|. \end{cases}$$

Furthermore, let ρ represent the ratio

$$\rho = m/M,$$

and observe that $|\rho| \leq 1$. Finally, let E represent the relevant extreme value of $u - \mathscr{A}$ on $\bar{I}$ (relevant in a sense made apparent below), which we define here by

$$E = \begin{cases} \sup_{\bar{I}}(u - \mathscr{A}) & \text{if } M < 0, \\ \inf_{\bar{I}}(u - \mathscr{A}) & \text{if } M \geq 0, \end{cases}$$

and when $M \neq 0$ let $\hat{E}$ be given by

$$\hat{E} = E/M.$$

Our approach to eliminate any possible overshoot or undershoot, as was exemplified by the example above, is to insert

$$\hat{P}(\tau_L \widehat{u_L}, \tau_R \widehat{u_R})(\theta) + \mathscr{A}, \tag{2.8}$$

for $P_j(x - x_{j+1/2})$ in the formula for $\mathscr{R}^{\Delta}(u)(x)$, using certain to be determined values of $0 \leq \tau_L, \tau_R \leq 1$ (recall that we have omitted subscripts for convenience). The decision to insert a quadratic of the form (2.8) into (2.1) is quite natural when one observes that first when $\tau_L = \tau_R = 1$ we have that

$$\hat{P}(\widehat{u_L}, \widehat{u_R})(\theta) + \mathscr{A} = P_j^1(x - x_{j+1/2}),$$

and second for *any* τ_L and τ_R we also have

$$\int_I [\hat{P}(\tau_L \widehat{u_L}, \tau_R \widehat{u_R})(\theta) + \mathscr{A}]\, dx = \int_I u(x)\, dx.$$

Moreover, since $\hat{P}(a,b)(\theta)$ is affine in a, b, we have that

$$\begin{aligned} &|P_j^1(x - x_{j+1/2}) - (\hat{P}(\tau_L \widehat{u_L}, \tau_R \widehat{u_R})(\theta) + \mathscr{A})| \\ &\quad = |\hat{P}((1-\tau_L)\widehat{u_L}, (1-\tau_R)\widehat{u_R})(\theta)|, \end{aligned}$$

which for $|\theta| \le \frac{1}{2}$ is bounded above by

$$\max((1-\tau_L)|\widehat{u_L}|, (1-\tau_R)|\widehat{u_R}|).$$

Therefore, inserting (2.8) into our interpolation formula $\mathscr{R}^{\Delta}(u)$ in place of $P_j^1(x - x_{j+1/2})$ will also yield a third-order accurate approximation provided that τ_L and τ_R are constructed so as to satisfy

$$(1-\tau_L)|\widehat{u_L}| = O(\Delta x^3), \qquad (1-\tau_R)|\widehat{u_R}| = O(\Delta x^3) \tag{2.9}$$

(in the event that $u(x)$ is smooth of course).

At this point it should be clear to the reader that if $P_j^1(x - x_{j+1/2})$ is monotone on I, it would satisfy the local estimates

$$\operatorname{Var}(P_j^1(x - x_{j+1/2}))|_{\bar{I}} \le \operatorname{Var}(u(x))|_{\bar{I}} \tag{2.10}$$

and

$$\begin{aligned} \sup_{\bar{I}}(P_j^1(x - x_{j+1/2})) &\le \sup_{\bar{I}}(u(x)), \\ \inf_{\bar{I}}(P_j^1(x - x_{j+1/2})) &\ge \inf_{\bar{I}}(u(x)). \end{aligned} \tag{2.11}$$

The goal now is to derive a recipe for determining τ_L and τ_R so that

$$\hat{P}(\tau_L \widehat{u_L}, \tau_R \widehat{u_R})(\theta) + \mathscr{A}$$

satisfies the estimates above on every cell I_j, regardless of whether $P_j^1(x - x_{j+1/2})$ is monotone on I_j or not. Then we need to verify that this recipe at the same time implies the error estimate (2.9).

Remark. When $u(x)$ is smooth we see upon differentiating (2.3) that $P_j^1(x - x_{j+1/2})$ is monotone on I unless for some x in I we have

$$u'(x) = \frac{1}{2}u'''(x)\left(\theta^2 - \frac{1}{12}\right)\Delta x^2 + O(\Delta x^3).$$

Therefore, it is necessary to modify $P_j^1(x - x_{j+1/2})$ only near certain (in a sense) nongeneric points. Loosely speaking, these points can be thought of as extreme points (or approximate extreme points) of u.

A straightforward calculation shows that the absolute value of the critical point of $P_j^1(x - x_{j+1/2})$, in the normalized variables defined above, is

$$|\theta_{\text{crit}}| = \frac{1-\rho}{6(1+\rho)}.$$

Therefore, $P_j^1(x - x_{j+1/2})$ fails to be monotone on I only when

$$\frac{1-\rho}{6(1+\rho)} < \frac{1}{2},$$

or equivalently when

$$-\tfrac{1}{2} < \rho \leq 1. \tag{2.12}$$

Again we calculate in the normalized variables to find that the critical value of $P_j^1(x - x_{j+1/2})$ is given by

$$P_{\text{crit}} = -M\frac{(1+\rho+\rho^2)}{3(1+\rho)} + \mathscr{A}.$$

Using the fact that $P_j^1(x - x_{j+1/2})$ is convex (resp. concave) when $M > 0$ (resp. $M < 0$), one easily determines that $P_j^1(x - x_{j+1/2})$ satisfies the local estimates (2.10) and (2.11) when, for $|\theta| \leq \frac{1}{2}$, it takes its values in the range of $M + \mathscr{A}$, $m + \mathscr{A}$ and $E + \mathscr{A}$. In the nontrivial case, $|\theta_{\text{crit}}| < \frac{1}{2}$, one easily checks that this fact is implied by the inequality

$$\frac{1+\rho+\rho^2}{3(1+\rho)} \leq -\hat{E} \tag{2.13}$$

(note that $\hat{E} \leq 0$). Therefore, even in the case when $P_j^1(x - x_{j+1/2})$ has its critical point inside the interval I_j, when inequality (2.13) is satisfied we are assured that the local estimates (2.10) and (2.11) are satisfied by $P_j^1(x - x_{j+1/2})$ (or equivalently by (2.8) with $\tau_L = \tau_R = 1$).

The only situation that remains to be considered is (2.12) satisfied, and (2.13) violated. To obtain the desired stability and accuracy in this situation, some care must be exercised choosing τ_L and τ_R. This task is divided into two separate cases.

Case 1. (2.12) is satisfied, (2.13) violated and $\rho \geq 0$.

We begin with a simple lemma given without proof.

LEMMA 2.1. *Let* $|\theta_c| < \frac{1}{2}$ *denote the critical point of* $\hat{P}(\widehat{u_L}, \widehat{u_R})(\theta) + \mathscr{A}$. *We then have that*

$$-\frac{1}{M}\hat{P}(\widehat{u_L}, \widehat{u_R})(\theta_c) \geq \frac{(1-\rho)^2 + 3(1+\rho)^2}{12(1+\rho)}.$$

Now define τ_+ by

$$\tau_+ = -\hat{E} \cdot \frac{3(1+\rho)}{1+\rho+\rho^2},$$

and observe that in this case $0 \leq \tau_+ < 1$. It is straightforward to check that if we set

$$\tau_L = \tau_+, \qquad \tau_R = \tau_+ \tag{2.14}$$

in this case, the critical point of (2.8) agrees with the critical point of $P_j^1(x - x_{j+1/2})$. More importantly, however, choosing τ_L and τ_R as prescribed above forces the critical value of (2.8) to become exactly $E + \mathscr{A}$, the relevant extreme value of u. These observations allow us to again conclude without proof:

LEMMA 2.2. *Choosing* τ_L *and* τ_R *as above, we have in the present case that the function*

$$P(x - x_{j+1/2}) = \begin{cases} u(x_j), & x = x_j, \\ \hat{P}(\tau_L\widehat{u_L}, \tau_R\widehat{u_R})(\theta) + \mathscr{A}, & x \in I^0, \\ u(x_{j+1}), & x = x_{j+1}, \end{cases}$$

defined on $\bar{I}$, *satisfies the local estimates* (2.10) *and* (2.11).

