SOME HIGHER ORDER DIFFERENCE SCHEMES ENFORCING AN ENTROPY INEQUALITY

M. S. Mock

1. INTRODUCTION

Weak solutions of the initial value problem for hyperbolic systems of conservation laws,

$$u_t + f(u)_x = 0, \qquad -\infty < x < \infty, t > 0; u(x, 0) \text{ given}, \tag{1.1}$$

are in general not unique. An entropy condition is imposed to select the physically relevant weak solution. It is clearly desirable to have such an entropy condition reflected in numerical methods for this type of problem; *i.e.,* to know that the limit of approximate solutions, when it exists, is the physically relevant solution.

In this context, the best understood methods are the "monotone" difference schemes. For these schemes, such a result is obtained in [3] for the case that (1.1) is a single equation, with u,f scalar valued. For the special example of Friedrichs' scheme, similar results for systems are also known [6]. Monotone schemes are also attractive in that they admit discrete representations of shock waves, at least for single equations [4]. Unfortunately, they are limited to first order accuracy, and thus are of limited practical importance.

In this paper we discuss some higher order schemes which also enforce an entropy condition. Our main results are for a second order scheme of Lax-Wendroff type. In the following, we shall assume either that (1.1) is a single equation or a system of dimension m with f_u symmetric. Smooth solutions of such systems satisfy an additional conservation law,

$$U_t + F_x = 0, \tag{1.2}$$

in which U, F are scalar valued functions given by

$$U = u^2, \quad F = 2(u \cdot f - \Phi) \tag{1.3}$$

where Φ is a scalar valued function satisfying $\Phi_u = f$. Discontinuous solutions of (1.1) do not satisfy (1.2). We shall, however, require such solutions to satisfy an entropy condition of the form

$$U_t + F_x \leq 0 \tag{1.4}$$

in the sense of distributions. For systems which are strongly nonlinear in the sense of [10], this condition is equivalent to the classical entropy condition of

Received March 14, 1977. Revision received September 2, 1977.

Michigan Math. J. 25 (1978).

Lax [7]; for systems which are genuinely nonlinear in the sense of [7], the condition (1.4) is still an equivalent entropy condition provided that the discontinuities are sufficiently small [6]. For the case of a single equation, m = 1, both of these nonlinearity conditions reduce to the requirement that f be a convex function of u.

A brief outline of our discussion follows. In Section 2, we discuss possible forms for the parabolic regularization of (1.1) and discuss the relation between an entropy condition of the form (1.4) and stability of a numerical method. Our main result is presented in section 3; we show that the limits of the solutions of a second order implicit scheme of Lax-Wendroff type satisfy the entropy condition (1.4). In section 4, we consider stationary discontinuities for a single conservation law, as described by this type of scheme. These solutions are of special interest because unphysical discontinuities occasionally observed, as in the numerical experiments reported in [3], tend to be of this type. The analysis reveals two mechanisms by which such difficulties can occur in practice. In particular, we show that it is possible for unphysical discontinuities to be linearly stable with respect to the difference scheme considered. This type of difficulty can be anticipated somewhat more generally, in view of the following observation: consider the case of a scalar equation (1.1), with $u = v + w$, where v is a weak solution of (1.1) of the special form

$$v(x, t) = \begin{cases} u_+, & x - st > 0 \\ u_-, & x - st < 0. \end{cases} \tag{1.5}$$

$$f(u_+) - f(u_-) = s(u_+ - u_-)$$

and w is a smooth perturbation, small in L_2 and C^1. From (1.1), we have

$$\begin{aligned} 0 &= ((v + w)_t + f(v + w)_x, w) = (w_t, w) - (f_u(v)\, w, w_x) + O(w^3) \\ &= (w_t, w) + (1/2)\, w(st, t)^2 (f_u(u_+) - f_u(u_-)) + O(w^3) \end{aligned} \tag{1.6}$$

where (,) is the L_2 scalar product in x. From (1.6), we see that v is stable, in the linear sense, if

$$f_u(u_+) > f_u(u_-) \tag{1.7}$$

which implies that v is *not* a physically admissible weak solution [12]. Some semi-empirical results of this type are also contained in [3]. In section 5 we describe a third order scheme for single equations, the limits of the solutions of which also satisfy (1.4). Although this scheme contains an extra dissipation term of fourth order, the order of L_2 dissipation of the scheme, in regions where the solution is smooth, is six. Some numerical experiments are described in Section 6.

2. REGULARIZATION AND STABILITY

In the difference schemes that follow, h denotes the space increment and k the time increment, with $\lambda = k/h$. $X = X_{(h)}$ is the space of continuous, piecewise

linear functions (with values in $\mathbb{R}^m$) with possibly discontinuous slope at the mesh points $x_j = jh$, $j = 0, \pm 1, \pm 2, \ldots$. For $n = 0, 1, 2, \ldots$, $u^n \in X$ is our approximation to $u(\cdot, t_n)$, with $t_n = nk$. Below we use $(,)$ for the real L_2 inner product in x, and $\|\cdot\|$ for the corresponding L_2 norm; other norms in x are denoted by subscripts. Generic constants are denoted by c.

With this notation, the monotone difference schemes correspond essentially to a parabolic regularization of (1.1), replacing the right side by a term $c\frac{h}{\lambda}u_{xx}$; this identification becomes quite explicit if we consider a finite element form of these schemes, as done in [9], [10]. An upper bound of order unity is required on $\lambda|f_u|$ for stability of these schemes; a more restrictive bound on the same quantity is required for the enforcement of an entropy condition such as (1.4) [3], [6]. This is as expected; the entropy condition (1.4) may be viewed as a rather strong stability condition, and is imposed even where the solution (more properly, the variable coefficient $f_u(u)$ of the linearized problem) is not smooth.

In contrast, Lax-Wendroff schemes may be viewed as using the second order term in a Taylor expansion of (1.1) in time as the regularization; *i.e.* replacing the right side of (1.1) by $(h\lambda(f_u^2 u_x)_x)/2$ in the neighborhood of a discontinuity. This type of dissipation has markedly different properties from the one discussed previously; not so much because of the nonlinearity, but because of the dependence on λ. In particular, the amount of regularization now increases with increasing λ. In practice, the amount of regularization controls the amount of entropy generation in the neighborhood of a given shock. Insufficient entropy generation leads to "overshooting" and the possible propagation of unphysical discontinuities. Indeed, the proposition given below shows that many schemes of Lax-Wendroff type cannot be expected to describe shocks accurately for sufficiently small values of λ.

We introduce the following scheme, which is a second order approximation to (1.1) where the solution is smooth: given $u^n \in X$, find $u^{n+1} \in X$ such that

$$(2.1)\quad (u^{n+1} - u^n + kf(u^n)_x, \phi) + \frac{k^2}{2}(f_u^2(u^{n+1})u_x^{n+1}, \phi_x) = 0 \qquad \text{for all } \phi \in X.$$

The approximation to the solution of (1.1) which is obtained by successive application of (2.1) with increasing n is denoted by $u_{(h,k)}$. This scheme is discussed extensively in sections 3 and 4.

