MATHEMATICS OF COMPUTATION, VOLUME 34, NUMBER 150
APRIL 1980, PAGES 391–400

Computation of Steady Shocks by Second-Order Finite-Difference Schemes

By Lasse K. Karlsen

Abstract. The computational stability of steady shocks which satisfy the entropy condition is considered for the scalar conservation law

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{1}{2}u^2\right) = 0.$$

It is shown that the computation of the pure initial value problem by Lax-Wendroff type schemes approaches a steady state if the initial data satisfies a specified condition, and that this condition is always satisfied for the corresponding initial-boundary value problem with a finite number of grid points. The effect of machine accuracy on the influence of the boundaries on the error near the shock is also discussed.

1. Introduction. In a recent paper by Harten et al. [1] the existence of steady discontinuous numerical solutions was considered for the nonlinear conservation law

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

It was proved that the Lax-Wendroff scheme is linearly unstable when applied to the initial value problem with initial data close to the exact steady discontinuous solution of the differential problem. In this paper we shall consider the case $f = \frac{1}{2}u^2$ a little further since this particular equation has some similarity to gas-dynamic shocks. For the same reason we shall only consider discontinuities which satisfy the entropy condition [2]. It will be shown that the computation of the pure initial value problem will reach a steady state for a restricted class of initial data and that this will always be the case for the initial-boundary value problem with a finite number of mesh points. In addition to the Lax-Wendroff scheme, some other second-order schemes are also considered.

2. The Initial Value Problem. As in [1] we consider perturbations w_j of the exact steady shock

$$u = \begin{cases} 1, & x, j \leqslant 0, \\ -1, & x, j > 0. \end{cases}$$

Linearization of the Lax-Wendroff operator

$$v_j^{n+1} = v_j^n - \tfrac{1}{2}h(f_{j+1}^n - f_{j-1}^n) + \tfrac{1}{2}h^2\left[a_{j+\frac{1}{2}}^n(f_{j+1}^n - f_j^n) - a_{j-\frac{1}{2}}^n(f_j^n - f_{j-1}^n)\right],$$

Received April 2, 1979.
AMS (MOS) subject classifications (1970). Primary 65M10; Secondary 76L05.

0025-5718/80/0000-0052/$03.50

where $h = \Delta t/\Delta x$ and $a = f' = u$ gives a linear perturbation operator of the form $W_j = T_w w_j$ for the application of the Lax-Wendroff operator once. After the change of variables

$$p_j = w_j + w_{1-j}, \qquad P_j = W_j + W_{1-j},$$
$$q_j = w_j - w_{1-j}, \qquad Q_j = W_j - W_{1-j},$$

one gets the two new operators $P_j = T_p p_j$ and $Q_j = T_q q_j$. Using the Kreiss theory [3], i.e. looking for solutions of the form $P_j = \rho p_j$, it is found that the T_p-operator has an eigenvector p with $\rho = 1$, and is weakly unstable since the L_2-norm $\| T_p^n \| \sim$ const $\sqrt{n}$, whereas the T_q-operator is stable.

The T_p-operator is given by

$$P_j = (\tfrac{1}{2}b + \tfrac{1}{2}b^2)p_{j+1} + (1 - b^2)p_j + (-\tfrac{1}{2}b + \tfrac{1}{2}b^2)p_{j-1}, \qquad j > 1, \tag{1}$$

$$P_1 = (1 + \tfrac{1}{2}b - \tfrac{1}{2}b^2)p_1 + (\tfrac{1}{2}b + \tfrac{1}{2}b^2)p_2, \tag{1a}$$

where $b = hf'(1) = h$ (this is of opposite sign to b in [1] since we consider only shocks which satisfy the entropy condition). Equation (1a) is equivalent to (1) with the boundary condition

$$p_0 = \frac{b + 1}{b - 1} p_1.$$

We shall now consider a new operator T_r which is obtained by the change of variable

$$r_j = (\tfrac{1}{2}b + \tfrac{1}{2}b^2)p_{j+1} - (-\tfrac{1}{2}b + \tfrac{1}{2}b^2)p_j, \tag{2}$$

which gives

$$R_j = (\tfrac{1}{2}b + \tfrac{1}{2}b^2)r_{j+1} + (1 - b^2)r_j + (-\tfrac{1}{2}b + \tfrac{1}{2}b^2)r_{j-1}, \qquad r_0 = 0.$$