Finally we verify that τ_L and τ_R as given by (2.14) satisfy the error estimate (2.9).

LEMMA 2.3. *Choosing τ_L and τ_R as above, we have in the present case that*

$$(1-\tau_L)|\widehat{u_L}| = O(\Delta x^3), \qquad (1-\tau_R)|\widehat{u_R}| = O(\Delta x^3)$$

whenever $u(x) \in C^3$.

Proof. First observe that formula (2.3) allows us to conclude that there is a bounded function, say $\eta(x)$, such that on I

$$\hat{P}(\widehat{u_L}, \widehat{u_R})(\theta) + \mathscr{A} = u(x) + \eta(x)\left(\frac{1}{4} - \theta^2\right)\Delta x^3.$$

From this we get that

$$\begin{aligned}\hat{P}(\tau_+\widehat{u_L}, \tau_+\widehat{u_R})(\theta) + \mathscr{A} \\ = u(x) + \eta(x)\left(\frac{1}{4} - \theta^2\right)\Delta x^3 + (1-\tau_+)M\left[-\frac{1}{M}\hat{P}(\widehat{u_L}, \widehat{u_R})(\theta)\right].\end{aligned}$$

At the critical point of $\hat{P}(\widehat{u_L}, \widehat{u_R})(\theta)$, say θ_c, we see that taking $\tau_L = \tau_R = \tau_+$, as we have done in this case, implies the inequality

$$\operatorname{sgn}(M)[\hat{P}(\tau_+\widehat{u_L}, \tau_+\widehat{u_R})(\theta_c) + \mathscr{A}] \le \operatorname{sgn}(M)u(\theta_c\Delta x + x_{j+1/2}),$$

and this implies that

$$(1-\tau_+)|M|\left[-\frac{1}{M}\hat{P}(\widehat{u_L}, \widehat{u_R})(\theta_c)\right] \le \left|\eta(\theta_c\Delta x + x_{j+1/2})\left(\frac{1}{4} - \theta_c^2\right)\right|\Delta x^3.$$

The result of Lemma 2.1 now makes the desired result obvious.

Case 2. (2.12) is satisfied, (2.13) violated and $\rho < 0$.

Rather than modifying both $\widehat{u_L}$ and $\widehat{u_R}$ as was done in the case above, here we modify only the end point value with maximum modulus. Consider the following. For $\rho < 0$ define τ_- by

$$\tau_- = -\frac{1}{2}[(\rho + 3\hat{E}) - (3(\hat{E} - \rho)(3\hat{E} + \rho))^{1/2}]$$

and observe that when $\rho < 0$ we have that $\hat{E} \le \rho$, thus showing that τ_- is real. Now define τ_L and τ_R by

$$\begin{aligned}\tau_L &= \begin{cases} 1 & \text{if } |\widehat{u_L}| < |\widehat{u_R}|, \\ \tau_- & \text{if } |\widehat{u_L}| \ge |\widehat{u_R}|, \end{cases} \\ \tau_R &= \begin{cases} \tau_- & \text{if } |\widehat{u_L}| < |\widehat{u_R}|, \\ 1 & \text{if } |\widehat{u_L}| \ge |\widehat{u_R}|. \end{cases}\end{aligned} \tag{2.15}$$

One easily computes that this particular choice of τ_L and τ_R gives

$$\theta_c = \pm\frac{1}{6}\left(\frac{\tau_- - \rho}{\tau_- + \rho}\right) \tag{2.16}$$

as the critical point of (2.8); the sign above depends on which normalized end point, $\widehat{u_L}$ or $\widehat{u_R}$, has the maximum modulus. Choosing τ_L and τ_R in this way also forces the critical value of (2.8) to be $E + \mathscr{A}$. Moreover, modifying only the end point with maximum modulus in this case does not introduce any new oscillations, as

could be the case if we modified here the end point with minimum modulus. Thus the function

$$P(x - x_{j+1/2}) = \begin{cases} u(x_j), & x = x_j, \\ \hat{P}(\tau_L \widehat{u_L}, \tau_R \widehat{u_R})(\theta) + \mathscr{A}, & x \in I^0, \\ u(x_{j+1}), & x = x_{j+1} \end{cases}$$

with τ_L and τ_R given by (2.15) has, in this case, only one local extremum in $\bar{I}$, and its extreme value is $E + \mathscr{A}$.

LEMMA 2.4. *In the present case, τ_- satisfies the inequalities $-2\rho \le \tau_- \le 1$.*

Proof. Set

$$\tau_-(s) = -\frac{1}{2}[(\rho + s) - (3(s - \rho)(3s + \rho))^{1/2}]$$

and compute that

$$\frac{d}{ds}\tau_-(s) \le 0 \quad \text{for } \rho \ge s \ge s_* = \frac{(1 + \rho + \rho^2)}{3(1 + \rho)}.$$

Since $\tau_-(\rho) = -2\rho$ and $\tau_-(s_*) = 1$ the final result is easily seen.

These observations above combine to give

LEMMA 2.5. *In the present case, choosing τ_L and τ_R as in* (2.15) *implies that the function $P(x - x_{j+1/2})$ satisfies the local estimates* (2.10) *and* (2.11).

Again we verify that τ_L and τ_R given by (2.15) satisfy the error estimate (2.9).

LEMMA 2.6. *Choosing τ_L and τ_R as above, we have in the present case that*

$$(1 - \tau_L)|\widehat{u_L}| = O(\Delta x^3), \qquad (1 - \tau_R)|\widehat{u_R}| = O(\Delta x^3)$$

whenever $u(x) \in C^3$.

Proof. We assume that $\widehat{u_R} = M$ since the proof is similar otherwise. Following the proof of Lemma 2.3, we arrive at the inequality

$$(1 - \tau_R)|\widehat{u_R}|[-\hat{P}(0,1)(\theta_c)] \le \left|\eta(\theta_c \Delta x + x_{j+1/2})\left(\frac{1}{4} - \theta_c^2\right)\right| \Delta x^3.$$

This inequality and some simplification gives us that

$$(1 - \tau_R)|\widehat{u_R}| \le \text{const}\left[\frac{2}{6\theta_c + 1}\right] \Delta x^3.$$

Using the result of Lemma 2.4 together with (2.16) implies that $\theta_c \ge 1/6$, and this inserted into the inequality above completes the proof of the lemma.

We conclude this section by condensing its main results into a theorem.

THEOREM 2.1. *In the interpolation formula* (2.1), *define $P_j(x - x_{j+1/2})$ by*

$$P_j(x - x_{j+1/2}) = \hat{P}(\tau_L \widehat{u_L}, \tau_R \widehat{u_R})(\theta) + \mathscr{A},$$

where τ_L and τ_R are given by the following recipe:

If $-1 \le \rho \le -1/2$, then

$$\tau_L = 1, \qquad \tau_R = 1.$$

If $-1/2 < \rho < 0$ and $\widehat{u_R} = M$, then

$$\tau_L = 1, \qquad \tau_R = \min(\tau_-, 1).$$

If $-1/2 < \rho < 0$ *and* $\widehat{u_L} = M$, *then*

$$\tau_L = \min(\tau_-, 1), \qquad \tau_R = 1.$$

If $0 \le \rho \le 1$, *then*

$$\tau_L = \min(\tau_+, 1), \qquad \tau_R = \min(\tau_+, 1).$$

Then the interpolation formula (2.1) *yields an approximation* $\mathscr{R}^\Delta(u)$ *that satisfies all properties* (2.2a) *through* (2.2d).

3.1. Evolution of the Approximation. In this section we develop a simple and accurate method to evolve a piecewise parabolic approximation to the solution of (1.1) from one time level to the next. This evolution scheme together with the reconstruction algorithm of Section 2 combine to yield a generically third-order accurate finite difference scheme. Moreover, the resulting approximation has nonincreasing variation in time and satisfies the same maximum principle as satisfied by the exact solution to (1.1).

To implement the approximation of Section 2, three basic pieces of information must be available for each cell. Specifically, the cell average and the left and right cell boundary point value of the function being approximated must be known. To obtain this information, we employ a staggered spatial mesh together with the method of characteristics. To simplify the presentation, we discuss our procedure for one time iteration only and note that succeeding time iterations follow in an analogous manner.