For a single equation (1.1), we can easily write (2.1) in terms of the discrete values $u_j^n = u^n(x_j)$,

$$(2.2)\quad \begin{aligned} &\frac{1}{6}(u_{j-1}^{n+1} + 4u_j^{n+1} + u_{j+1}^{n+1}) - \frac{\lambda^2}{2}(g(u_{j-1}^{n+1}) - 2g(u_j^{n+1}) + g(u_{j+1}^{n+1})) \\ &\quad = \frac{1}{6}(u_{j-1}^n + 4u_j^n + u_{j+1}^n) - \lambda\int_0^1 [f((1-\xi)u_j^n + \xi u_{j+1}^n) \\ &\quad - f((1-\xi)u_{j-1}^n + \xi u_j^n)]\,d\xi. \end{aligned}$$

where $g(u)$ satisfies $g_u = f_u^2$.

PROPOSITION. *Suppose the scheme* (2.1) *is applied to a Riemann problem, with initial data of the form*

$$u(x,0) = \begin{cases} u_+, & x > 0 \\ u_-, & x < 0 \end{cases}, \qquad u_+, u_- \text{ constant}, \tag{2.3}$$

and suppose f *is strongly nonlinear, in the sense of* [10]. *Suppose that the approximate solutions* $u_{(h,k)}$ *converge boundedly almost everywhere, to a limit function* $\tilde{u}(x,t)$, *as* $h,k \to 0$ *with* $k = o(h)$. *Assume in addition, that the variation of the discrete solutions satisfies*

$$\sup_n (\operatorname{var}_x u_{(h,k)}) \text{ bounded uniformly in } h, k. \tag{2.4}$$

Then the limit function $\tilde{u}$ *is continuous.*

Several remarks precede the proof. We do not anticipate actual computations with h, k refined in this manner, but rather with $\lambda = h/k$ fixed; our inference is simply that λ must not be chosen too small, since shock waves should be permitted in the solution of such problems.

The proposition is stated for the scheme (2.1) and the initial data (2.3) for simplicity; the result is somewhat more general. The proof relies essentially only on the consistency of the scheme with (1.1) and the fact that the entropy generation within compact regions of the x,t plane approaches zero as the mesh is refined in this manner, under the assumption (2.4). This is true rather generally for schemes of Lax-Wendroff type.

Proof of proposition. With the special form of initial data (2.3), the method (2.1) is homogeneous in h; *i.e.*, the values of $u_{(h,k)}$ depend on h, k only through the dependence on λ, and $u_{(h,k)}(x,t) = u_{(\alpha h,\alpha k)}(\alpha x, \alpha t)$ for all $\alpha > 0$. Because of this homogeneity, bounded convergence as $h, k \to 0$, $k = o(h)$ is equivalent to convergence as $k \to 0$ with h fixed, followed by convergence as $h \to 0$. Specifically, let $u_{(h)}$ be the continuous time Galerkin approximation to (1.1): for all $t > 0$, $u_{(h)}(\cdot,t) \in X$ and satisfies

$$(u_{(h),t} + f(u_{(h)})_x, \phi) = 0 \qquad \text{for all } \phi \in X. \tag{2.5}$$

Equation (2.5) defines an infinite system of ordinary differential equations for the $u_{(h)}(x_j, \cdot)$, $j = 0, \pm 1, \ldots$, as functions of time. We assume that $u_{(h,k)}(\cdot, 0)$ and $u_{(h)}(\cdot, 0)$ are chosen as reasonable approximations to the initial data (2.3). Since (2.1) is a consistent finite difference approximation to (2.5), and since the $u_{(h,k)}$ are uniformly bounded by assumption, it follows that $u_{(h,k)} \to u_{(h)}$ uniformly over compact regions of the (x,t) plane, as $\lambda \to 0$ with h fixed. The convergence of $u_{(h)}$ to $\tilde{u}$, boundedly almost everywhere as $h \to 0$, then follows from a triangle inequality,

$$\|u_{(h)} - \tilde{u}\|_{L_\infty} \le \|u_{(h)} - u_{(h,k)}\|_{L_\infty} + \|u_{(h,k)} - \tilde{u}\|_{L_\infty}; \tag{2.6}$$

in (2.6), we first choose λ sufficiently small to make the first right hand term small, and then h sufficiently small for the second term.

It follows from [8] that $\tilde{u}$ is a weak solution of (1.1). We claim that $U(\tilde{u})$ is a weak solution of (1.2). Let ζ be a nonnegative scalar valued C_0^∞ function of (x,t), and let ϕ in (2.5) be the piecewise linear interpolate, at each value of t, of $u_{(h)}(\cdot, t)\zeta(\cdot, t)$. In Lemma 3.4 below, it will be shown that $\eta = u_{(h)}\zeta - \phi$ is uniformly of $O(h)$. With this choice of ϕ, (2.5) becomes

$$\frac{1}{2}\left(U(u_{(h)})_t + F(u_{(h)})_x, \zeta\right) = -\left(u_{(h),t} + f(u_{(h)})_x, \eta\right). \tag{2.7}$$

Using (2.5) directly to estimate $u_{(h),t}$, it follows that the magnitude of the right side of (2.7) is less than $ch \sup_t (\operatorname{var}_x u_{(h)})$. Since $u_{(h)}$ must also satisfy (2.4), our claim follows by passing to the limit as $h \to 0$.

Now suppose there is a discontinuity in $\tilde{u}$, between two states u_ℓ and u_r. Denoting the speed of the discontinuity by s, we must then have

$$s(u_r - u_\ell) + f(u_\ell) - f(u_r) = 0 \quad \text{and}$$
$$s(U(u_r) - U(u_\ell)) + F(u_\ell) - F(u_r) = 0$$

simultaneously. For a strongly nonlinear f, as assumed, this is impossible, because of the equivalence of the entropy condition (1.4) with that of [7]. Since the latter condition excludes discontinuities with the "wrong" polarity, it follows that $s(U(u_r) - U(u_\ell)) + F(u_\ell) - F(u_r)$ cannot be zero. This completes the proof.

3. MAIN THEOREM

In the following, we assume that the initial data $u(\cdot,0)$ is of bounded total variation, approaching $u_+(u_-)$ as $x \to +\infty(-\infty)$. For L a multiple of h, $\psi_L \in X$ is given by

$$\psi_L(x) = \begin{cases} u_-, & x < -L, \\ u_- + (u_+ - u_-)(1 + x/L), & -L \le x \le L, \\ u_+, & x > L; \end{cases} \tag{3.1}$$

we also assume $u(\cdot,0) - \psi_L \in L_2$. Our main results for the scheme (2.1) are described by the following three theorems:

THEOREM 3.1. *Suppose* $u^n - \psi_L \in L_2$, *and suppose any one of the following conditions holds:*

(i) $m = 1$ (*i.e.,* (1.1) *is a single equation*);

(ii) $u_+ = u_-$;

(iii) $|f_u(u)| \le c(1 + |u|)$, *where* $|\cdot|$ *denotes vector and matrix norms on* $\mathbb{R}^m$;

then there exists $u^{n+1} \in X$ *satisfying* (2.1), *and* $u^{n+1} - \psi_L \in L_2$.

THEOREM 3.2. *In the case* $m = 1$, $u^n - \psi_L \in L_2$, *the solution* u^{n+1} *is unique and depends continuously* (*in* L_2) *on* u^n.