From the argument in [1] for the T_q-operator it follows that $T_r^*(-b)$ is stable and, hence, so is T_r since $T_r(b) = T_r^*(-b)$. As $n \longrightarrow \infty$, we obtain $\| T_r^n \| \longrightarrow 0$ and from the definition of r in Eq. (2) the steady error profile

$$p_{j+1} = \lambda p_j, \qquad \lambda = \frac{b - 1}{b + 1}.$$

To obtain this steady result we must obviously require the initial data for the T_r-operator to be bounded, i.e. $\|r\| \leqslant K$, which from the definition of the L_2-norm [4] is equivalent to

$$\Delta x \tfrac{1}{4} b^2 (b + 1)^2 \sum_{j=1}^{\infty} \left(p_{j+1} - \frac{b - 1}{b + 1} p_j \right)^2 \leqslant K. \tag{3}$$

From the von Neumann condition $(b - 1)/(b + 1) \leqslant 0$, and the right-hand side of (3) is therefore bounded if the initial data p_j is oscillatory, $p_j p_{j+1} \leqslant 0$, and decays at least as fast as $((b - 1)/(b + 1))^j$ as $j \longrightarrow \infty$. If condition (3) is satisfied, the computation of the pure initial value problem will approach a steady state.

In the computed example for T_p^n in [1] condition (3) is satisfied since $p_j \equiv 0$ for $j \geqslant 50$ in their initial data. The information contained in the initial data for

$j \geqslant 50$ is, however, carried along characteristics running towards the shock and will not reach the shock before approximately $n = 50/b$ (≈ 55 for $b = .9$). $\| T_p^n \|$ will, therefore, only grow as $\sqrt{n}$ initially; and when the characteristic from $j = 50$ reaches the shock, the computation settles down to a steady-state error profile. This is shown in Figure 1 which is the development of $\| T_p^n \phi \|^2$ with the initial data ϕ of [1]. The approach to the steady state is not apparent in [1] because the computation is only shown to $n = 40$.

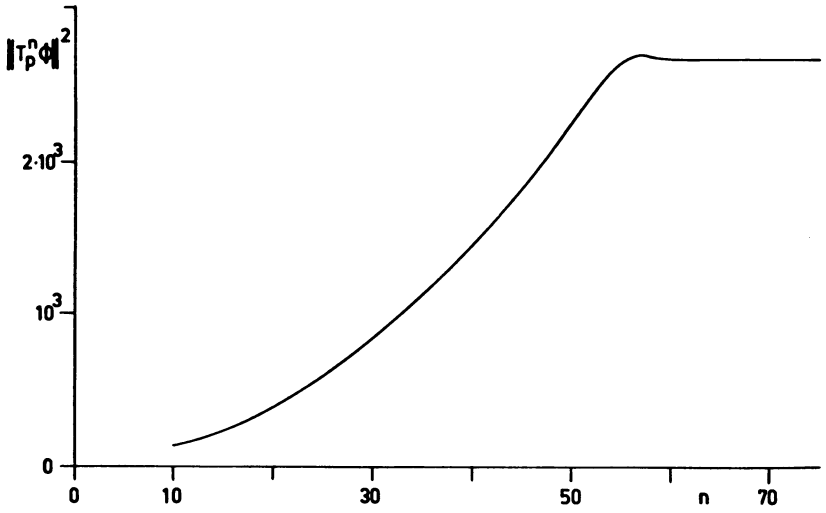

FIGURE 1

Computed p-operator with data of [1], $b = .9$

3. The Initial-Boundary Value Problem. From the previous pure initial value problem we construct an initial-boundary value problem by adding the boundary conditions $v_{-N_2}^n = 1$ and $v_{N_1}^n = -1$ ($N_1, N_2 > 0$) for all n. For the Lax-Wendroff scheme we know from the previous section that the following eigenvector has $\rho = 1$, ($W_j = \rho w_j$):