Let $T(u_0)(x,t)$, $t \ge 0$, denote the solution to the differential equation

$$\frac{\partial}{\partial t}u + \frac{\partial}{\partial x}f(u) = 0, \qquad u(x,0) = u_0(x). \tag{3.1}$$

As in the previous section, partition the real line into nonoverlapping intervals $\bigcup_j I_j$, with $I_j = [x_j, x_{j+1})$, and approximate $u_0 \in BV$ by

$$v^0(x) = \mathscr{R}^\Delta(u_0)(x). \tag{3.2}$$

Consider the staggered partition $\bigcup_j I_{j-1/2}$, $I_{j-1/2} = [x_{j-1/2}, x_{j+1/2})$, along with its associated staggered reconstruction algorithm $\mathscr{R}^S$. The object of our evolution technique is to determine a piecewise parabolic approximation of $T(v^0)(x, \Delta t/2)$. This goal is accomplished by computing a slight perturbation of $\mathscr{R}^S(T(v^0)(\cdot, \Delta t/2))$.

To avoid abusing notation too much, we assume that $v^0(x)$ is given by

$$v^0(x) = \bar{\mathscr{R}}^\Delta(u_0)(x), \tag{3.3}$$

where $\overline{\mathscr{R}}^\Delta$ (and $\overline{\mathscr{R}}^S$ as well) denotes a "preconditioned" version of $\mathscr{R}^\Delta$. By preconditioned we mean specifically that the data which $\mathscr{R}^\Delta$ is applied to is modified (in a way we discuss in detail at the end of this section) so that for all u in the range of u_0 and all j,

$$|f''(u)|\Delta t \max_{\bar{I}_j} \left| \frac{d}{dx} P_j(x - x_{j+1/2}) \right| < 2. \tag{3.4a}$$

Written in terms of the normalized variables (2.5), (3.4a) becomes

$$|f''(u)|\lambda \max_{|\theta| \le 1/2} \left| \frac{d}{d\theta} \hat{P}(\tau_L \widehat{u_L}, \tau_R \widehat{u_R})(\theta) \right| < 2, \tag{3.4b}$$

where $\lambda = \Delta t/\Delta x$. Essentially what we do below to enforce condition (3.4) is to push extremely large gradients of u out of the quadratic P_j and into jump discontinuities at cell interfaces. A condition much like (3.4) is made in [10]. An additional condition that we assume throughout is the Courant condition; that is, for all u in the range of u_0 we assume the ratio λ is taken so that

$$|f'(u)|\lambda < 1. \tag{3.5}$$

With (3.4) and (3.5) we have:

LEMMA 3.1. *The characteristic equation*

$$\frac{d}{dt}x(s,t) = f'(v^0(s)), \qquad x(s,0) = s \in I^0_{j-1},$$

has a unique solution $x(s_{j-1},t)$ *such that*

$$x(s_{j-1},\Delta t/2) = x_{j-1/2}, \qquad s_{j-1} \in I^0_{j-1}.$$

Proof. Consider finding the root of the function

$$H(s) = f'(v^0(s))\Delta t + 2(s - x_{j-1/2}).$$

According to (3.5) we have that

$$H(x_{j-1}+0) = f'(v^0(x_{j-1}+0))\Delta t - \Delta x < 0,$$
$$H(x_j - 0) = f'(v^0(x_j - 0))\Delta t + \Delta x > 0,$$

and (3.4) implies that for every $s \in I^0_{j-1}$

$$\frac{d}{ds}H(s) = f''(v^0(s))\Delta t\,\frac{d}{ds}P_{j-1}(s - s_{j-1/2}) + 2 > 0.$$

Therefore, $H(s)$ has a unique root in I^0_{j-1}, call it s_{j-1}, and

$$x(s_{j-1},t) = s_{j-1} + f'(v^0(s_{j-1}))t$$

defines the desired characteristic.

The result of the previous lemma guarantees that through each point $(x_{j-1/2},\Delta t/2)$ there is a unique backward characteristic of (3.1) (with $u(x,0) = \bar{\mathscr{R}}^{\Delta}(u_0)(x)$), intersecting the line $t = 0$ within cell I_{j-1}. The divergence theorem applied to (3.1) over the trapezoid defined by the points $(x_{j-1/2},\Delta t/2)$, $(x_{j+1/2},\Delta t/2)$, $(s_{j-1},0)$ and $(s_j,0)$ gives the formula

$$\begin{aligned}\frac{1}{\Delta x}\int_{I_{j-1/2}} & T(v^0)(x,\Delta t/2)\,dx \\ &= \frac{1}{\Delta x}\int_{s_{j-1}}^{s_j} v^0(x)\,dx - \frac{\lambda}{2}\Delta^+\left(f(v^0(s_{j-1})) - f'(v^0(s_{j-1}))v^0(s_{j-1})\right).\end{aligned} \tag{3.6}$$

Therefore the average of $T(v^0)(x,\Delta t/2)$ on every cell $I_{j-1/2}$ is given exactly by the right-hand side of (3.6). Moreover, the cell endpoint values of $T(v^0)(x,\Delta t/2)$ are given exactly by

$$T(v^0)(x_{j-1/2},\Delta t/2) = v^0(s_{j-1}), \qquad T(v^0)(x_{j+1/2},\Delta t/2) = v^0(s_j). \tag{3.7}$$

For convenience we adopt the following notation. Let

$$\begin{aligned} &u_1 = v^0(s_{j-1}), \quad u_2 = v^0((s_{j-1}+x_j)/2), \quad u_3 = v^0(x_j - 0), \\ &u_4 = v^0(x_j + 0), \quad u_5 = v^0((x_j + s_j)/2), \quad u_6 = v^0(s_j), \end{aligned} \tag{3.8}$$

and rewrite the right-hand side of (3.6) with the aid of Simpson's rule as

$$
\begin{aligned}
\mathscr{A}_{j-1/2} &= F(u_1, u_2, u_3, u_4, u_5, u_6) \\
&= \frac{1}{12}(1 - f'(u_6)\lambda)(u_6 + 4u_5 + u_4) \\
&\quad + \frac{1}{12}(1 + f'(u_1)\lambda)(u_3 + 4u_2 + u_1) \\
&\quad - \frac{\lambda}{2}[(f(u_6) - f'(u_6)u_6) - (f(u_1) - f'(u_1)u_1)].
\end{aligned} \tag{3.9}
$$

We pause now to give a direct proof of a result concerning the monotone and conservative operator $F(\cdots)$.

LEMMA 3.2. *Let*

$$M = \max_{[s_{j-1}, s_j]}(v^0(\cdot)), \qquad m = \min_{[s_{j-1}, s_j]}(v^0(\cdot)),$$

and suppose that conditions (3.4) *and* (3.5) *are satisfied. Then, when* $\mathscr{A}_{j-1/2}$ *is given by* (3.9), *we have that*

$$m \le \mathscr{A}_{j-1/2} \le M.$$

We prove the lemma using a more restrictive version of (3.4). That is, we replace 2 on the right-hand side of (3.4) with 2/3. The interested reader can easily relax this assumption. We choose the route of simplicity here over generality, for ease of presentation.

Proof. First one easily checks that

$$F(m, m, m, m, m, m) = m, \qquad F(M, M, M, M, M, M) = M.$$

Condition (3.5) gives us that F is an increasing function in arguments two through five. Moreover, a simple calculation gives us that

$$
\begin{aligned}
\frac{\partial}{\partial u_1}F(\cdots) &= \frac{1}{12}[(1 + f'(u_1)\lambda) + f''(u_1)\lambda(u_3 + 4u_2 - 5u_1)], \\
\frac{\partial}{\partial u_6}F(\cdots) &= \frac{1}{12}[(1 - f'(u_6)\lambda) - f''(u_6)\lambda(u_4 + 4u_5 - 5u_6)],
\end{aligned}
$$

and another simple calculation gives us that

$$-5(u_1 - u_3) + 4(u_2 - u_3) = \frac{3}{2}\Delta x(1 + f'(u_1)\lambda)\frac{d}{dx}P_{j-1}(x^* - x_{j-1/2}),$$

with $x^* = x_j + \frac{2}{3}(s_{j-1} - x_j)$, and

$$-5(u_6 - u_4) + 4(u_5 - u_4) = -\frac{3}{2}\Delta x(1 - f'(u_6)\lambda)\frac{d}{dx}P_j(x' - x_{j+1/2}),$$

with $x' = x_j + \frac{2}{3}(s_j - x_j)$. Putting these identities together we find that

$$
\begin{aligned}
\frac{\partial}{\partial u_1}F(\cdots) &\ge \left(\frac{1 + f'(u_1)\lambda}{12}\right)\left(1 - \frac{3}{2}\left|f''(u)\Delta t\frac{d}{dx}P_{j-1}\right|\right), \\
\frac{\partial}{\partial u_6}F(\cdots) &\ge \left(\frac{1 - f'(u_6)\lambda}{12}\right)\left(1 - \frac{3}{2}\left|f''(u)\Delta t\frac{d}{dx}P_j\right|\right).
\end{aligned}
$$

Modified condition (3.4) now implies that F is also increasing in argument one and six. Arguments u_1 through u_6 can now be perturbed to m, as well as M, in such a way as to maintain condition (3.4), therefore establishing the desired result.

