

# A Finite Difference Scheme for a System of Two Conservation Laws with Artificial Viscosity\*

By David Hoff

**Abstract.** In this paper we analyze an implicit finite difference scheme for the mixed initial-value Dirichlet problem for a system of two conservation laws with artificial viscosity. The system we consider is a model for isentropic flow in one space dimension. First, we show that, under certain conditions on the mesh, the scheme is stable in the sense that it possesses an invariant set (defined by the so-called Riemann invariants). We obtain this result as an extension of the same stability theorem for the Lax-Friedrichs scheme in the inviscid case. Second, we show that the approximants remain bounded and, in fact, decay to the boundary values as  $t \rightarrow \infty$ . Finally, we obtain two  $O(\Delta x^2)$  error bounds; the first grows exponentially in time while the second, which requires that the data have small oscillation, is independent of time.

**1. Introduction.** In this paper we analyze certain finite difference schemes for the system of equations

$$(1.1) \quad \begin{aligned} u_t + p(v)_x &= du_{xx}, \\ v_t - u_x &= dv_{xx} \quad \text{for } (x, t) \in (0, L) \times (0, \infty) \end{aligned}$$

with data

$$(1.2) \quad u(x, 0) = u_0(x), \quad v(x, 0) = v_0(x)$$

and boundary conditions

$$(1.3) \quad \begin{aligned} u(x, t) &= 0, \\ v(x, t) &= v_b > 0 \quad \text{for } x = 0, L. \end{aligned}$$

If  $d = 0$ , then (1.1) describes one-dimensional isentropic gas motion in which  $u$  is the velocity,  $1/v$  is the density, and  $p(v)$  is the pressure. Putting  $d > 0$  has the effect of artificially smoothing out discontinuities. Except for the first result of Section 2, we shall require that  $d > 0$ .

$p(v)$  is to be  $C^2$  and positive for  $v$  positive and is assumed to satisfy

$$(1.4) \quad p' < 0, \quad p'' > 0 \quad \text{and} \quad \lim_{v \rightarrow \infty} v^3 |p'(v)| = \infty.$$

---

Received July 28, 1978.

AMS (MOS) subject classifications (1970). Primary 65M10, 65M15.

\*This research was performed while the author was visiting the Courant Institute of Mathematical Sciences, New York University.

For example, we could take  $p(v) = kv^{-\gamma}$ , where  $k > 0$  and  $\gamma < 2$ .

Denote the so-called Riemann invariants by

$$(1.5) \quad \begin{aligned} r_c(u, v) &= u + \int_c^v \sqrt{-p'(s)} \, ds \quad \text{and} \\ s_c(u, v) &= u - \int_c^v \sqrt{-p'(s)} \, ds. \end{aligned}$$

Then the set

$$(1.6) \quad S_c \equiv \{(u, v) : s_c(u, v) \leq 0 \leq r_c(u, v)\}$$

is *invariant* for (1.1)–(1.3). This means that if the data and boundary values are in  $S_c$  for all  $x$ , then any smooth solution is also in  $S_c$  for all  $x$  and  $t$ , as long as it is defined. For a proof see [1, p. 385].

Notice that, by (1.4),  $S_c$  is convex. Also, if  $c' < c$ , then  $S_{c'} \subseteq \text{interior}(S_c)$ . See Figure 1.

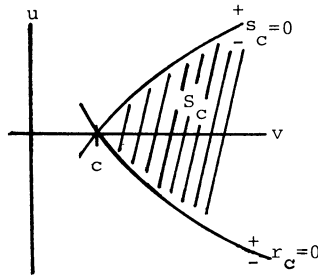


FIGURE 1

So for appropriate data and boundary values,  $v(x, t)$  will be bounded below and the nonlinear term  $p(v)$  will be under control. If we assume in addition that  $d > 0$ , then the necessary a priori bounds on the solution can be derived which can be used to show that a smooth solution of (1.1)–(1.3) exists for all time. See [2].