$$w_j = \tfrac{1}{2}(p_j + q_j) = A\lambda^{j-1} + \tfrac{1}{2}q_0, \qquad j > 0,$$

$$w_j = \tfrac{1}{2}(p_{1-j} - q_{1-j}) = A\lambda^{-j} - \tfrac{1}{2}q_0, \qquad j \leqslant 0,$$

where $\lambda = (b - 1)/(b + 1)$ and A, q_0 are constants. The boundary conditions give

$$w_{N_1} = 0: \quad A\lambda^{N_1 - 1} + \tfrac{1}{2}q_0 = 0,$$

$$w_{-N_2} = 0: \quad A\lambda^{N_2} - \tfrac{1}{2}q_0 = 0,$$

or

$$[\lambda^{N_1 - 1} + \lambda^{N_2}]A = 0.$$

Since the square-bracketed term is generally $\neq 0$, an eigenvector $w \neq 0$ would seemingly not exist. However, for reasonable N_i and b close to 1 the square bracket is smaller than the accumulated round-off error. An eigenvector $w \neq 0$ will, therefore,

exist in most practical computations. This effect of N_i will be shown in an example below.

In this case $q_0 \longrightarrow 0$ and the eigenvector is the same as in the pure initial value problem. The primary effect of adding the boundary condition for reasonably large N_1 and N_2 is, therefore, similar to having the initial data zero for $j \geqslant N_1$, and $j \leqslant -N_2$. The above condition (3) for a steady error profile is, therefore, fulfilled.

We expect the steady state to be reached after approximately $n_s = (1/b)\max(N_1, N_2)$ steps. For a fixed-length interval we have $\max(N_1, N_2) \sim \Delta x^{-1}$ and, hence, $n_s \sim 1/b\Delta x$. For bounded initial data the maximum error increases as $\sqrt{n_s} \sim \Delta x^{-1/2}$ as the mesh is refined. However, the L_2-norm

$$\|w\| = \Delta x \left(\sum_j w_j^2 \right)^{1/2} \sim \Delta x^{1/2};$$

and the initial-boundary value problem, therefore, converges in the L_2 sense. These results may be of doubtful value for the nonlinear operator, since the linearization breaks down when the error becomes large.

4. The MacCormack Scheme. A commonly used scheme for gas-dynamic problems is the one suggested by MacCormack (MC) [5]

$$v_j^{\overline{n+1}} = v_j^n - \sigma h(f_{j+\sigma}^n - f_j^n),$$
$$v_j^{n+1} = \tfrac{1}{2}(v_j^n + v_j^{\overline{n+1}} - \sigma h(f_j^{\overline{n+1}} - f_{j-\sigma}^{\overline{n+1}})),$$

where $\sigma = 1$ gives forward-backward differencing (FB) and conversely for $\sigma = -1$ (BF). This scheme is interesting because it is unsymmetrical at the shock and tends to give better results for gas-dynamic shocks than the Lax-Wendroff scheme. The p-operator is the same as for the Lax-Wendroff scheme. The q-operator is, however, different at the shock

$$Q_1 = (1 - \tfrac{1}{2}b - \tfrac{1}{2}b^2)q_1 + (\tfrac{1}{2}b + \tfrac{1}{2}b^2)q_2 \pm b^2 p_1.$$

The first two terms on the right-hand side are the same as for Lax-Wendroff. The last term (+ for FB and − for BF) couples the q-operator to the p-operator. This equation is equivalent to the following boundary condition at the shock: $q_0 = q_1 \pm 2bp_1/(b-1)$. For $p_j \neq 0$ we must, therefore, have when $\rho = 1$, $q_j = \pm bp_j$. The steady-state error profile for MC is, therefore, given by

$$w_j = \tfrac{1}{2}p_j(1 \pm b) = A(1 \pm b)\left(\frac{b-1}{b+1}\right)^j, \qquad j > 0,$$

$$w_j = \tfrac{1}{2}p_{1-j}(1 \mp b) = A(1 \mp b)\left(\frac{b-1}{b+1}\right)^{-j}, \qquad j \leqslant 0.$$

Depending on the choice of differencing (FB of BF), the error is smaller on one side of the shock by the factor $(b-1)/(b+1)$ compared to the other side. This of course is a well-known phenomenon for the MC scheme. The implications for the initial-boundary value problem are similar to the LW scheme.