Next we show that the reconstruction-evolution algorithm developed above is third-order accurate in regions where the smooth initial datum u_0 satisfies

$$|u_0'(x)| \geq \frac{1}{24}|u_0'''(x)|\Delta x^2 + O(\Delta x^3), \tag{3.10}$$

and in regions where u_0 violates (3.10), loosely speaking near local extrema, we lose at most one order of accuracy. In the present context, what we mean by an rth-order accurate scheme is that if we define $\bar{u}_{j-1/2}$ to be the $j-1/2$ cell average of the exact solution to (3.1) at $t = \Delta t/2$, starting with smooth datum u_0, that is,

$$\bar{u}_{j-1/2} = \frac{1}{\Delta x}\int_{I_{j-1/2}} T(u_0)(x, \Delta t/2)\, dx,$$

then the rth-order scheme gives a $j-1/2$ cell average, say $\mathscr{A}_{j-1/2}$, which satisfies

$$|\mathscr{A}_{j-1/2} - \bar{u}_{j-1/2}| = O(\Delta x^{r+1}). \tag{3.11}$$

At the present time this notion of accuracy has not been shown to have rigorous theoretical significance. Nevertheless, extensive numerical evidence demonstrates the dramatic improvement of the quality of approximations coming from certain higher-order methods; see [7], [8] for example.

LEMMA 3.3. *Suppose that $u_0 \in C^4$ and that Δt is taken sufficiently small so that $\mathscr{R}^{\Delta}(u_0)$ satisfies* (3.4). *Moreover assume that $u_0(x)$ satisfies* (3.10) *for all $x \in I_{j-1} \cup I_j$. Then we have that*

$$|\mathscr{A}_{j-1/2} - \bar{u}_{j-1/2}| \leq \text{const}\, \Delta x^4,$$

where $\mathscr{A}_{j-1/2}$ is given by the reconstruction algorithm of Section 2 combined with the evolution operator (3.9), *and $\bar{u}_{j-1/2}$ is the $j-1/2$ cell average of the exact solution to* (3.1) *at $t = \Delta t/2$.*

Proof. Since assumption (3.10) implies that no modification of P^1_{j-1} or P^1_j is performed in the reconstruction, we have that in $I_{j-1} \cup I_j$

$$v^0(x) = \begin{cases} P^1_{j-1}(x - x_{j-1/2}), & x \in I_{j-1}, \\ P^1_j(x - x_{j+1/2}), & x \in I_j; \end{cases}$$

see the remark of the previous section. The divergence theorem allows us to rewrite $\mathscr{A}_{j-1/2}$ as

$$\begin{aligned} \mathscr{A}_{j-1/2} = {} & \frac{1}{\Delta x}\left[\int_{x_{j-1/2}}^{x_j} P^1_{j-1}(x - x_{j-1/2})\, dx + \int_{x_j}^{x_{j+1/2}} P^1_j(x - x_{j+1/2})\, dx\right] \\ & - \frac{1}{\Delta x}\left[\int_0^{\Delta t/2} (f(T(v^0)(x_{j+1/2}, t)) - f(T(v^0)(x_{j-1/2}, t)))\, dt\right], \end{aligned} \tag{3.12}$$

and also to write $\bar{u}_{j-1/2}$ as

$$\begin{aligned} \bar{u}_{j-1/2} = {} & \frac{1}{\Delta x}\left[\int_{x_{j-1/2}}^{x_{j+1/2}} u_0(x)\, dx\right] \\ & - \frac{1}{\Delta x}\left[\int_0^{\Delta t/2} (f(T(u_0)(x_{j+1/2}, t)) - f(T(u_0)(x_{j-1/2}, t)))\, dt\right]. \end{aligned} \tag{3.13}$$

Identity (2.3) gives us immediately that

$$\frac{1}{\Delta x}\left|\int_{x_{j-1/2}}^{x_j} P_{j-1}^1(x-x_{j-1/2})\,dx+\int_{x_j}^{x_{j+1/2}} P_j^1(x-x_{j+1/2})\,dx\right.$$
$$\left.-\int_{x_{j-1/2}}^{x_{j+1/2}} u_0(x)\,dx\right| \tag{3.14}$$
$$\leq \text{const}\,\Delta x^4,$$

where the constant above depends on $\sup_{I_{j-1}\cup I_j}(|u_0^{\mathrm{IV}}|)$. The quantity

$$E_{j+1/2}^f(t)\equiv f(T(v^0)(x_{j+1/2},t))-f(T(u_0)(x_{j+1/2},t))$$

can be investigated over the time interval $0\leq t\leq \Delta t/2$ using characteristics. One easily finds that

$$T(v^0)(x_{j+1/2},t)=P_j^1(s_{j+1/2}^v(t)-x_{j+1/2}),$$
$$T(u^0)(x_{j+1/2},t)=u_0(s_{j+1/2}^u(t)),$$

where $s_{j+1/2}^v(t)$ and $s_{j+1/2}^u(t)$ satisfy

$$x_{j+1/2}=s_{j+1/2}^v(t)+f'(v^0(s_{j+1/2}^v(t)))t,$$
$$x_{j+1/2}=s_{j+1/2}^u(t)+f'(u_0(s_{j+1/2}^u(t)))t.$$

The characteristic equations above, along with the error estimate (2.2d), combine to imply that

$$|s_{j+1/2}^v(t)-s_{j+1/2}^u(t)|\leq \text{const}\,\Delta x^3 t, \tag{3.15}$$

provided that $P_j^1(x-x_{j+1/2})$ satisfies the slope condition (3.4) and $0\leq t\leq \Delta t/2$. It should be noted at this point that (3.15) remains valid whether assumption (3.10) is satisfied or not. Adding and subtracting $f(v^0(s_{j+1/2}^u(t)))$ into $E_{j+1/2}^f(t)$ yields the identity

$$E_{j+1/2}^f(t)=f'(u_0(s_{j+1/2}^u(t)))(v^0(s_{j+1/2}^u(t))-u_0(s_{j+1/2}^u(t)))$$
$$+O(1)\Delta x^3 t+O(\Delta x^6), \tag{3.16}$$

which is again valid whether assumption (3.10) is satisfied or not. The same analysis yields an expansion for $E_{j-1/2}^f(t)$ with $j-1/2$ replacing $j+1/2$ in the formula above. Moreover,

$$s_{j+1/2}^u(t)-s_{j-1/2}^u(t)=\Delta x+O(\Delta x t), \tag{3.17}$$

which as above is valid when (3.4) is satisfied and $0\leq t\leq \Delta t/2$. Finally, noting that $P_{j-1}^1(x-x_{j-1/2})$ and $P_j^1(x-x_{j+1/2})$ both satisfy smooth error formulas (see identity (2.3)), we easily conclude with the aid of (3.17) that

$$\frac{1}{\Delta x}\left|\int_0^{\Delta t/2}(E_{j+1/2}^f(t)-E_{j-1/2}^f(t))\,dt\right|\leq \text{const}\,\Delta x^2\Delta t^2,$$

which completes the proof of the lemma.

To see that the order of accuracy can be no less than two, regardless of whether u_0 satisfies (3.10) in $I_{j-1}\cup I_j$ or not, simply return to Eqs. (3.14) and (3.16) of the lemma above and recall from (2.2d) that

$$u(x)=\mathscr{R}^{\Delta}(u)(x)+O(\Delta x^3),$$

for any smooth function $u(x)$.