THEOREM 3.3. *Suppose the solutions of* (2.1) *converge boundedly almost everywhere, to a limit function* $\tilde{u}$, *as* $h,k \to 0$ *with* λ *fixed. Suppose in addition that the total variation of the discrete approximations satisfies*

$$h \sup_n (\operatorname{var}_x u^n) \to 0 \qquad \text{as } h \to 0. \tag{3.2}$$

Then the limit function $\tilde{u}$ *is a weak solution of* (1.1), *and satisfies*

$$U(\tilde{u})_t + F(\tilde{u})_x \leq 0 \tag{3.3}$$

in the sense of distributions, for U, F *given by* (1.3).

Theorem 3.3 is, of course, the entropy condition enforced by the scheme (2.1). In contrast to the results for conditionally stable schemes [3], [6], no bound on λ is needed. Also in contrast to the monotone schemes, (3.3) is not enforced at each time step, but only in the weaker sense

$$\int\int (U(\tilde{u})\,\zeta_t + F(\tilde{u})\,\zeta_x)\,dx\,dt \geq 0 \tag{3.4}$$

for any nonnegative C_0^∞ function ζ of (x,t).

The essential ingredient in the proofs of Theorems 3.1, 3.3 is an energy estimate for schemes of the type (2.1). Such estimates are obtained for linear problems, for similar schemes in [9]; the essential requirement on the scheme (2.1) is that we be able to obtain such an estimate without assuming smoothness of the variable coefficients $f_u(u)$ of the linearized problem.

Proof of Theorem 3.1. We will solve (2.1) in finite intervals in x, and then pass to the limit that the interval becomes all of $\mathbb{R}$. For nonnegative integer j, let X_j be the subspace of X such that the elements of X_j have compact support in $[-L - hj, L + hj]$, with L fixed. We consider the following problem: find $\omega_j \in X_j$ such that

$$(\omega_j + \psi_L - u^n + kf(u^n)_x, \phi) + \frac{k^2}{2}(f_u^2(\omega_j + \psi_L)(\omega_j + \psi_L)_x, \phi_x) = 0 \tag{3.5}$$

$$\text{for all } \phi \in X_j.$$

Our approximation to u^{n+1} is $\omega_j + \psi_L$, which satisfies boundary conditions at $\pm(L + hj)$. It suffices to obtain an a priori estimate of the form

$$\|\omega_j\| \leq K, \qquad K \text{ independent of } j. \tag{3.6}$$

The existence of ω_j satisfying (3.5) will then follow by an easy application of finite dimensional degree theory, for example by deforming k to zero. Then the sequence $\{\omega_j\}_{j=0}^\infty$ is bounded in $H = X \cap L_2(-\infty, \infty)$ and so has a weakly convergent subsequence, with limit $w \in H$ satisfying (3.6). But H is locally finite dimensional, so by an easy diagonalization argument we have $\omega_j \to \omega$ uniformly on compact

sets. From (3.5) it follows that $u^{n+1} = \omega + \psi_L$ satisfies (2.1) for all test functions ϕ with compact support. But X is of countable dimension, and such ϕ span X.

Choosing $\phi = \omega_j$ in (3.5), we obtain the following, after some algebraic manipulation and where $\delta = \omega_j + \psi_L - u^n$ and F is given by (1.3),

$$\begin{aligned}\|\omega_j\|^2 &= \|u^n - \psi_L\|^2 + k^2\|f_u(u^n)(u^n - \psi_L)_x\|^2 \\ &\quad - k(F(u_+) - F(u_-)) - k^2\|f_u(\omega_j + \psi_L)\omega_{j,x}\|^2 \\ &\quad - \|\delta + kf_u(u^n)(u^n - \psi_L)_x\|^2 - 2k(f_u(u^n)\psi_{L,x}, \omega_j) \\ &\quad - k^2(f_u^2(\omega_j + \psi_L)\psi_{L,x}, \omega_{j,x}).\end{aligned} \tag{3.7}$$

By virtue of the locally finite dimensional nature of the space X and the assumption that $u^n - \psi_L \in L_2$, it follows that u^n and u_x^n are uniformly bounded with respect to j (not h). Then (3.7) may be estimated using Schwarz,

$$\begin{aligned}\|\omega_j\|^2 \le c(L) + cL^{-1/2}\|\omega_j\| - k^2\|f_u(\omega_j + \psi_L)\omega_{j,x}\|^2 \\ + k^2|(f_u^2(\omega_j + \psi_L)\psi_{L,x}, \omega_{j,x})| \\ \le c(L) + k^2\|f_u(\omega_j + \psi_L)\psi_{L,x}\|^2\end{aligned} \tag{3.8}$$

In case (i), the equation (2.2) satisfies a maximum principle, so ω_j is uniformly bounded and the last term in (3.8) is $O(L^{-1})$. In case (ii), $\psi_{L,x} \equiv 0$. In case (iii), the last term in (3.8) is estimated by

$$(c/L^2)\int_{-L}^{L}(1 + \omega_j^2)\,dx \le c + (c/L^2)\|\omega_j\|^2;$$

choosing L sufficiently large, an estimate of the form (3.6) follows.

Proof of Theorem 3.2. As noted above (2.2) satisfies a maximum principle; without loss of generality, we may assume that f_u is uniformly bounded. Let b, v, w be elements of H, with values $b_j = b(x_j)$ etc. Since H is isomorphic to ℓ_2, let $T: \ell_2 \to \ell_2$ be given by

$$\begin{aligned}T(w)_j &= (w_{j-1} + 4w_j + w_{j+1})/6 \\ &\quad - \lambda^2[g(w_{j-1} + \psi_L(x_{j-1})) - 2g(w_j + \psi_L(x_j)) + g(w_{j+1} + \psi_L(x_{j+1}))]/2.\end{aligned} \tag{3.9}$$

The Frechet derivation of T is given by

$$\begin{aligned}(T'(w)v)_j &= (v_{j-1} + 4v_j + v_{j+1})/6 \\ &\quad - \lambda^2[f_u^2(w_{j-1} + \psi_L(x_{j-1}))v_{j-1} - 2f_u^2(w_j + \psi_L(x_j))v_j \\ &\quad + f_u^2(w_{j+1} + \psi_L(x_{j+1}))v_{j+1}]/2.\end{aligned} \tag{3.10}$$

We will show that as a mapping of $\ell_2 \to \ell_2$, $T'(w)$ is inverse bounded independently of w, so that by a global form of the implicit function theorem, *cf.* [13; p. 16], T is a homeomorphism of ℓ_2 onto ℓ_2. Indeed, if $T'(w)v = b$, setting

$$a_j = \lambda^2 f_u^2 (w_j + \psi_L(x_j))/2,$$

we have

$$2(a_j + 1/3)v_j = b_j + (a_{j-1} - 1/6)v_{j-1} + (a_{j+1} - 1/6)v_{j+1}, \qquad \text{for all } j. \tag{3.11}$$

Squaring both sides of (3.11) and summing over j, we obtain

$$\sum_j (a_j + 1/3)^2 v_j^2 \le ((1/4) + \varepsilon^{-1}) \sum_j b_j^2 + (1 + \varepsilon) \sum_j (a_j - 1/6)^2 v_j^2 \tag{3.12}$$

for any $\varepsilon > 0$. Since the a_j are nonnegative and bounded, we can choose ε sufficiently small that

$$(1 + \varepsilon) \sup_j \left(\frac{a_j - 1/6}{a_j - 1/3} \right)^2 < 1,$$

then

$$\sum_j v_j^2 \le c \sum_j b_j^2$$

is immediate from (3.12). This completes the proof of Theorem 3.2.