We therefore take  $u_0$  and  $v_0$  to be smooth functions satisfying the boundary conditions (1.3) such that

$$\begin{bmatrix} u_0(x) \\ v_0(x) \end{bmatrix} \in S_c$$

for all  $x$  and for some  $c > 0$ , which will be fixed throughout. It follows then that  $\begin{bmatrix} 0 \\ v_b \end{bmatrix} \in S_c$  also.

Now let

$$(1.7) \quad U = \begin{bmatrix} u \\ v \end{bmatrix}, \quad F(U) = \begin{bmatrix} p(v) - p(v_b) \\ -u \end{bmatrix}, \quad U_b = \begin{bmatrix} 0 \\ v_b \end{bmatrix} \quad \text{and} \quad U_0 = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}.$$

Then (1.1)–(1.3) can be rewritten in the form

$$\begin{aligned}
 U_t &= dU_{xx} - F(U)_x, \\
 U(x, 0) &= U_0(x), \\
 (1.8) \quad U(x, t) = U_b &= \begin{bmatrix} 0 \\ \vdots \\ v_b \end{bmatrix} \text{ for } x = 0, L.
 \end{aligned}$$

We construct a finite difference scheme as follows. Let  $x_k = k\Delta x$  for  $k = 0, \dots, N + 1$ , where

$$(1.9) \quad (N + 1)\Delta x = L.$$

And let  $t_n = n\Delta t$  for  $n = 0, 1, \dots$ . We shall approximate the solution

$$\begin{bmatrix} U(x_1, t_n) \\ \vdots \\ U(x_N, t_n) \end{bmatrix} \text{ by the vector } U^n \equiv \begin{bmatrix} U_1^n \\ \vdots \\ U_N^n \end{bmatrix} \in R^{2N},$$

where, in analogy with (1.8),  $U^n$  is computed from the finite difference scheme

$$(1.10) \quad \frac{U^n - U^{n-1}}{\Delta t} = L(U^n)$$

with  $L(U)$  being defined by

$$(1.11) \quad L(U)_k = \frac{d}{\Delta x^2} (U_{k+1} - 2U_k + U_{k-1}) + \frac{1}{2\Delta x} [F(U_{k-1}) - F(U_{k+1})].$$

We make the convention that  $U_0 = U_{N+1} = U_b$  so that  $F(U_0) = F(U_{N+1}) = 0$ . Notice that  $L(U)$  is defined only if each  $v$ -component of  $U$  is positive. For our purposes it is sufficient to regard  $L$  as mapping  $S_c^N$  into  $R^{2N}$ .

In Section 2 we prove that, under appropriate mesh conditions, (1.10) has a unique solution  $U^n \in S_c^N$  provided that  $U^{n-1} \in S_c^N$ . Thus, the approximants exist for all  $n$ . In the course of doing this, we also establish the invariance of  $S_c^N$  for the following auxiliary schemes (in reverse order):

$$(1.12) \quad \frac{d}{dt} U(t) = L(U) \quad (\text{semidiscrete});$$

$$(1.13) \quad \frac{U^n - U^{n-1}}{\Delta t} = L(U^{n-1}) \quad (\text{explicit});$$

and for the conservation-law case  $d = 0$ ,

$$(1.14) \quad \frac{U_k^n - \frac{1}{2}(U_{k+1}^{n-1} + U_{k-1}^{n-1})}{\Delta t} = L(U^{n-1})_k = \frac{F(U_{k-1}^{n-1}) - F(U_{k+1}^{n-1})}{2\Delta x}.$$

Also in Section 2 we discuss briefly a practical method for computing the solution of (1.10).

In Section 3 we imitate the energy estimates of [2] and show that the solution

of (1.10) decays to  $U_b$  uniformly in  $x$ , exponentially in time. In Section 4 we derive two  $O(\Delta x^2)$  error estimates in the  $L^2$ -norm: the first grows exponentially in time and the second, proved under the hypothesis that the set  $\{v_0(x): x \in [0, L]\}$  has small diameter, is independent of time. The result of Section 3 then shows that the second error estimate eventually becomes applicable.