Next we discuss the piecewise parabolic reconstruction of $T(v^0)(x, \Delta t/2)$. The evolution procedure described above yields the quantities

$$\int_{I_{j-1/2}} T(v^0)(x, \Delta t/2)\, dx, \tag{3.18a}$$

$$T(v^0)(x_{j-1/2}, \Delta t/2), \tag{3.18b}$$

$$T(v^0)(x_{j+1/2}, \Delta t/2), \tag{3.18c}$$

in each cell $I_{j-1/2}$. This is enough information to construct the basic parabolic approximation $P^1_{j-1/2}(x - x_j)$. To implement the full reconstruction of Section 2, we need information concerning the values of

$$\max_{I_{j-1/2}} (T(v^0)(\cdot, \Delta t/2)), \qquad \min_{I_{j-1/2}} (T(v^0)(\cdot, \Delta t/2)).$$

However, these values are in general difficult to calculate exactly. In place of them we use instead

$$\max_{[s_{j-1}, s_j]} (v^0(\cdot)), \tag{3.19a}$$

$$\min_{[s_{j-1}, s_j]} (v^0(\cdot)), \tag{3.19b}$$

to calculate τ_L and τ_R; see Theorem 2.1. Over cell $I_{j-1/2}$, (3.19a) and (3.19b) serve as an upper and lower bound for the maximum and minimum of $T(v^0)(x, \Delta t/2)$, respectively. Therefore, using (3.19) does not affect the accuracy of the approximation. In fact, if v^0 is continuous across the interface between cells I_{j-1} and I_j, (3.19) gives the maximum and minimum of $T(v^0)(x, \Delta t/2)$ on $I_{j-1/2}$ exactly.

Define the reconstruction of $T(v^0)(x, \Delta t/2)$ by

$$\mathscr{R}^S(T(v^0)(\cdot, \Delta t/2))(x) = \sum_j P_{j-1/2}(x - x_j) + v^0(s_{j-1})\delta(x - x_{j-1/2}), \tag{3.20}$$

where $P_{j-1/2}(x - x_j)$ is obtained from the algorithm of Theorem 2.1, using (3.18) and (3.19) on the staggered mesh $\bigcup_j I_{j-1/2}$. The results and techniques of Lemma 3.2 and Theorem 2.1 now combine to make the proof of the following lemma obvious.

LEMMA 3.4. *Given that $u_0 \in BV$, and that $v^0(x)$ and $\mathscr{R}^S(T(v^0)(\cdot, \Delta t/2))(x)$ are obtained by the methods described above, we have the estimates*

(i) $$\begin{aligned} &\operatorname{Var}(\mathscr{R}^S(T(v^0)(\cdot, \Delta t/2))) \le \operatorname{Var}(v^0) \le \operatorname{Var}(u_0), \\ &\sup(\mathscr{R}^S(T(v^0)(\cdot, \Delta t/2))) \le \sup(v^0) \le \sup(u_0), \end{aligned}$$

(ii) $$\inf(\mathscr{R}^S(T(v^0)(\cdot, \Delta t/2))) \ge \inf(v^0) \ge \inf(u_0).$$

We have assumed throughout that the reconstruction algorithm applied to preconditioned data yields an approximation that satisfies properties (2.2a)–(2.2d) as well as satisfying property (3.4) (a particular preconditioning method is described at the end of this section). Therefore, defining v^1 by

$$v^1(x) = \overline{\mathscr{R}}^S(T(v^0)(\cdot, \Delta t/2))(x)$$

and succeeding v^n analogously, we have that

$$\text{Var}(v^n) \leq \text{Var}(u_0), \tag{3.21a}$$

$$\sup(v^n) \leq \sup(u_0), \qquad \inf(v^n) \geq \inf(u_0), \tag{3.21b}$$

for all $n \geq 0$. Extend the discrete time approximation v^n to all $t \geq 0$ by

$$v^\Delta(x,t) = \sum_n T(v^n)(x, t - n\Delta t/2)\chi_n(t), \tag{3.22}$$

where

$$\chi_n(t) = \begin{cases} 1, & n\Delta t \leq 2t \leq (n+1)\Delta t, \\ 0, & \text{otherwise.} \end{cases}$$

The usual techniques combine to show that every sequence $\{v^\Delta\}$, with Δx and Δt tending to zero, with fixed ratio $\Delta t/\Delta x$ satisfying the Courant condition, has an L^1 convergent subsequence on any compact subset of $\mathbf{R} \times \mathbf{R}^+$; see [5], [12]. In addition, we have

THEOREM 3.1. *If $v = \lim_{\Delta x \to 0} v^\Delta$, with $\Delta t/\Delta x$ fixed and satisfying the Courant condition* (3.5), *and if $u_0 \in BV$, then v is a weak solution to* (3.1).

Proof. Let $\varphi \in C^\infty(\mathbf{R} \times \mathbf{R}^+)$ have compact support. A straightforward calculation gives the identity

$$\begin{aligned} \int_{\mathbf{R}\times\mathbf{R}^+} (v^\Delta \varphi_t + f(v^\Delta)\varphi_x)\,dx\,dt + \int_{\mathbf{R}} u_0\varphi(x,0)\,dx \\ = \sum_{n=1} \int_{\mathbf{R}} (T(v^n) - \bar{\mathscr{R}}^\Delta(T(v^n)))\varphi(x,t^n)\,dx \\ + \int_{\mathbf{R}} (u_0 - \bar{\mathscr{R}}^\Delta(u_0))\varphi(x,0)\,dx. \end{aligned}$$

The idea is to show that the right-hand side of the identity above tends to zero as Δx tends to zero. Using the result of Lemma 3.2 in [13], together with properties (2.2a) and (2.2b), we easily find that for any $u \in BV$

$$\int_{\mathbf{R}} |u - \bar{\mathscr{R}}^\Delta(u)|\,dx \leq 2\Delta x\,\text{Var}(u).$$

This inequality allows us to bound the absolute value of the identity above by

$$\left[\sum_{n=1} \left(\max_{|x-y|\leq\Delta x} |\varphi(x,t^n) - \varphi(y,t^n)| \right) + \max|\varphi(x,0)| \right] \cdot 2\Delta x\,\text{Var}(u_0),$$

which tends to zero as Δx tends to zero.

3.2. Preconditioning of the Data. The preconditioning algorithm we describe below serves two purposes. First, it mollifies the basic piecewise quadratic interpolation in regions of large variation. Second, it is designed to limit the slope of the approximation on the interior of grid cells so that (3.4) is guaranteed to be satisfied. Moreover, our preconditioning is designed so that the error formula (2.3) does not degrade through terms of order Δx^3. Since preconditioning will be the main topic of future work, we give here only a rough sketch of the particular technique we use in the numerical examples of the next section.

Let $v(x_j)$ represent the point values of the approximation at cell interfaces x_j (see (3.18) and note the obvious change of notation). Moreover, let $\mathscr{A}_j$ be given by (3.18a). Formula (2.8) implies that

$$\begin{aligned}(v(x_{j+1}) - \mathscr{A}_j) + O(\Delta x^4)\\ = (\mathscr{A}_{j-1} - v(x_{j-1})) + 2(\mathscr{A}_{j-1} - \mathscr{A}_j) + 4(\mathscr{A}_j - v(x_j)),\end{aligned} \tag{3.23a}$$

and

$$\begin{aligned}(v(x_j) - \mathscr{A}_j) + O(\Delta x^4)\\ = (\mathscr{A}_{j+1} - v(x_{j+2})) + 2(\mathscr{A}_{j+1} - \mathscr{A}_j) + 4(\mathscr{A}_j - v(x_{j+1})).\end{aligned} \tag{3.23b}$$