The proof of Theorem 3.3 requires the following:

Lemma 3.4. *For a sequence of values of* h *approaching zero, let* $v_h \in X_{(h)}$ *be bounded independently of* x, h, *and let* ζ *be a nonnegative scalar valued* C_0^∞ *function of* x. *Let* $\phi_h \in X_{(h)}$ *be the piecewise linear interpolate of the product* $v_h \zeta$, *and let* $\eta_h = v_h \zeta - \phi_h$. *Then* $|\eta_h|$ *is uniformly of* $O(h)$, *and* $|\eta_{h,x}|$ *is uniformly bounded. Furthermore,* $\eta_h = \gamma_h + \sigma_h$, *with* $|\sigma_h| \le O(h^{5/4})$, *and*

$$\int_{\Omega_h} \frac{\gamma_h^2(x)}{\zeta(x)}\, dx = O(h^{3/2}) \tag{3.13}$$

where Ω_h *is the support of* γ_h *and is contained in the support of* ζ.

Proof. By direct computation,

$$\eta_h(x) = \sum_j v_h(x_j)(\zeta(x) - \zeta(x_j))\, \Psi((x - x_j)/h) \tag{3.14}$$

where $\Psi(\xi) = 1 - |\xi|$, $|\xi| < 1$, and 0 otherwise. The bounds on $|\eta_h|$ and $|\eta_{h,x}|$ are immediate from (3.14). Let γ_h be given by the same sum (3.14), but restricted to those values of j for which $\inf_{x \in [x_{j-1}, x_{j+1}]} \zeta(x) \ge h^{1/2}$; the remaining terms in the sum give σ_h. Since ζ has compact support, independent of h, (3.13) follows. We claim that $\sup_{\zeta(x) < h^{1/2}} |\zeta_x(x)| = O(h^{1/4})$; otherwise, since $|\zeta_{xx}|$ is bounded, a partial

Taylor expansion shows that ζ changes sign, which is impossible. Then using (3.14), it follows that $|\sigma_h| \le O(h^{5/4})$.

Proof of Theorem 3.3. Let ζ denote a nonnegative C_0^∞ function of (x,t). In (2.1), set $\phi = 2u^{n+1}\zeta(\cdot, t_{n+1}) - \eta(\cdot, t_{n+1})$, the piecewise linear interpolate of $2u^{n+1}\zeta(\cdot, t_{n+1})$. With U, F given by (1.3) and $\delta = u^{n+1} - u^n$, we obtain after some manipulations,

$$
\begin{aligned}
&(U(u^{n+1}) - U(u^n) + kF(u^n)_x, \zeta(\cdot, t_{n+1})) \\
&\quad = (-\delta^2 - 2k\delta f(u^n)_x - k^2 (f_u^2(u^{n+1}) u_x^{n+1}) u_x^{n+1}, \zeta(\cdot, t_{n+1}) \\
&\qquad -k^2 (f_u^2(u^{n+1}) u_x^{n+1}, u^{n+1}\zeta(\cdot, t_{n+1})_x) + (\delta + k(f(u^n)_x, \eta(\cdot, t_{n+1})) \\
&\qquad + k^2 (f_u^2(u^{n+1}) u_x^{n+1}, \eta(\cdot, t_{n+1})_x)/2 \\
&\quad = -((\delta + kf(u^n)_x)^2, \zeta(\cdot, t_{n+1})) - k^2 ((f_u^2(u^{n+1}) u_x^{n+1}) u_x^{n+1} \\
&\qquad - (f_u^2(u^n) u_x^n) u_x^n, \zeta(\cdot, t_{n+1})) \\
&\qquad -k^2 (f_u^2(u^{n+1}) u_x^{n+1}, u^{n+1}\zeta(\cdot, t_{n+1})_x) + (\delta + kf(u^n)_x, \eta(\cdot, t_{n+1})) \\
&\qquad + k^2 (f_u^2(u^{n+1}) u_x^{n+1}, \eta(\cdot, t_{n+1})_x)/2.
\end{aligned} \tag{3.15}
$$

In (3.15) we use the boundedness of u^{n+1} and the bound on its variation (3.2), to get the third right hand term of $o(h)$; the last term is of this same order, using the boundedness of $\eta(\cdot, t_{n+1})_x$, Lemma 3.4. We use Lemma 3.4 to estimate the fourth term

$$
\begin{aligned}
&|(\delta + kf(u^n)_x, \gamma(\cdot, t_{n+1}) + \sigma(\cdot, t_{n+1})) \\
&\quad \le \varepsilon((\delta + kf(u^n)_x)^2, \zeta(\cdot, t_{n+1})) \\
&\quad + \frac{c}{\varepsilon} \int_{\Omega^{n+1}} \gamma^2(\cdot, t_{n+1}) \zeta^{-1}(\cdot, t_{n+1})\, dx + O(h^{5/4}) \\
&\quad \le \varepsilon((\delta + kf(u^n)_x)^2, \zeta(\cdot, t_{n+1})) + O(h^{5/4}).
\end{aligned} \tag{3.16}
$$

Using (3.16), (3.15) becomes

$$
\begin{aligned}
&(U(u^{n+1}) - U(u^n) + kF(u^n)_x, \zeta(\cdot, t_{n+1})) \\
&\quad \le -k^2 (((f_u^2(u^{n+1}) u_x^{n+1}) u_x^{n+1} - (f_u^2(u^n) u_x^n), \zeta(\cdot, t_{n+1})) + o(h)
\end{aligned} \tag{3.17}
$$

We next sum (3.17) over n, over the support (in time) of ζ; the right hand term is summed by parts, to give $k^2 \sum_n ((f_u^2(u^n) u_x^n) u_x^n, \zeta(\cdot, t_n) - \zeta(\cdot, t_{n+1}))$; using the smoothness of ζ in t, this term is of order

$$
\begin{aligned}
k^3 \sum_n \| f_u(u^n) u_x^n \|^2 &\le c\, k^2 \sum_n \operatorname*{var}_x (u^n) \\
&\le c\, k \sup_n (\operatorname*{var}_x (u^n)) \le o(1)
\end{aligned}
$$

using (3.2). Thus

$$k \sum_n ((U(u^{n+1}) - U(u^n))/k + F(u^n)_x, \zeta(\cdot, t_{n+1})) \leq o(1) \tag{3.18}$$

as $h,k \to 0$ from which (3.4) follows. It is shown in [8] that the limit function $\tilde{u}$ is a weak solution of (1.1); thus Theorem 3.3 is proved.

For an isolated discontinuity between two states u_ℓ (on the left) and u_r (on the right), (1.3, 1.4) and the Rankine-Hugoniot relation imply

$$\int_{u_r}^{u_\ell} f \cdot du \leq (1/2)(u_\ell - u_r) \cdot (f(u_\ell) + f(u_r)), \tag{3.19}$$

where the line integral is independent of path by the symmetry of f. For a single equation, (3.19) is a weaker requirement than the Oleinik condition [12].

4. STATIONARY DISCONTINUITIES

In this section we discuss the approximations to stationary discontinuities obtained from the scheme (2.1). We restrict attention to the case of a single equation, $m = 1$. Interest in such problems arises because the unphysical discontinuities which are occasionally observed with non-monotone schemes, such as those reported in [3] and below, frequently are stationary. Indeed, our analysis reveals two mechanisms by which this can occur, and in both cases suggests sufficiently large values of λ as a remedy.