Nonconservative, Switched MacCormack. In an earlier study using a similar linearization procedure for the steady-state solution of the MC scheme [6] it was found that the error could be removed if the differencing was switched across the shock: FB for $j \leqslant 0$ and BF for $j > 0$. This gives the same q-operator as the LW scheme, but the p-operator is now different at the shock

$$P_1 = \left(1 - \frac{1}{2}b - \frac{3}{2}b^2\right)p_1 + \left(\frac{1}{2}b + \frac{1}{2}b^2\right)p_2,$$

which is equivalent to the new boundary condition

$$p_0 = -p_1.$$

In the p-equation with $P_j = \rho p_0 \lambda^j$ this boundary condition requires $\lambda = -1$ and $\rho = 1 - 2b^2$. For $b \leqslant 1$ we get $|\rho| \leqslant 1$, the scheme is, therefore, stable; and the error will decrease as n increases. The same result is also valid for the opposite switching BF for $j \leqslant 0$, FB for $j > 0$. These two schemes, however, have the disadvantage that they are not conservative at the shock, which causes an incorrect shock speed when used in a transient flow with a moving shock. Some conservative switching schemes will, therefore, be considered next.

5. Conservative, Switched Schemes.

Switched MacCormack. The two schemes discussed above with opposite differencing on both sides of the shock can be made fully conservative if a special shock operator is applied for $j = 0$

1) FB $j < 0$, BF $j > 0$,

$$v_0^{n+1} = v_0^n - \tfrac{1}{2}h(\overline{f_1^{n+1}} - \overline{f_{-1}^{n+1}}),$$

$$W_0 = (\tfrac{1}{2}b + \tfrac{1}{2}b^2)w_1 + w_0 + (\tfrac{1}{2}b + \tfrac{1}{2}b^2)w_{-1},$$

2) BF $j < 0$, FB $j > 0$,

$$v_0^{n+1} = v_0^n - \tfrac{1}{2}h(f_1^n - f_{-1}^n),$$

$$W_0 = \tfrac{1}{2}bw_1 + w_0 + \tfrac{1}{2}bw_{-1}.$$

It is easily shown that an eigenvector for the linearized operator of scheme 1) which has $\rho = 1$ is given by

$$w_j = \lambda w_{j-1}, \qquad j > 1,$$

$$w_1 = -\lambda w_0,$$

$$w_j = \lambda w_{j+1}, \qquad j < 0,$$

where $\lambda = (b - 1)/(b + 1)$. The result for scheme 2) is similar, being the mirror image of scheme 1) in the shock. The effect of the shock operator is clearly to regain the possibility of a steady-state error. As explained below this is due to the fact that the scheme is now again fully conservative.

Upwind Differencing Schemes. A hybrid scheme which has been found to give improved results for gas-dynamic shocks is one using the upwind differencing scheme suggested by Warming and Beam (UW) [7]. This scheme is fully conservative and uses a shock point operator. For gas-dynamic applications they combine the UW scheme with the MC scheme as follows:

$$v_j^{\overline{n+1}} = v_j^n - h(f_j^n - f_{j-1}^n),$$

$$j<0 \quad v_j^{n+1} = \tfrac{1}{2}(v_j^n + v_j^{\overline{n+1}}) - \tfrac{1}{2}h(f_j^n - 2f_{j-1}^n + f_{j-2}^n) - \tfrac{1}{2}h(f_j^{\overline{n+1}} - f_{j-1}^{\overline{n+1}}),$$

$$v_0^{n+1} = v_0^n - \tfrac{1}{2}h(f_0^n - 2f_{-1}^n + f_{-2}^n) - \tfrac{1}{2}h(f_1^{\overline{n+1}} - f_{-1}^{\overline{n+1}}),$$