**2. The Invariance of  $S_c^N$ .** In this section we prove that, under appropriate conditions on the mesh,  $S_c^N$  is invariant for each of the schemes (1.10), (1.12)–(1.14). This result is most easily established for the explicit scheme (1.14), in which the computation of  $U_k^n$  involves the values of  $U^{n-1}$  at only two nodes. Fortunately, the invariance of  $S_c^N$  for (1.14) will lead in a natural way to the same invariance for the other schemes.

We denote the mesh ratios by

$$(2.1) \quad \alpha = \frac{\Delta t}{\Delta x} \quad \text{and} \quad \beta = \frac{\Delta t}{\Delta x^2}.$$

**THEOREM 2.1.** *If  $\alpha \leq 1/\sqrt{-p'(c)}$ , then  $S_c^N$  is invariant for (1.14).*

*Proof.* We introduce the function  $\bar{U}: S_c^2 \rightarrow R^2$  by

$$\bar{U}(U_1, U_2) = \frac{1}{2}(U_1 + U_2) + \frac{\alpha}{2}[F(U_1) - F(U_2)].$$

More specifically, letting  $U_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix}$  and  $U = \begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix}$ ,

$$(2.2) \quad \bar{u}(U_1, U_2) = \frac{1}{2}(u_1 + u_2) + \frac{\alpha}{2}[p(v_1) - p(v_2)],$$

$$\bar{v}(U_1, U_2) = \frac{1}{2}(v_1 + v_2) + \frac{\alpha}{2}(u_2 - u_1).$$

The scheme (1.14) may therefore be written

$$U_k^n = \bar{U}(U_{k-1}^{n-1}, U_{k+1}^{n-1});$$

and the assertion of the theorem is that  $U_1, U_2 \in S_c$  implies  $\bar{U}(U_1, U_2) \in S_c$ .

By way of preliminaries, we have from (1.4) and our hypothesis on  $\alpha$  that

$$(2.3) \quad \alpha\sqrt{-p'(v)} \leq 1 \quad \text{for } c \leq v,$$

with equality only if  $v = c$ . Also by (1.4),

$$(2.4) \quad v_1 \leq v_2 \quad \text{implies} \quad \sqrt{-p'(v_1)} \geq \sqrt{-p'(v_2)}.$$

Finally, we define

$$(2.5) \quad g(v) = \int_c^v \sqrt{-p'(s)} \, ds$$

so that

$$S_c = \left\{ \begin{bmatrix} u \\ v \end{bmatrix} : -g(v) \leq u \leq g(v) \right\}$$

(see (1.5)–(1.6)).

The general case of the theorem will follow easily from the special case that  $U_1, U_2 \in \partial S_c$ . This special case is proved in Steps I–IV below, and is extended to the general case in Steps V and VI.

*Step I.* The theorem is true if  $u_i = g(v_i), i = 1, 2$ .

To prove this, fix  $U_1 = [g(v_1)]$  and allow  $U_2 = [g(v)]$  to vary as  $v$  increases from  $c$ . Then  $\bar{U}(U_1, U_2)$  is the curve  $\bar{U}(v)$  in  $R^2$  given by

$$(2.6) \quad \begin{aligned} \bar{u}(v) &= \frac{1}{2} [g(v_1) + g(v)] + \frac{\alpha}{2} [p(v_1) - p(v)], \\ \bar{v}(v) &= \frac{1}{2} (v_1 + v) + \frac{\alpha}{2} [g(v) - g(v_1)]. \end{aligned}$$

We need to show that  $c \leq v$  implies  $\bar{U}(v) \in S_c$ . This will follow from the following four facts: (See Figure 2.)

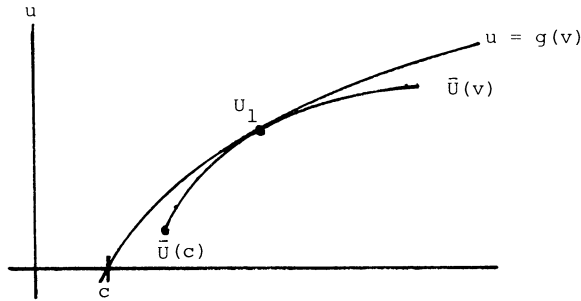


FIGURE 2

- (a)  $U(c) \in S_c$ ;
- (b)  $c \leq v \leq v_1$  implies that  $\bar{U}(v)$  is rising more rapidly than the curve  $u = g(v)$  at the same value of  $v$ ;
- (c)  $\bar{U}(v_1) = U_1$ ;
- (d)  $v_1 \leq v$  implies that  $\bar{U}(v)$  is rising less rapidly than the curve  $u = g(v)$  at the same value of  $v$ .