We seek solutions of the discrete system

$$\begin{aligned} \lambda [g(u_{j+1}) - 2g(u_j) + g(u_{j-1})]/2 = \int_0^1 [f((1-\xi)u_j + \xi u_{j+1}) \\ - f(1-\xi)u_{j-1} + \xi u_j)]\, d\xi, \end{aligned} \tag{4.1}$$

satisfying

$$u_j \to u_+ (u_-) \qquad \text{as } j \to +\infty (-\infty), \tag{4.2}$$

where u_+, u_- satisfy $f(u_+) = f(u_-)$. Choosing the additive constant in f so that

$$f(u_+) = f(u_-) = 0, \tag{4.3}$$

we can simplify (4.1) to a one-step relation,

$$\frac{1}{q-p} \int_p^q f(u)\, du = \frac{\lambda}{2} \int_p^q f_u^2(u)\, du, \tag{4.4}$$

where $p = u_j$ and $q = u_{j+1}$ for any value of j.

Theorem 3.3 enforces an entropy condition on the solutions of (4.1, 4.2); from (3.18) and (4.3), we have

$$\int_{u_-}^{u_+} f(u)\,du \geq 0. \tag{4.5}$$

Our results on the existence of solutions of (4.1, 4.2) may be summarized by the following four theorems. In the cases where f is assumed convex, we also assume that the zeros of f_{uu}, if they exist, are isolated.

THEOREM 4.1. *For* f *convex, given any* $\lambda > 0$ *and any* $u_0 \in (u_+, u_-)$, *there exists a solution of* (4.1, 4.2), *assuming the value* u_0 *at* $x = 0$, *if and only if* (4.5) *is satisfied.*

THEOREM 4.2. *Suppose* f *is convex in the interval* $[u_+, u_-]$, *and suppose that*

$$\lambda \sup_{u \in [u_+, u_-]} |f_u(u)| < 1; \tag{4.6}$$

then for any $u_0 \in (u_+, u_-)$ *there are no solutions of* (4.1, 4.2) *bounded within the interval* $[u_+, u_-]$, *and assuming the value* u_0 *at* $x = 0$.

THEOREM 4.3. *For general smooth* f, *assume that* λ *is sufficiently large, depending on* u_+, u_-. *For every* u_0 *between* u_+ *and* u_-, *there exists a monotone solution of* (4.1, 4.2), *assuming the value* u_0 *at* $x = 0$, *if* $f_u(u_+)$ *and* $f_u(u_-)$ *are nonzero and the Oleinik condition,*

$$(u_+ - u_-) f(u) > 0 \qquad \textit{for all } u \textit{ between } u_+ \textit{ and } u_-, \tag{4.7}$$

is satisfied.

THEOREM 4.4. *Suppose that* u_+, u_-, λ *satisfy* (4.3), (4.5), *and*

$$\frac{1}{u_+ - u_-} \int_{u_-}^{u_+} f(u)\,du = \frac{\lambda}{2} \int_{u_-}^{u_+} f_u^2(u)\,du; \tag{4.8}$$

then a sufficient condition for the solution of (4.1, 4.2) *given by*

$$u_j = \begin{cases} u_+, & j > 0 \\ u_-, & j \leq 0 \end{cases} \tag{4.9}$$

to be stable (in the linearized sense), with respect to the scheme (2.1), *is*

$$f_u(u_+) \geq f_u(u_-), \tag{4.10a}$$

$$\lambda^2 [f_u^2(u_+) - f_u^2(u_-)]^2 \leq -4 f_u(u_+) f_u(u_-) [1 + \lambda f_u(u_+) - \lambda f_u(u_-)]. \tag{4.10b}$$

Several remarks precede the proofs.

For f convex, (4.5) and (4.7) are equivalent, and Theorem 4.1 states that discrete shock profiles exist exactly when they should. However, if λ is sufficiently small that (4.6) is satisfied, then overshooting generally occurs on at least one side of the discontinuity. The overshooting gets worse as λ is further reduced, and can lead to the propagation of unphysical discontinuities, for nonconvex f outside the interval $[u_+,u_-]$. Existence of a discrete shock profile may then fail, at least for some values of λ, in which case multiple discontinuities will be propagated numerically; this cannot be physically correct. We note that (4.6) is essentially the Courant stability condition for conditionally stable schemes, and is weaker than criteria usually employed in practice. In contrast, for sufficiently large λ, it is not necessary to have f convex, but only that contact discontinuities be avoided.

In general, solutions of (4.1, 4.2) of the special form (4.9) which satisfy (4.5) but not (4.7) are possible. As such discontinuities occasionally arise in computations [3], their stability is of considerable interest. The pair of equations (4.10) is a sufficient condition for their stability. Equation (4.10a) is very similar to (1.7), and is incompatible with the Oleinik condition (4.7), except for the case of a contact discontinuity, $f_u(u_+) = f_u(u_-) = 0$. In the case where the zeros of f are all simple, values of u_+,u_- for which (4.10) is satisfied have an even number of zeros of f between them.

In practice, the values of u_+,u_- are determined by the choice of λ, because of the requirement (4.8) for the existence of such a solution (4.9). Theorem 4.3 shows that the trouble can be avoided by choosing λ sufficiently large, but in general this will require an unconditionally stable scheme, such as (2.1).

In contrast, such stability cannot be expected for solutions of (4.1, 4.2) satisfying the Oleinik condition (4.7). For $f_u(u_+) < 0$ and $f_u(u_-) > 0$, the existence of marginal modes in the linearized scheme (4.14) is immediate. Such modes are expected from consideration of the translation properties of discrete shock profiles.

The proof of Theorem 4.1 requires only application of the intermediate value theorem to (4.4); we omit the details.

Proof of Theorem 4.2. Without loss of generality, we take $f_{uu} \geq 0$, $f < 0$ in (u_+,u_-), $f_u(\bar{u}) = 0$ where $u_+ < \bar{u} < u_-$. For $p \in (u_+,\bar{u}]$, it suffices to show that no value of $q \in [u_+,p)$ satisfies (4.4). Integrating (4.4) by parts gives

$$f(p) = (p-q)\int_0^1 \xi f_u(u(\xi))\,d\xi - \frac{\lambda}{2}\int_q^p f_u^2(u)\,du, \qquad \xi = \frac{u-q}{p-q},$$

$$= \frac{f(p)-f(q)}{2} + (p-q)\int_0^1 (\xi - 1/2) f_u(u(\xi))\,d\xi - \frac{\lambda}{2}\int_q^p f_u^2\,du$$

$$= -f(q) + 2(p-q)\int_0^{1/2} \tau\left[f_u\left(\frac{p+q}{2} + \tau(p-q)\right)\right.$$

$$\left. - f_u\left(\frac{p+q}{2} - \tau(p-q)\right)\right] d\tau$$

$$- \lambda \int_q^p f_u^2(u)\,du, \qquad \tau = \xi - 1/2,$$

$$> -f(q) - \lambda \int_q^p f_u^2\,du > f(p)$$

if (4.6) holds and $f(q) \in (f(p),0]$.

The only remaining possibility of a solution of (4.1, 4.2) bounded within $[u_+,u_-]$ is a special solution of the form (4.9), which can only happen if (4.8) holds. Such solutions do not assume the prescribed value u_0 at $x = 0$.