$j>0$ MC **BF.**

The linearized operator for the UW scheme is

$$W_j = \left(1 - \frac{3}{2}b + \frac{1}{2}b^2\right)w_j + (2b - b^2)w_{j-1} + \left(-\frac{1}{2}b + \frac{1}{2}b^2\right)w_{j-2}.$$

This scheme has no unstable eigenvectors. For $\rho = 1$ the eigenvectors are

$$w_j = \lambda w_{j+1} \quad \text{with } \lambda = 1 \text{ and } \lambda = \frac{3-b}{1-b}.$$

Since $|\lambda| \geqslant 1$ for $b \leqslant 2$ (which is the von Neumann condition for this scheme), the scheme is stable. For $j \geqslant 0$ the following eigenvector with $\rho = 1$ can exist:

$$w_1 = -\lambda w_0,$$

$$w_j = \lambda w_{j-1}, \qquad j>0, \qquad \lambda = (b-1)/(b+1),$$

which is typical for the MC scheme.

Finally we consider the use of the upwind scheme on both sides of the shock, except at $j = 0$ where the conservative shock operator is given by

$$v_0^{n+1} = v_0^n - \frac{1}{2}h(2f_1^n - f_2^n + f_1^{\overline{n+1}} - 2f_{-1}^n + f_{-2}^n - f_{-1}^{\overline{n+1}}),$$

$$W_0 = \left(-\frac{1}{2}b + \frac{1}{2}b^2\right)w_2 + \left(\frac{3}{2}b - \frac{1}{2}b^2\right)w_1 + w_0 + \left(\frac{3}{2}b - \frac{1}{2}b^2\right)w_{-1} + \left(-\frac{1}{2}b + \frac{1}{2}b^2\right)w_{-2}.$$

Because the UW scheme is stable for $j \neq 0$, all w_j $(j \neq 0) = 0$, and the last equation gives $W_0 = w_0$. The only remaining error in the steady state must, therefore, be at $j = 0$.

Steady-State Errors of Conservative Schemes. In a conservative scheme the quantity v is conserved, i.e. its sum is constant except for fluxes at the boundaries. This means that the sum of the errors (the perturbation from the exact steady state) is similarly conserved. Conservative schemes, therefore, generally have remaining errors in the steady state. The total sum of these errors is the total initial error plus a contribution from fluxes at the boundaries.

The unconservative, switched MacCormack scheme is completely stable because the extra term introduced by the switching acts as a sink as long as the errors are different from zero.

6. Computed Results. As a test case we have computed the specified initial-boundary value problem with $N_1 = 15, N_2 = 9, b = .9$ and a constant initial error $w_j = .01, -N_2 < j < N_1, w_{-N_2} = w_{N_1} = 0$. The development of $\|w\|^2$ with n for the schemes discussed above is shown in Figure 2. For all the conservative schemes $\|w\|^2$ grows approximately linearly with n up to $n \approx N_1/b$ and then approaches the value dictated by the steady-state error profile as expected. Also shown is the immediate convergence of the nonconservative, switched MC scheme to the exact steady-state solution. A comparison between the nonlinear and linear steady eigenvectors of w is shown in Table 1. Considering the relatively large maximum error the agreement is reasonably good, except possibly at the shock points $j = 0, 1$.