*Proof of (b).* Compute from (2.5) and (2.6)

$$\frac{d\bar{u}}{dv} = \frac{1}{2} g'(v) - \frac{\alpha}{2} p'(v) = \sqrt{-p'(v)} \left( \frac{1}{2} + \frac{\alpha}{2} \sqrt{-p'(v)} \right) \quad \text{and}$$

$$\frac{d\bar{v}}{dv} = \frac{1}{2} + \frac{\alpha}{2} \sqrt{-p'(v)},$$

so that

$$(2.7) \quad \frac{d\bar{u}}{d\bar{v}}(\bar{U}(v)) = \sqrt{-p'(\bar{v})}.$$

(b) is proved by comparing the number in (2.7) with the slope of the curve  $u = g(v)$  at  $\bar{v}$ , which by (2.5) is  $\sqrt{-p'(\bar{v})}$ . (b) thus asserts that  $\sqrt{-p'(v)} \geq \sqrt{-p'(\bar{v})}$ , and this will follow from (2.4) if we can show that

$$(2.8) \quad v \leq \bar{v}(v).$$

From (2.6) we have

$$\begin{aligned} \bar{v}(v) &= \frac{1}{2}(v_1 + v) + \frac{\alpha}{2} \int_{v_1}^v \sqrt{-p'(s)} \, ds \\ &= \frac{1}{2}(v_1 + v) + \frac{\alpha}{2} \sqrt{-p'(\xi)}(v - v_1) \geq \frac{1}{2}(v_1 + v) + \frac{1}{2}(v - v_1) = v \end{aligned}$$

by (2.3) and the hypothesis of (b) that  $v - v_1 \leq 0$ . This establishes (2.8) and proves (b).

*Proof of (a).* Knowing that (b) is true, we need only show that  $\bar{U}(c)$  is not below the curve  $u = -g(v)$ . Compute from (2.6)

$$\begin{aligned} \bar{u}(c) &= \frac{1}{2}g(v_1) + \frac{\alpha}{2} \int_c^{v_1} p'(s) \, ds \\ &= \frac{1}{2}g(v_1) - \frac{\alpha}{2} \int_c^{v_1} \sqrt{-p'(s)} \sqrt{-p'(s)} \, ds \\ &= \left( \frac{1}{2} - \frac{\alpha}{2} \sqrt{-p'(\xi)} \right) \int_c^{v_1} \sqrt{-p'(s)} \, ds \geq 0 \end{aligned}$$

by (2.3). And from (2.8),  $c \leq \bar{v}(c)$ . Thus,  $\bar{U}(c) \in S_c$  (see Figure 2).

(c) is obvious, and (d) is proved just like (b) but with certain inequalities reversed.

*Step II.* The theorem is true if  $u_1 = g(v_1)$  and  $u_2 = -g(v_2)$ .

Again we fix  $U_1 = [g(v_1)]$  and vary  $U_2 = [-g(v)]$  for  $c \leq v$  to obtain a curve  $\bar{U}(U_1, U_2) = \bar{U}(v)$  given by

$$\begin{aligned} \bar{v}(v) &= \frac{1}{2}(v_1 + v) - \frac{\alpha}{2} [g(v) + g(v_1)], \\ \bar{u}(v) &= \frac{1}{2} [g(v_1) - g(v)] + \frac{\alpha}{2} [p(v_1) - p(v)]. \end{aligned} \tag{2.9}$$

Notice that  $\bar{U}(c)$  in this case coincides with the  $\bar{U}(c)$  of Step I and is, therefore, in  $S_c$ . Step II will follow by establishing the existence of  $v^* \geq c$  such that

- (a)  $c \leq v \leq v^*$  implies  $d\bar{u}(\bar{U}(v))/d\bar{v} \leq -g'(\bar{v}(v))$ ;
- (b)  $\bar{U}(v^*) \in S_c$ ;
- (c)  $v^* \leq v$  implies  $-g'(\bar{v}(v)) \leq d\bar{u}(\bar{U}(v))/d\bar{v} < 0$ . (See Figure 3.)