Proof of Theorem 4.3. Again we take $f < 0$ in (u_+,u_-), $u_+ < u_-$. For $p \in [u_+,u_-]$ we show the existence of $q \in (u_+,p)$ for λ sufficiently large. Integrating (4.4) by parts and using Schwarz, we obtain

$$\begin{aligned} f(q) + \frac{\lambda}{2} \int_q^p f_u^2(u)\,du &= \int_q^p f_u(u) \frac{p-u}{p-q}\,du \\ &\le \left(\frac{p-q}{3}\right)^{1/2} \left(\int_q^p f_u^2(u)\,du\right)^{1/2}. \end{aligned} \tag{4.11}$$

Application of the intermediate value theorem to (4.11) gives the existence of $q \in [u_+,p)$ if

$$\lambda^2 \int_{u_+}^p f_u^2(u)\,du \ge 4(p-u_+)/3,$$

so that a sufficient condition on λ is

$$\lambda^2 \inf_{u_+<p<u_-} \left(\int_{u_+}^p f_u^2(u)\,du\right) \Big/ (p-u_+) \ge 4/3 \tag{4.12}$$

plus a similar condition with u_+ replaced by u_-. Such λ exists provided that $f_u(u_+)$ and $f_u(u_-)$ are not zero.

The converse of this theorem also holds, with (4.7) replaced by the weaker statement

$$(u_+ - u_-)f(u) \ge 0 \qquad \text{for all } u \text{ between } u_+, u_-. \tag{4.13}$$

A partial Taylor expansion of (4.4) shows that p,q cannot both be arbitrarily close to u_+ (u_-) if $f_u(u_+)$ ($f_u(u_-)$) is zero. Thus instead of approaching the asymptotic value, a solution of (4.1, 4.2) in this case would have to achieve it exactly for some finite value of j, and u_0 could not be arbitrarily chosen.

Proof of Theorem 4.4. In (2.1), let $u^n = u + v^n$, where $u \in X$ is of the form (4.9), with (4.8) satisfied, and $v^n \in X \cap L_2$ is small. The linearized equation for v^{n+1} is then

$$(4.14)\qquad (v^{n+1} - v^n + k(f_u(u)v^n)_x, \phi) + k^2((f_u^2(u)v^{n+1})_x, \phi_x)/2 = 0$$
$$\text{for all } \phi \in X.$$

In the following, we use the homogeneity in h to simplify the notation by setting $h = 1$, $k = \lambda$; we will also use the abbreviations $y = v_0^{n+1}$, $z = v_1^{n+1}$, $\delta = v^{n+1} - v^n$. Choosing $\phi = v^{n+1}$ in (4.14), we obtain after some manipulations,

$$(4.15)\qquad \|v^{n+1}\|^2 - \|v^n\|^2 = -\|\delta\|^2 + 2\lambda(f_u(u)v^n, v_x^{n+1}) - \lambda^2((f_u^2(u)v^{n+1})_x, v_x^{n+1}).$$

We now use the fact that u is piecewise constant, except in the interval (0,1). We split the integrals in the right side of (4.15), into $\int_{-\infty}^{0} + \int_0^1 + \int_1^{\infty}$. For $\int_1^{\infty}$, we have

$$(4.16)\qquad \begin{aligned} &-\int_1^{\infty} \delta^2\,dx + 2\lambda f_u(u_+)\int_1^{\infty}(v^{n+1} - \delta)v_x^{n+1}\,dx - \lambda^2 f_u^2(u_+)\int_1^{\infty}(v_x^{n+1})^2\,dx \\ &= -\int_1^{\infty}(\delta + \lambda f_u(u_+)v_x^{n+1})^2\,dx - \lambda f_u(u_+)z^2. \end{aligned}$$

Similarly, $\int_{-\infty}^{0}$ gives a contribution

$$(4.17)\qquad \int_{-\infty}^{0} = -\int_{-\infty}^{0}(\delta + \lambda f_u(u_-)v_x^{n+1})^2\,dx + \lambda f_u(u_-)y^2.$$

In the interval (0,1), $u_x = u_+ - u_-$, v_x^n, δ_x, and $v_x^{n+1} = z - y$ are constants. Simply because δ is linear and δ_x constant, the first term in (4.15) gives

$$(4.18)\qquad \int_0^1 \delta^2\,dx \geq \left(\int_0^1 \delta_x^2\,dx\right)\Big/ 12.$$

The third term in (4.15) may be integrated to obtain

$$(4.19)\qquad -\lambda^2\int_0^1 (f_u^2(u)v^{n+1})_x v_x^{n+1}\,dx = -\lambda^2(z - y)(zf_u^2(u_+) - yf_u^2(u_-)).$$

The second term in (4.15) is

$$(4.20)\qquad \begin{aligned} 2\lambda\int_0^1 f_u(u)v^n v_x^{n+1}\,dx &= 2\lambda(z - y)\int_0^1 f_u(u(x))v^n(x)\,dx \\ &= -2\lambda((z - y)/(u_+ - u_-))(v_1^n - v_0^n)\int_0^1 f(u(x))\,dx, \\ &= -\lambda^2(z - y)(v_1^n - v_0^n)\int_0^1 f_u^2(u(x))\,dx, \end{aligned}$$

integrating by parts and then using (4.8). Combining (4.16–4.20), we can estimate (4.15) by

$$\begin{aligned}\|v^{n+1}\|^2 - \|v^n\|^2 \le \int_0^1 (-\delta_x^2/12 + A\,\delta_x(z-y) - A(z-y)^2)\,dx \\ - \lambda f_u(u_+)z^2 + \lambda f_u(u_-)y^2 - \lambda^2(z-y)(zf_u^2(u_+) - yf_u^2(u_-)).\end{aligned} \tag{4.21}$$

where

$$A = \lambda^2 \int_0^1 f_u^2(u(x))\,dx. \tag{4.22}$$

We next show that $A \le 1/3$, which implies that the integral in the right side of (4.21) is nonpositive. Integrating (4.8) by parts, using $f(u_+) = f(u_-) = 0$, and then Schwarz gives

$$\begin{aligned}\frac{\lambda}{2}\left|\int_{u_-}^{u_+} f_u^2(u)\,du\right| &= \left|\frac{1}{u_+ - u_-}\int_{u_-}^{u_+}\left(\frac{u_+ + u_-}{2} - u\right) f_u(u)\,du\right| \\ &\le \left|\frac{u_+ - u_-}{12}\right|^{1/2}\left|\int_{u_-}^{u_+} f_u^2(u)\,du\right|^{1/2},\end{aligned}$$

from which (4.22) follows easily.

The remainder of the right side of (4.21) is a homogeneous quadratic form in y,z; (4.10) is the sufficient condition that it is nonpositive definite. Thus the proof of Theorem 4.4 is complete.