Denote the right-hand sides of (3.23a) and (3.23b) by $\widetilde{u_R}$ and $\widetilde{u_L}$, respectively. Now define $\overline{u_R}$ and $\overline{u_L}$ by the following recipe. For the case when $|v(x_{j+1}) - \mathscr{A}_j| > |v(x_j) - \mathscr{A}_j|$, set

$$\begin{aligned}\overline{u_R} &= \max(|v(x_j) - \mathscr{A}_j|, \min(|v(x_{j+1}) - \mathscr{A}_j|, |\widetilde{u_R}|))\\ &\quad \cdot \operatorname{sgn}(v(x_{j+1}) - \mathscr{A}_j),\\ \overline{u_L} &= v(x_j) - \mathscr{A}_j,\end{aligned} \tag{3.24a}$$

and when $|v(x_j) - \mathscr{A}_j| \geq |v(x_{j+1}) - \mathscr{A}_j|$ set

$$\begin{aligned}\overline{u_R} &= v(x_{j+1}) - \mathscr{A}_j,\\ \overline{u_L} &= \max(|v(x_{j+1}) - \mathscr{A}_j|, \min(|v(x_j) - \mathscr{A}_j|, |\widetilde{u_L}|))\\ &\quad \cdot \operatorname{sgn}(v(x_j) - \mathscr{A}_j).\end{aligned} \tag{3.24b}$$

Again, we have omitted the subscripts on the left-hand side for convenience. So we have by using (3.23) and (3.24) that

$$\overline{u_R} = v(x_{j+1}) - \mathscr{A}_j + O(\Delta x^4)$$

and

$$\overline{u_L} = v(x_j) - \mathscr{A}_j + O(\Delta x^4).$$

Therefore, if we replace the point values $v(x_j)$ and $v(x_{j+1})$ in cell I_j by $\overline{u_L} + \mathscr{A}_j$ and $\overline{u_R} + \mathscr{A}_j$, respectively, we do not affect the results of Theorem 2.1. This defines our mollification procedure.

To limit the slope of $P_j(x - x_{j+1/2})$ so that (3.4) is satisfied, fix $S > 0$ so that

$$S < 2/(\lambda \cdot \max|f''|).$$

Next, define d by

$$d = \begin{cases} \min(|\overline{u_L}|, |\overline{u_R}|, S/2) & \text{if } \overline{u_L} \cdot \overline{u_R} < 0,\\ 0 & \text{otherwise,}\end{cases}$$

and set

$$u_{L^{2\text{nd}}} = d \cdot \operatorname{sgn}(\overline{u_L}), \qquad u_{R^{2\text{nd}}} = d \cdot \operatorname{sgn}(\overline{u_R}).$$

By construction, $u_{L^{2\text{nd}}}$ and $u_{R^{2\text{nd}}}$ yield a function that automatically satisfies the slope condition (3.4). Next, define

$$\begin{aligned}S_2 &= 2 \cdot d,\\ S_3 &= 2 \cdot \max(|2\overline{u_R} + \overline{u_L}|, |2\overline{u_L} + \overline{u_R}|),\\ \eta &= \max\left(\frac{S_3 - S}{S_3 - S_2}, 0\right),\end{aligned}$$

and note that $0 \leq \eta \leq 1$. S_3 is nothing more than the maximum of $|dP_j(\theta)/d\theta|$ on the interval I_j. We have from Section 2 that

$$S_3 - S = \left| \frac{du}{dx} + O(\Delta x) \right| \Delta x - S,$$

which therefore implies that $\eta = 0$ for smooth functions when Δx is sufficiently small. Defining the final preconditioning by the modified end points

$$\widehat{u_L} = (1-\eta)\overline{u_L} + \eta u_{L^{2\text{nd}}}, \qquad \widehat{u_R} = (1-\eta)\overline{u_R} + \eta u_{R^{2\text{nd}}}, \tag{3.25}$$

one easily verifies that the reconstruction will now satisfy the slope condition (3.4).

4. Numerical Experiments. Understandably, one could surmise on first reading that the precondition-reconstruction-evolution algorithm described in the sections above is complicated and possibly difficult to program. Needless to say, we believe that this is not the case. In our FORTRAN program we use separate subroutines to perform the preconditioning and reconstruction step. The main routine performs the evolution steps along with program initialization and output of the approximation. The preconditioning and evolution routines are straightforward and we believe warrant no further discussion. A simple to read, while admittably not very efficient, FORTRAN code for the reconstruction step of Section 2 is given at the end of this section. The preconditioning and reconstruction routines return modified cell boundary interface values for the approximation and leave fixed the approximation cell average.

The first numerical example we present is simple linear advection defined by the equation

$$\frac{\partial}{\partial t}u + \frac{\partial}{\partial x}u = 0, \tag{4.1}$$

with periodic boundary conditions

$$u(0) = u(1),$$

and initial datum

$$u_0(x) = \sin(2\pi x).$$

For this example we take the ratio $\Delta t/\Delta x$ to be 0.8. Figures 1a and 1b show the cell averages (circles) of our third-order method compared to the exact solution to (4.1) (the solid lines), using 10 grid points on $[0,1]$, after respectively two cycles and six cycles; that is $t = 2$ and $t = 6$. Notice the very slight decay of the peaks after six cycles. The fine resolution found here is retained after many more cycles. Figures 2a and 2b show the performance of the first-order Lax-Friedrichs scheme using the same parameters as above. Figure 3 demonstrates the first-order Lax-Friedrichs scheme on the example above after six cycles, this time using 100 grid points. Fixing $\eta \equiv 1$ in the preconditioning step (3.25) defines a second-order TVD scheme; variants of this second-order scheme will be the topic of a future paper. Figures 4a and 4b illustrate the performance of this second-order scheme with 10 grid points, again after two and six cycles respectively. To our knowledge, no other TVD scheme yields approximations of the quality exemplified by our third-order method. This is not surprising since ours is at present the only TVD scheme that is guaranteed to retain second-order accuracy at extreme points of the approximation.

Numerical results of comparable quality have been previously obtained from ENO schemes [8], [10], however various theoretical questions concerning the stability of ENO approximations remain unresolved. Letting the problem above run through 100 cycles gives Figure 5a, our third-order TVD method, and Figure 5b, our *second-order* TVD method. The fact that the cell averages of 5b appear to be flat is no illusion; its variation is on the order of 10^{-3}.

The second example we consider here is the nonlinear problem

$$\frac{\partial}{\partial t}u + \frac{\partial}{\partial x}(u^2/2) = 0, \tag{4.2}$$

with the same boundary conditions and initial datum as above. An easy calculation verifies that the solution to (4.2) remains smooth until $t = 1/2\pi$, after which a shock forms. We use this example simply to indicate the convergence of our third-order method. Again we take $\Delta t/\Delta x = 0.8$. Figures 6a, 6b and 6c illustrate the third-order method at $t = 0.16 \simeq 1/2\pi$, using respectively 20, 40 and 80 grid points. Figures 7a, 7b and 7c illustrate the third-order method at $t = 0.32$, again using 20, 40 and 80 grid points.

The final problem we examine is given by

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

where $f(u)$ is the nonconvex function

$$f(u) = \frac{u^2}{u^2 + (1-u)^2/4}.$$

The initial datum for this problem is assumed to lie in the interval $[0, 1]$, and the ratio $\Delta t/\Delta x$ is fixed so that

$$\max_{u\in[0,1]} |f'(u)|\frac{\Delta t}{\Delta x} = 0.8.$$

This problem has particular interest since certain TVD schemes have been found to frequently "latch" onto entropy violating solutions.