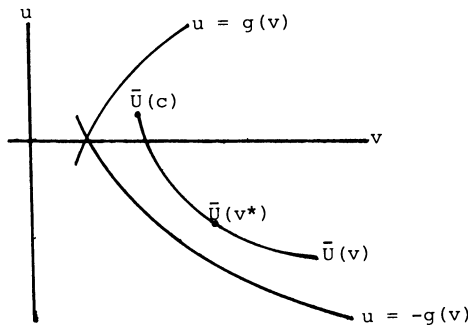


FIGURE 3

Let  $f(v) = v + \alpha g(v)$ . Then  $v^*$  is defined as the unique solution of

$$(2.10) \quad f(v^*) = v_1 - \alpha g(v_1).$$

To see that there is such a  $v^*$  notice that  $f'(v) > 0$ ,  $f(c) = c$ ,  $f(\infty) = \infty$ , and from (2.5) and (2.3)

$$v_1 - \alpha g(v_1) = v_1 - \alpha \sqrt{-p'(\xi)} (v_1 - c) \geq v_1 - (v_1 - c) = c.$$

Just as in Step I we compute

$$(2.11) \quad \frac{d\bar{u}}{dv}(\bar{U}(v)) = -\sqrt{-p'(v)}.$$

*Proof of (a).* (a) therefore asserts that  $\sqrt{-p'(v)} \geq \sqrt{-p'(\bar{v}(v))}$  for  $v \leq v^*$ . This inequality will follow from (2.4) if we can show that

$$(2.12) \quad v \leq \bar{v}(v).$$

To do this notice that  $v \leq v^*$  implies  $f(v) \leq f(v^*)$  so that  $v + \alpha g(v) \leq v_1 - \alpha g(v_1)$ . Halve this inequality and add  $\frac{1}{2}v$  to obtain

$$v \leq \frac{1}{2}(v + v_1) - \frac{\alpha}{2} [g(v_1) + g(v)] = \bar{v}(v).$$

This establishes (2.12) and proves (a). The proof of (c) is similar.

*Proof of (b).* Since (a) is true and since  $\bar{U}(c) \in S_c$ , we need only show that  $\bar{U}(v^*)$  is above the curve  $u = -g(u)$ . That is, we need to prove that  $\bar{u}(v^*) \geq -g(\bar{v}(v^*))$ . It follows easily from (2.9) and (2.10) that  $\bar{v}(v^*) = v^*$ . The requirement then is that

$$(2.13) \quad \bar{u}(v^*) \geq -g(v^*).$$

To prove this, substitute

$$(2.14) \quad g(v_1) = -g(v^*) + \frac{1}{\alpha} (v_1 - v^*)$$

(this is (2.10)) into (2.9) to obtain

$$\begin{aligned} \bar{u}(v^*) &= -g(v^*) + \frac{1}{2\alpha} [(v_1 - v^*) + \alpha^2(p(v_1) - p(v^*))] \\ &= -g(v^*) + \frac{1}{2\alpha} [1 + \alpha^2 p'(\xi)] (v_1 - v^*) \geq -g(v^*) \end{aligned}$$

by (2.3) and the fact that  $v_1 \geq v^*$ , which is obvious from (2.14). Thus, (2.13) holds, (b) is proved, and Step II is completed.

*Steps III and IV.* These are the same as Steps I and II but with  $g$  replaced by  $-g$  in the hypotheses. The proofs are similar and so are omitted.

*Step V.* The theorem is true if  $U_1 \in \partial S_c$ .