5. A THIRD ORDER SCHEME

For a single equation (1.1), we discuss a third order scheme which satisfies an analog of Theorem 3.3. Let $Y = Y_{(h)}$ denote the Hermite cubic space; *i.e.* the space of piecewise cubic polynomials in x, with continuous first derivatives at the mesh points x_j [14]. For $\phi \in Y$, $\phi_{xx} \in L_\infty$. Given $u^n \in Y$, we obtain $u^{n+1} \in Y$ from

$$\begin{aligned}(u^{n+1} - u^n + kf(u^n)_x - k^2 g(u^n)_{xx}/2, \phi) + k^3(f_u^3(u^{n+1})u_x^{n+1}, \phi_{xx})/6 \\ + k^4(g(u^{n+1})_{xx}, (f_u^2(u^{n+1})\phi_x)_x)/8 = 0 \qquad \text{for all } \phi \in Y,\end{aligned} \tag{5.1}$$

where $g_u = f_u^2$ as above. For the Hermite cubics, the continuous time Galerkin approximation applied to linear symmetric first under hyperbolic systems gives accuracy $O(h^3)$ in L_2 [1]; in this sense we call the scheme (5.1) accurate to $O(h^3 + k^3)$ in regions where the solution is smooth. The last term in (5.1) is an additional source of artificial dissipation.

We shall show that this scheme enforces the following entropy condition:

THEOREM 5.1. *Suppose the solutions of* (5.1) *converge boundedly almost everywhere to a limit* $\tilde{u}$ *as* $h,k \to 0$ *with* $\lambda = k/h$ *fixed. Suppose in addition that*

(5.2) $\|u_x^n\|_{L_1}$ and $h\|u_{xx}^n\|_{L_1}$ *are bounded independently of* n, h, k.

Then

$$U(\tilde{u})_t + F(\tilde{u})_x \leq 0 \tag{5.3}$$

in the sense of distributions, where U, F *are given by* (1.3).

The proof requires the following lemma:

LEMMA 5.2. *For a sequence of values of* h *approaching zero, let* $v_h \in Y_{(h)}$ *satisfy*

$$\|v_h\|_{L_\infty} + \|v_{h,x}\|_{L_1} + h\|v_{h,xx}\|_{L_1} \leq c, \tag{5.4}$$

c *independent of* h. *Let* ζ *be a* C_0^∞ *function of* x, *and let* $\phi_h \in Y_{(h)}$ *interpolate the value and first derivative of* v_h ζ; *i.e.,*

$$v_h(x_j)\,\zeta(x_j) = \phi_h(x_j),\ v_h(x_j)\,\zeta_x(x_j) + v_{h,x}(x_j)\,\zeta(x_j) = \phi_{h,x}(x_j) \quad \text{for all } j. \tag{5.5}$$

Let $\eta_h = \phi_h - v_h\zeta$, *then* $\|\eta_h\|_{L_\infty} = O(h)$, $\|\eta_{h,x}\|_{L_\infty} = O(1)$, $\|\eta_{h,xx}\|_{L_\infty} = O(h^{-1})$; *furthermore,* $\eta_h = \gamma_h + \sigma_h$, *with* $\|\sigma_h\|_{L_\infty} = O(h^{5/4})$ *and* γ_h *satisfies* (3.13).

The proof is completely analogous to the proof of Lemma 3.4, and is therefore omitted. The proof of Theorem (5.1) is similar so that of Theorem 3.3; in (5.1), we choose $\phi = u^{n+1}\zeta(\cdot, t_{n+1}) + \eta^{n+1}$, obtaining after collecting terms, and several partial integrations,

$$\begin{aligned}
&(U(u^{n+1}) - U(u^n) + kF(u^n)_x, \zeta(\cdot, t_{n+1})) \\
&\quad = -(\delta^2 + 2k\delta f(u^n)_x - k^2\delta g(u^n)_{xx} + k^2(f(u^n)_x)^2 \\
&\quad + k^3 f_u^3(u^{n+1})\,u_x^{n+1}u_{xx}^{n+1}/3 + k^4(g(u^{n+1})_{xx})^2/4, \zeta(\cdot, t_{n+k})) \\
&\quad - (k^2F(u^n)_x/2 + 2k^3f_u^3(u^{n+1})(u_x^{n+1})^2/3, \zeta_x(\cdot, t_{n+1})) \\
&\quad - (k^3u^{n+1}f_u^3(u^{n+1})\,u_x^{n+1}/3, \zeta_{xx}(\cdot, t_{n+1})) \\
&\quad - (k^4g(u^{n+1})_{xx}/4, 2f(u^{n+1})(u^{n+1}f(u^{n+1}))_x\zeta_x(\cdot, t_{n+1}) \\
&\quad + f^2(u^{n+1})\,u^{n+1}\zeta_{xx}(\cdot, t_{n+1})) \\
&\quad - 2(\delta + kf(u^n)_x - k^2g(u^n)_{xx}/2, \eta^{n+1}) \\
&\quad + k^3(f_u^3(u^{n+1})\,u_x^{n+1}, \eta_{xx}^{n+1} - k^4(g(u^{n+1})_{xx}/4, (f_u^3(u^{n+1})\eta_x^{n+1})_x))/3
\end{aligned} \tag{5.6}$$

where $\delta = u^{n+1} - u^n$.

Integration by parts gives the following identity;

$$
\begin{aligned}
(f_u^3 u_x u_{xx}, \zeta) &= -3(f_u^2 f_{uu} u_x^3, \zeta)/2 - (f_u^3 u_x^2, \zeta_x)/2 \\
&= -((f_u^2 u_x)_x, f_u u_x \zeta) - (f_u^2 u_x, f_{uu} u_x^2 \zeta) - (f_u^3 u_x^2, \zeta_x) \\
&= -3((f_u^2 u_x)_x, f_u u_x \zeta) - 2(f_u^3 u_x^2, \zeta_x).
\end{aligned} \tag{5.7}
$$

Setting $u = u^{n+1}$, $\zeta = \zeta(\cdot, t_{n+1})$ in (5.7) and combining with (5.6), we obtain

$$
\begin{aligned}
&(U(u^{n+1}) - U(u^n) + kF(u^n)_x, \zeta(\cdot, t_{n+1})) \\
&\quad = -((\delta + kf(u^n)_x - k^2 g(u^n)_{xx}/2)^2, \zeta(\cdot, t_{n+1})) \\
&\quad - 2(\delta + kf(u^n)_x - k^2 g(u^n)_{xx}/2, \eta^{n+1}) \\
&\quad + k^3 (f(u^{n+1})_x g(u^{n+1})_{xx} - f(u^n)_x g(u^n)_{xx}, \zeta(\cdot, t_{n+1})) \\
&\quad - k^4 ((g(u^{n+1})_{xx})^2 - (g(u^n)_{xx})^2, \zeta(\cdot, t_{n+k}))/4 + O(h^2)
\end{aligned} \tag{5.8}
$$

using the boundedness of u^n, u^{n+1}, and the various bounds on the derivatives of u^n, u^{n+1}, η^{n+1} obtained from (5.4) and Lemma 5.2. In (5.8), we replace η^{n+1} by $\gamma^{n+1} + \sigma^{n+1}$ and proceed as in Theorem 3.3. The last two terms in (5.8) are summed by parts in time, as above, and we recover (3.18). Then Theorem 5.1 is proved.

In regions where the solution is smooth, the L_2 dissipation of this scheme is given by the first right hand term of (5.8). This will be of $O(h^6)$, even though a fourth order dissipation term was included in (5.1). This is quite different from the analogous results for third order finite element schemes for linear problems [9].

6. NUMERICAL EXPERIMENTS

A series of simple numerical experiments was conducted, in an attempt to obtain at least qualitative answers to several questions raised by the above analysis. One area of investigation was the possible extention of the results of Section 4 to include moving discontinuities. More specifically, we would like to understand the effect of motion of a discontinuity on overshooting, and to obtain generalizations of Theorems 4.2, 4.3. A second question in this area is whether motion allows contact discontinuities to be followed by schemes such as (2.1). It is also unclear whether the stable, unphysical discontinuities described by Theorem 4.4 have moving analogs.