We consider first the datum,

$$u_0(x) = \begin{cases} 1, & x < 1/4, \\ 0, & x \geq 1/4. \end{cases}$$

Figures 8a, 8b and 8c compare respectively our third-order TVD method, our second-order TVD method and the Lax-Friedrichs method, using 40 spatial grid points, to the exact solution at $t = 0.404$. The most notable difference between the third-order approximation and the second-order approximation is at the interface between the constant state on the left and the rarefaction wave. Additionally, we have found that on this example certain well-known schemes tend to "overcompress" in the region between the rarefaction wave and the moving discontinuity. As is seen in Figure 8, there is no evidence of this phenomenon with our methods. Figure 9 illustrates the problem with datum

$$u_0(x) = \begin{cases} 0, & x < 1/4, \\ 1, & 1/4 \leq x < 1/2, \\ 0, & 1/2 \leq x, \end{cases}$$

displayed at $t = 0.202$. Figure 9a demonstrates the performance of our third-order method using 40 grid points, 9b demonstrates the performance of our second-order method using 40 grid points, and 9c demonstrates the performance of the Lax-Friedrichs schemes, this time however using 200 grid points.

```
      SUBROUTINE RECON(UMIN,UMAX,AVE,UL,UR)
C*
C*    THIS SUBROUTINE ALTERS THE QUADRATIC INTERPOLATION:
C*     P = 3*(UL+UR-2*AVE)*THETA**2 + (UR-UL)*THETA + (AVE-(UL+UR-2*AVE)/4),
C*     THETA = (X-XMID)/DX, XMID = (XL+XR)/2, XL < X < XR,
C*  BY MODIFYING THE CELL BOUNDARY VALUES UL AND UR IN SUCH A WAY AS
C*  TO PRESERVE THE CELL AVERAGE AVE, MAINTAIN THIRD ORDER ACCURACY
C*  WHEN APPLIED TO A FUNCTION U WITH CELL MINIMUM (MAXIMUM) VALUE
C*  UMIN (UMAX), AND TO FURTHERMORE GUARANTEE THAT THE RESULTING
C*  INTERPOLATION P HAS ITS CELL VARIATION BOUNDED BY THE CELL VAR-
C*  IATION OF U.
C*
      REAL MAXMOD, MINMOD

      ULHAT  = UL - AVE
      URHAT  = UR - AVE
      AULHAT = ABS( ULHAT )
      AURHAT = ABS( URHAT )

      IF(AULHAT.GE.AURHAT) THEN
        MAXMOD = ULHAT
        MINMOD = URHAT
      ELSE
        MAXMOD = URHAT
        MINMOD = ULHAT
      ENDIF

      IF(MAXMOD.EQ.0.0) RETURN

      RHO    = MINMOD/MAXMOD

      IF(RHO.LE.-0.5) RETURN

      CRITVAL = -MAXMOD*( 1.0+RHO+RHO*RHO )/( 3.0*(1.0+RHO) )
      IF(MAXMOD.GT.0.0) THEN
        E = UMIN - AVE
        IF(CRITVAL.GE.E) RETURN
      ELSE
        E = UMAX - AVE
        IF(CRITVAL.LE.E) RETURN
      ENDIF

      EHAT =  E/MAXMOD
      IF(RHO.LT.0.0) THEN
        TAUMIN = 1.0
        TAUMAX = 0.5*(-(RHO+3.0*EHAT)+SQRT( 3.0*(RHO+3.0*EHAT)*(EHAT-RHO) ))
      ELSE
        TAU    = -EHAT*( 3.0*(1.0+RHO)/(1.0+RHO+RHO*RHO) )
        TAUMIN =  TAU
        TAUMAX =  TAU
      ENDIF

      IF(AULHAT.GE.AURHAT) THEN
        UL = TAUMAX*ULHAT + AVE
        UR = TAUMIN*URHAT + AVE
      ELSE
        UL = TAUMIN*ULHAT + AVE
        UR = TAUMAX*URHAT + AVE
      ENDIF

      RETURN
      END
```