To prove this, fix  $U_1 \in \partial S_c$  and let  $U_2 = [v_2^u]$  with  $u$  varying between  $-g(v_2)$  and  $g(v_2)$  to obtain a curve  $\bar{U}(U_1, U_2) = \bar{U}(u)$ . From (2.2) this curve is a line whose endpoints, by Steps I–IV, are in  $S_c$ . The whole line is in  $S_c$  because  $S_c$  is convex.

*Step VI. The general case.* Fix  $U_2 \in S_c$  and let  $U_1 = [v_1^u]$  with  $u$  varying

between  $-g(v_1)$  and  $g(v_1)$ . Again the resulting curve  $\bar{U}(u) = \bar{U}(U_1, U_2)$  is a line whose endpoints, by Step V, are in  $S_c$ . Hence, so is the whole line.  $\square$

Analogous one-sided estimates for the scheme (1.14) for more general conservation systems are derived in [3]. There the condition on  $\alpha$  is more stringent than that of Theorem 2.1, which is the usual Courant-Friedrichs-Lewy condition.

We turn now to the other schemes discussed in Section 1, assuming that  $d > 0$ .

**COROLLARY 1.** *If  $\beta \leq 1/2d$  and  $\Delta x \leq 2d/\sqrt{-p'(c)}$ , then  $S_c^N$  is invariant for the explicit scheme (1.13).*

*Proof.* Using the definition of  $L$ , (1.11), (1.13) can be written

$$U_k^n = U_k^{n-1} + \beta d(U_{k+1}^{n-1} - 2U_k^{n-1} + U_{k-1}^{n-1}) + \frac{\alpha}{2}(F_{k-1}^{n-1} - F_{k+1}^{n-1})$$

$$= (1 - 2\beta d)U_k^{n-1} + 2\beta d \left[ \frac{1}{2}(U_{k+1}^{n-1} + U_{k-1}^{n-1}) + \frac{\Delta x}{4d}(F_{k-1}^{n-1} - F_{k+1}^{n-1}) \right].$$

If we assume by induction that  $U_k^{n-1} \in S_c$  for all  $k$ , then the brackets is in  $S_c$  by the theorem and our hypothesis on  $\Delta x$ . The assumption about  $\beta$  then shows that  $U_k^n$  is a convex combination of points of  $S_c$ .  $\square$

**COROLLARY 2.** *If  $\Delta x \leq 2d/\sqrt{-p'(c)}$ , then  $S_c^N$  is invariant for the semidiscrete scheme (1.12).*

*Proof.* (1.13) is Euler's method for the o.d.e. (1.12).  $\square$

Before extending this result to the scheme of interest, (1.10), we introduce the following notation. For  $U \in S_c^N$  define

$$(2.15) \quad r_c^k(U) = r_c(U_k) = u_k + \int_c^{v_k} \sqrt{-p'(s)} ds, \quad \text{and}$$

$$s_c^k(U) = s_c(U_k) = u_k - \int_c^{v_k} \sqrt{-p'(s)} ds.$$

Then

$$(2.16) \quad S_c^N = \{U: s_c^k(U) \leq 0 \leq r_c^k(U) \text{ for } k = 1, \dots, N\};$$

see (1.6).

We shall require the following technical lemma.

**LEMMA.** *Assume that  $\Delta x \leq 2d/\sqrt{-p'(c)}$ . Let  $U \in S_c^N$  and suppose that  $r_c^k(U) = 0$  (or  $s_c^k(U) = 0$ ) for some  $k$ . Then*

$$\langle \nabla r_c^k(U), L(U) \rangle \geq 0 \quad (\text{or } \langle \nabla s_c^k(U), L(U) \rangle \leq 0).$$

(Here  $\nabla$  is the gradient with respect to  $U$ , and  $\langle \cdot, \cdot \rangle$  is the usual inner product on  $R^{2N}$ .)