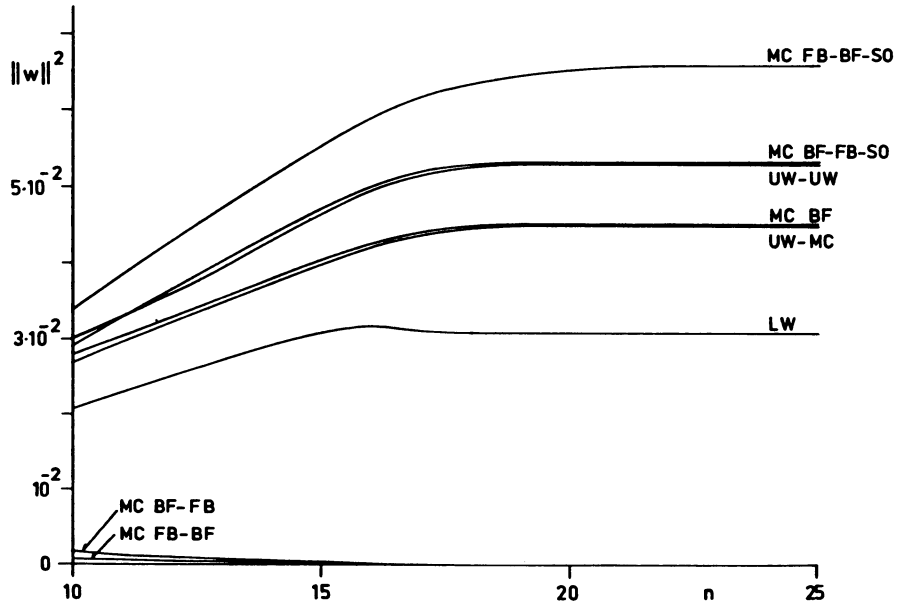

FIGURE 2

Computed $\|w\|^2$ *of initial-boundary value problem with* $b = .9$

An estimate of the total error in the steady state can be obtained for the conservative schemes. For the LW and MC schemes the error flux per step at the left boundary is

$$\tfrac{1}{2}b(1 - b)w^n_{-N_2+1}$$

as obtained from the linear approximation. The contribution from the left boundary to the total remaining error is, therefore,

$$\tfrac{1}{2}b(1 - b)(w^0_{-N_2+1} + w^1_{-N_2+1} + \cdots).$$

Because $w^n_{-N_1+1}$ rapidly approaches zero as n increases, this sum can be estimated from a few steps and was found to be $(1/b)w^0_{-N_1+1}$ for this special constant initial error. The contribution from the right boundary is the same as the left for the MC and LW operators.

For the upwind scheme the exact flux terms were used when points outside the boundaries were needed in the algorithm. The result of this procedure is that the UW operator has zero error flux at the boundary, and the initial total error is, therefore, conserved. Table 2 shows a comparison between the linear estimate and the nonlinear computed total error in the steady state. For the MC scheme it can be seen that a smaller nonlinear effect is obtained when the predictor step is taken towards the boundary. This effect does, however, depend on the type of initial values chosen.

TABLE 1

Summary of steady-state eigenvectors

Upper values: Linear operator
Lower values: Nonlinear operator

Scheme	w_0	$\frac{w_1}{w_0}$	$\frac{w_2}{w_1}$	$\frac{w_0}{w_{-1}}$	$\frac{w_{-1}}{w_{-2}}$
LW	$1.22\text{-}10^{-1}$ $1.01\text{-}10^{-1}$	1 1.41	$-5.26\text{-}10^{-2}$ $-7.99\text{-}10^{-2}$	-19 -33.2	-19 -18.8
MC (BF)	$2.31\text{-}10^{-1}$ $2.10\text{-}10^{-1}$	$5.26\text{-}10^{-2}$ $1.50\text{-}10^{-1}$	$-5.26\text{-}10^{-2}$ $-4.40\text{-}10^{-2}$	-19 -35.7	-19 -18.9
MC FB-BF-SO	$2.31\text{-}10^{-1}$ $2.49\text{-}10^{-1}$	$5.26\text{-}10^{-2}$ $1.81\text{-}10^{-1}$	$-5.26\text{-}10^{-2}$ $-4.06\text{-}10^{-2}$	-19 -5.53	-19 -24.6
MC BF-FB-SO	$2.31\text{-}10^{-1}$ $2.31\text{-}10^{-1}$	$-5.26\text{-}10^{-2}$ $-2.08\text{-}10^{-3}$	$-5.26\text{-}10^{-2}$ $-5.23\text{-}10^{-2}$	19 48.1	-19 -19.1
UW + MC	$2.20\text{-}10^{-1}$ $2.10\text{-}10^{-1}$	$5.26\text{-}10^{-2}$ $1.49\text{-}10^{-1}$	$-5.26\text{-}10^{-2}$ $-4.41\text{-}10^{-2}$	∞ $\sim 10^6$	— —
UW + UW	$2.30\text{-}10^{-1}$ $2.30\text{-}10^{-1}$	0 $\sim 10^{-6}$	— —	∞ $\sim 10^6$	— —