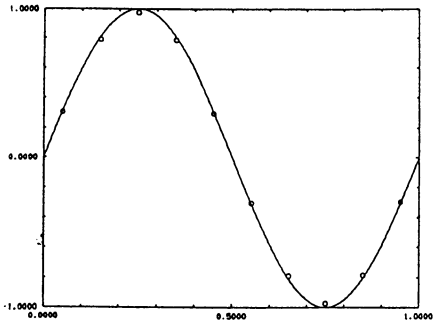

FIGURE 1a

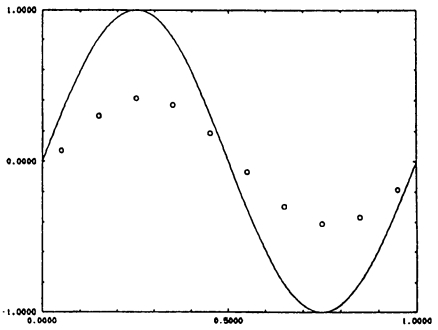

FIGURE 2a

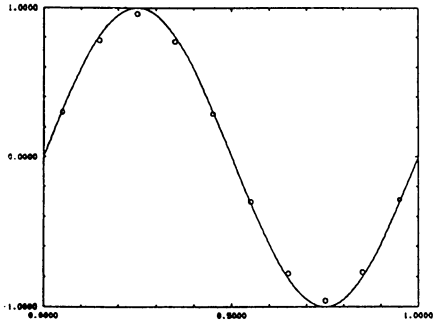

FIGURE 1b

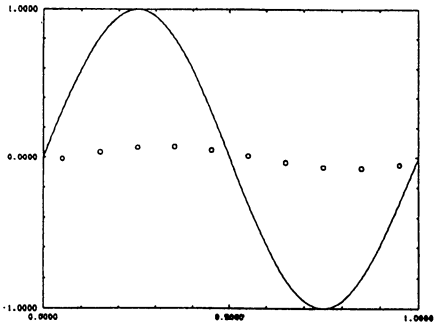

FIGURE 2b

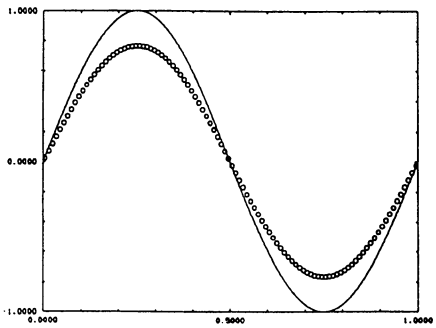

FIGURE 3

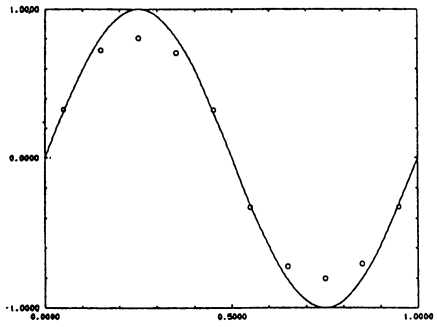

FIGURE 4a

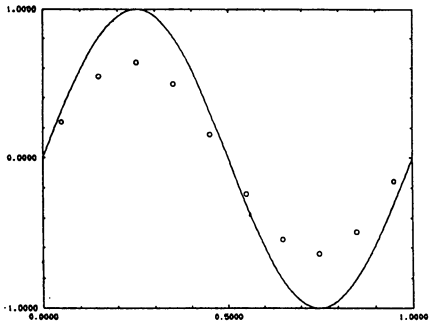

FIGURE 5a

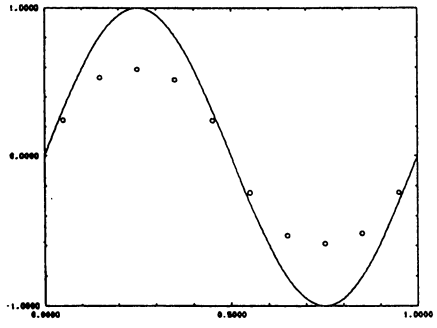

FIGURE 4b

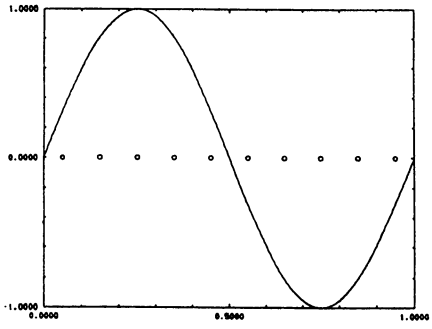

FIGURE 5b

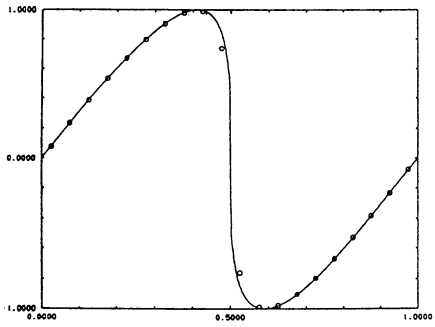

FIGURE 6a

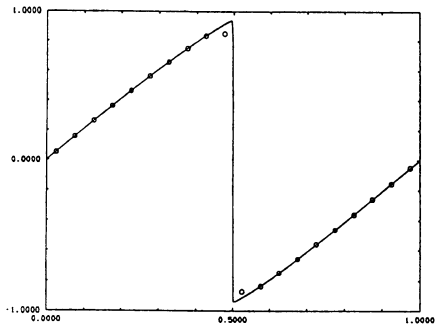

FIGURE 7a

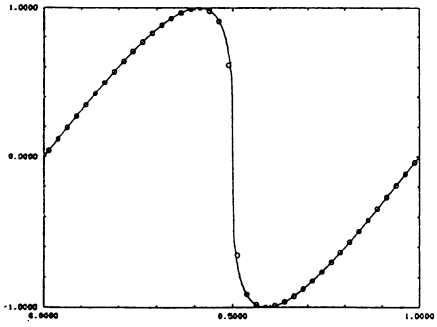

FIGURE 6b

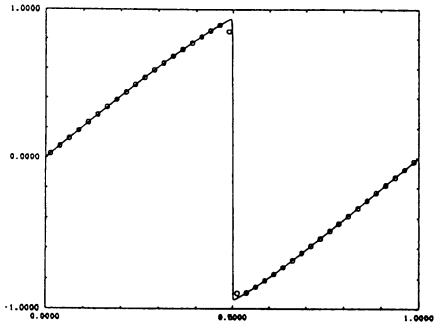

FIGURE 7b

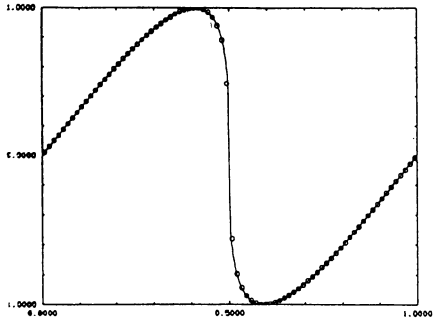

FIGURE 6c

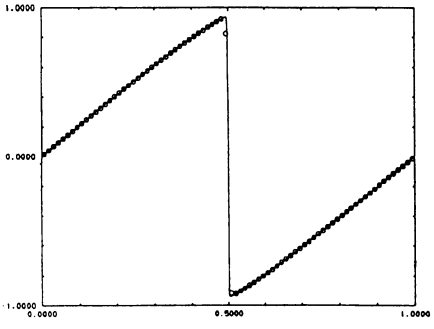

FIGURE 7c

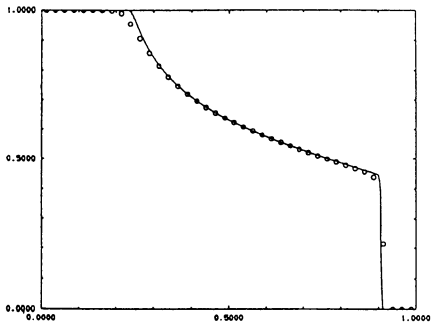

FIGURE 8a

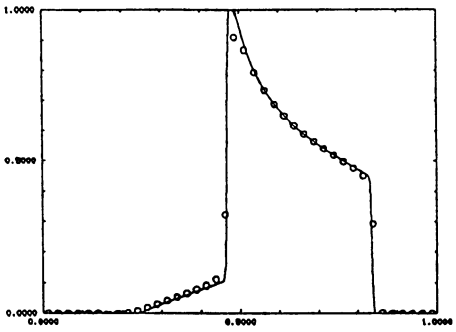

FIGURE 9a

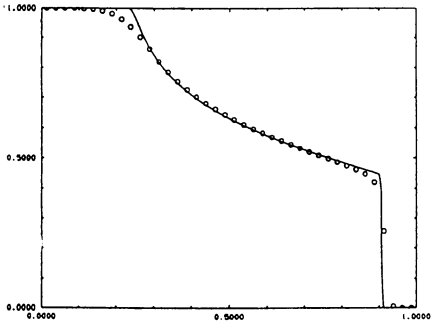

FIGURE 8b

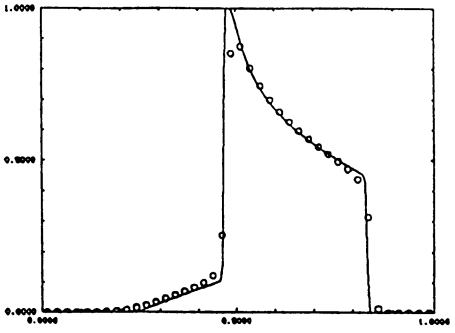

FIGURE 9b

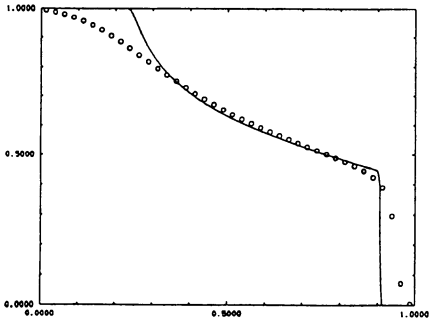

FIGURE 8c

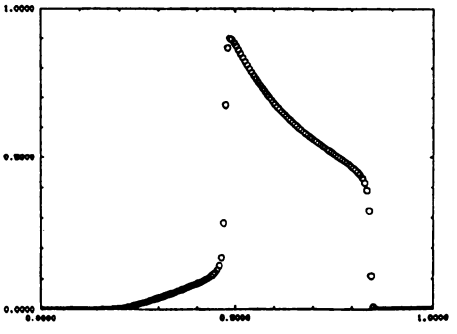

FIGURE 9c

Department of Mathematics
University of Houston
Houston, Texas 77004

1. M. BEN-ARTZI & J. FALCOVITZ, "A second-order Godunov-type scheme for compressible fluid dynamics," *J. Comput. Phys.*, v. 55, 1984, pp. 1–32.

2. S. R. CHAKRAVARTHY, A. HARTEN & S. OSHER, *Essentially Non-Oscillatory Shock-Capturing Schemes of Arbitrarily High Accuracy*, AIAA 24th Aerospace Sciences Meeting, January 6–9, 1986, Reno, Nevada.

3. P. COLELLA & P. R. WOODWARD, "The piecewise-parabolic method (PPM) for gas-dynamical simulations," *J. Comput. Phys.*, v. 54, 1984, pp. 174–201.

4. S. D. CONTE & C. DE BOOR, *Elementary Numerical Analysis*, 3rd ed., McGraw-Hill, New York, 1980.

5. M. CRANDALL & A. MAJDA, "Monotone difference approximations for scalar conservation laws," *Math. Comp.*, v. 34, 1980, pp. 1–22.

6. S. K. GODUNOV, "A finite difference method for the numerical computation of discontinuous solutions of the equations of fluid dynamics," *Mat. Sb.*, v. 47, 1959, pp. 271–295.

7. A. HARTEN, "High resolution schemes for hyperbolic conservation laws," *J. Comput. Phys.*, v. 49, 1983, pp. 357–393.

8. A. HARTEN, B. ENGQUIST, S. OSHER & S. R. CHAKRAVARTHY, "Uniformly high order accurate essentially non-oscillatory schemes III," *J. Comput. Phys.*, v. 71, 1987, pp. 231–303.

9. A. HARTEN, J. M. HYMAN & P. D. LAX, "On finite difference approximations and entropy conditions for shocks," *Comm. Pure Appl. Math.*, v. 29, 1976, pp. 297–322.

10. A. HARTEN & S. OSHER, "Uniformly high-order accurate non-oscillatory schemes I," *SIAM J. Numer. Anal.*, v. 24, 1987, pp. 279–309.

11. P. D. LAX, "Shock waves and entropy," *Contributions to Nonlinear Functional Analysis* (E. H. Zarantonello, ed.), Academic Press, New York, 1971, pp. 603–634.

12. R. SANDERS, "On convergence of monotone finite difference schemes with variable spatial differencing," *Math. Comp.*, v. 40, 1983, pp. 91–106.

13. R. SANDERS, "The moving grid method for nonlinear hyperbolic conservation laws," *SIAM J. Numer. Anal.*, v. 22, 1985, pp. 713–728.

14. P. K. SWEBY, "High resolution schemes using flux limiters for hyperbolic conservation laws," *SIAM J. Numer. Anal.*, v. 21, 1984, pp. 995–1011.

15. B. VAN LEER, "Towards the ultimate conservative scheme, II. Monotonicity and conservation combined in a second order scheme," *J. Comput. Phys.*, v. 14, 1974, pp. 361–376.