A second area of investigation concerns the value of λ, and whether such higher order schemes really need to be implicit: how much trouble is there if $\lambda << \lambda_0$, where

$$
\lambda_0 \sup_{u \in (u_+, u_-)} \|f_u(u)\| = 1, \tag{6.1}
$$

and whether there is any discernible advantage in taking $\lambda > \lambda_0$.

Our experiments were confined to single equations (1.1), using different forms of the flux function f, and involved three different schemes: the scheme described by (2.1, 2.2); an explicit form of this scheme, given by

$$
\begin{aligned}
u_j^{n+1} = u_j^n - \lambda \int_0^1 [f((1-\xi)u_j^n + \xi u_{j+1}^n) - f((1-\xi)u_{j-1}^n + \xi u_j^n)]\, d\xi \\
+ \lambda^2 [g(u_{j-1}^n) - 2g(u_j^n) + g(u_{j+1}^n)]/2;
\end{aligned} \tag{6.2}
$$

and a modified two-step Lax-Richtmyer scheme,

$$
\begin{aligned}
u_j^{n+1} &= (1-6\theta)u_j^n + 4\theta(u_{j-1}^n + u_{j+1}^n) - \theta(u_{j-2}^n + u_{j+2}^n) - \lambda[f(v_j^n) - f(v_{j-1}^n)] \\
v_j^n &= (u_j^n + u_{j+1}^n)/2 - \lambda[f(u_{j+1}^n) - f(u_j^n)]/2,
\end{aligned} \tag{6.3}
$$

in which, typically, $\theta = 0.1$.

The scheme (6.2) is obtained from (2.2) by lumping the mass matrices, at both time levels, and moving the second order term to the backward time level. The lumping of the mass matrices causes no difficulty; indeed, it is a stabilizing mechanism [9], and does not affect any of the results of section 3. However, we have not obtained a version of Theorem 3.3 with the second order term at the backward time level. In particular, the assumption of an upper bound on λ may not be sufficient for this purpose.

It is possible that explicit, conditionally stable schemes of higher order accuracy can be made compatible with entropy inequalities by using higher order regularization, *e.g.* $-c(h^3/\lambda)u_{xxxx}$ in the right side of (1.1). An example of an explicit second order scheme based on this form of regularization is given by (6.3), with $\theta > 0$. The ordinary Lax-Richtmyer scheme, corresponding to $\theta = 0$ in (6.3), is completely incapable of describing stationary discontinuities. For θ positive and λ sufficiently small, the results of [2], [5], [11] suggest that this scheme will approximate discontinuities if the Oleinik condition is satisfied, and furthermore be less sensitive to contact discontinuities than (2.2) or (6.2).

Our experiments utilized initial data of the form (2.3), corresponding to a Riemann problem. For such data, convergence as $h,k \to 0$ with λ fixed is immediate by homogeneity, and Theorem 3.3 can easily be applied. For each time step, a value of λ was obtained from a relation of the form

$$
\lambda_n \sup_x |f_u(u^n(x))| = \beta \tag{6.4}
$$

with $\beta < 1$ for the conditionally stable schemes.

Our results were not very surprising, and we will discuss them only qualitatively. For $\beta < 1$ in (6.4), the schemes (2.2) and (6.2) gave almost identical results, and no advantage of the implicit scheme (2.2) was observed. However, in some cases better results could be obtained with the implicit scheme by using $\beta > 1$.

In all of our calculations involving moving shocks (as opposed to contact discontinuities), overshooting occurred on one side of the discontinuity for the schemes (2.2), (6.2), for all values of β employed. Conditions of the form (4.12) are not sufficient to get monotone profiles for moving shocks. However, the overshooting could typically be reduced to roughly one percent of the discontinuity by using λ satisfying (4.12), in the scheme (2.2). For the schemes (2.2), (6.2),

the overshooting gets worse as λ is reduced, as expected. The scheme (6.3) overshoots on both sides of the discontinuity, and the amount of overshooting is relatively insensitive to the value of λ.

For convex f, no unphysical solutions were observed with any of these schemes. This result is in contrast with some experiments reported in [3]; we are inclined to attribute the differences to somewhat different approximation of the spatial derivatives.

For nonconvex f, sufficiently large overshooting can cause a discontinuity, which should be propagaged intact, to be split into two or three discontinuities, each of which is propagated separately. In all of our experiments, only one of each group of discontinuities so obtained failed to satisfy the Oleinik condition. Whether such a splitting will occur depends strongly on the values of f outside the interval of the original discontinuity. Cases were observed in which a discontinuity was propagated correctly only if the value of λ was sufficiently large, using the scheme (2.2).

Unphysical solutions of the special form (4.9) were observed; they were always stationary. Their appearance could be caused by overshooting, as described above, or by the failure of a computation to properly split up a given initial discontinuity.

We did not observe either of the schemes (2.2), (6.2) correctly propagating a contact discontinuity, moving or stationary. The scheme (6.3) was clearly superior in this respect, although it is somewhat unpredictable by virtue of an overshooting tendency which is not easily controlled in practice.

REFERENCES

1. T. Dupont, *Galerkin methods for first order hyperbolics: an example.* SIAM J. Numer. Anal. 10 (1973), 890–899.
2. I. M. Gel'fand, *Some problems in the theory of quasilinear equations.* Amer. Math. Soc. Transl. Ser. 2, No. 29 (1963), 295–381.
3. A. Harten, J. M. Hyman and P. D. Lax, *On finite-difference approximations and entropy conditions for shocks.* Comm. Pure Appl. Math. 29 (1976), 297–322.
4. G. Jennings, *Discrete shocks.* Comm. Pure Appl. Math. 27 (1974), 25–37.
5. N. Kopell and L. N. Howard, *Bifurcations and trajectories joining critical points.* Advances in Math. 18 (1975), no. 3, 306–358.
6. P. D. Lax, *Shock waves and entropy.* Contributions to nonlinear functional analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971), pp. 603–634. Academic Press, New York, 1971.
7. ———, *Hyperbolic systems of conservation laws.* II. Comm. Pure Appl. Math. 10 (1957), 537–566.
8. P. D. Lax and B. Wendroff, *Systems of conservation laws.* Comm. Pure Appl. Math. 13 (1960), 217–237.
9. M. S. Mock, *Explicit finite element schemes for first order symmetric hyperbolic systems.* Numer. Math. 26 (1976), 367–378.

10. ———, *Discrete shocks and genuine nonlinearity,* Michigan Math. J.

11. ———, *On fourth order dissipation and single conservation laws.* Comm. Pure Appl. Math. 29 (1976), no. 4, 383–388.

12. O. A. Oleinik, *Uniqueness and stability of the generalized solution of the Cauchy problem for a quasilinear equation.* Amer. Math. Soc. Transl. Ser. 2, No. 33 (1963), 285–290.

13. J. T. Schwartz, *Nonlinear functional analysis.* Gordon and Breach, New York, 1969.

14. G. Strang and G. Fix, *An analysis of the finite element method.* Prentice-Hall, Englewood Cliffs, N.J., 1973.

Department of Mathematics
Rutgers University
New Brunswick, New Jersey 08903