Finally, a test has been made on the importance of the term $S = [\lambda^{N_1-1} + \lambda^{N_2}]$ in the discussion above concerning the difference between errors obtained for the pure initial value and the initial-boundary value problems. The computations were performed on a PDP-15/76 computer which has a precision of $10^{-7} - 10^{-8}$. For $b = .9$ we must have $N_1 \approx 5$ to obtain S of similar magnitude. A test has been made for the UW + MC scheme with $N_1 = 5$ and $N_1 = 3$. For $N_1 = 5$ the scheme seems to converge very slowly towards the exact steady state, after the initial transient. Instead of a steady state, there is a slow decrease in $|w_0|$ at the rate of approximately 10^{-7} per time step. For $N_1 = 3$ S is as large as 10^{-4} which is well within machine accuracy. As expected the error now decreases and after approximately 3600 steps converges to

a new steady state with $|w_0|$ as small as 10^{-4}. In this result the error near the N_1 boundary is $\sim 10^{-8}$ which is machine accuracy. This test shows the importance of the spatial decay rates of the error eigenvectors towards the boundaries. If this rate is sufficiently rapid the boundary conditions have no effect on the errors in the interior.

TABLE 2

Comparison of steady-state total errors

Sceme	Linear estimate	Nonlinear computed
LW	.231	.2310
MC (BF)	.231	.2399
MC FB-BF-SO	.231	.2490
MC BF-FB-SO	.231	.2310
UW + MC (BF)	.2305	.2394
UW + UW	.23	.2300

In conclusion it is interesting to note that a reduction of the steady-state error for any of the conservative schemes must be accompanied by a corresponding error flux at the boundaries. If, for example, an artificial viscosity is applied without destroying the conservative property, then its effect must be felt at the boundaries in order to reduce the total error, otherwise it would just accomplish a redistribution of the errors. This implies that the initial errors must be able to propagate away from the shock and pass through the boundaries.

In the present scalar differential problem the characteristics run away from the boundaries, so there is no mechanism by which signals can propagate in the required manner. It is, therefore, doubtful if this simple problem is a good model for more complicated systems such as for example one-dimensional gasdynamic flow. For a gasdynamic shock the situation is similar on the upstream supersonic side since all characteristics run downstream. On the subsonic side, however, there are characteristics in both directions, which means that errors can also propagate away from the shock towards the downstream boundary.

Department of Aeronautics
The Royal Institute of Technology
S-100 44 Stockholm 70, Sweden

1. 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.

2. O. A. OLEINIK, "Discontinuous solutions of nonlinear differential equations," *Uspehi Mat. Nauk*, v. 12, 1957, pp. 3–73; English transl., *Amer. Math. Soc. Transl.* (2), v. 26, 1963, pp. 95–172.

3. H.-O. KREISS, "Difference approximations for initial-boundary value problems for hyperbolic differential equations," in *Numerical Solutions of Nonlinear Partial Differential Equations* (D. Greenspan, Ed.), Wiley, New York, 1966, pp. 114–166.

4. R. D. RICHTMEYER & K. W. MORTON, *Difference Methods for Initial Value Problems*, Interscience, New York, 1967.

5. R. W. MacCORMACK, *The Effect of Viscosity in Hypervelocity Impact Cratering*, AIAA Paper 69-354, 1969.

6. L. K. KARLSEN, *A Criterion for the Existence of Erroneous Modes Near Steady Shocks in Conservative Finite-Difference Computations*, Trans. Roy. Inst. Techn., TRITA-FPT-009, Stockholm, 1974.

7. R. F. WARMING & R. M. BEAM, "Upwind second-order difference schemes and applications in unsteady aerodynamic flows," *AIAA J.*, v. 14, 1976, pp. 1241–1249.