*Proof.* Let  $U(t)$  be the solution of the semidiscrete scheme (1.12) with  $U(0) = U$ . Then for small positive  $t$ ,  $U(t) \in S_c^N$  by Corollary 2 above, so that  $0 \leq r_c^k(U(t))$  for such  $t$ . From Taylor's theorem, then,

$$0 \leq r_c^k(U(t)) = r_c^k(U) + t \langle \nabla r_c^k(U), L(U) \rangle + O(t^2).$$



Since  $r_c^k(U) = 0$ , the result follows. The proof is similar in the case that  $s_c^k(U) = 0$ .  $\square$

If  $f$  is a real-valued function of a vector  $U$ , denote by  $H_f$  the Hessian matrix of  $f$  with respect to  $U$ . A simple computation then shows that

$$(2.17) \quad H_{r_c^k} \leq 0 \leq H_{s_c^k}.$$

The following theorem is the main result of this section.

**THEOREM 2.2.** *Assume that  $\Delta x \leq 2d/\sqrt{-p'(c)}$  and that  $\alpha \leq [\max(1, -p'(c))]^{-1}$ . Then given  $U^{n-1} \in S_c^N$ , there is a unique  $U^n \in S_c^N$  solving the implicit scheme (1.10)*

*Proof.* We attempt to solve

$$(2.18) \quad U(\tau) = U^{n-1} + \tau L(U(\tau))$$

for  $\tau \in [0, \Delta t]$  by integrating the o.d.e.

$$(2.19) \quad [I - \tau J_L(U(\tau))] \frac{d}{d\tau} U(\tau) = L(U(\tau)), \quad U(0) = U^{n-1}$$

( $J_L$  is the Jacobian matrix of  $L$  with respect to  $U$ ).

It is obvious that (2.19) has a solution for  $\tau$  near 0. If the solution fails to exist up to  $\tau = \Delta t$ , then one of the following must occur:

- (a)  $U(\tau)$  exits the set on which  $L$  is defined and Lipschitz (some  $v$ -component of  $U(\tau)$  approaches 0);
- (b)  $I - \tau J_L$  fails to be invertible;
- (c)  $|U(\tau)|$  becomes infinite.

(a) is precluded by showing that, in fact,  $U(\tau) \in S_c^N$ , which we now do. Temporarily assume that

$$(2.20) \quad \Delta x < 2d/\sqrt{-p'(c)}.$$

If  $U(\tau)$  does not remain in  $S_c^N$ , then for some  $\tau_0 \geq 0$  and for some  $k$ ,  $r_c^k(U(\tau_0)) = 0$  (or  $s_c^k(U(\tau_0)) = 0$ ) with  $r_c^k(U(\tau_0 + \epsilon)) < 0$  (or  $s_c^k(U(\tau_0 + \epsilon)) > 0$ ) for small positive  $\epsilon$ . We shall consider only the first case. It is then an easy matter to show that there is a  $c_1 < c$  with

$$(2.21) \quad \Delta x \leq 2d/\sqrt{-p'(c_1)}$$

(possible by (2.20)) and a  $\tau_1 > 0$  with  $r_{c_1}^k(U(\tau_1)) = 0$ ; see Figure 4.

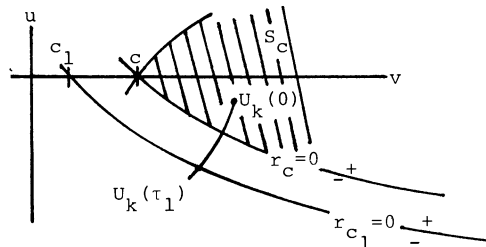


FIGURE 4

We would then have from (2.17) and (2.18) that

$$r_{c_1}^k(U^{n-1}) = r_{c_1}^k(U(0)) \leq r_{c_1}^k(U(\tau_1)) - \tau_1 \langle \nabla r_{c_1}^k(U(\tau_1)), L(U(\tau_1)) \rangle.$$

But  $r_{c_1}^k(U(\tau_1)) = 0$ , and the lemma applied to  $U(\tau_1)$  with  $c$  replaced by  $c_1$  (this is where we use (2.21)) shows that  $r_{c_1}^k(U^{n-1}) \leq 0$ . This is false because since  $U^{n-1} \in S_c^N$ ,  $r_c^k(U^{n-1}) \geq 0$ ; and this implies that  $r_{c_1}^k(U^{n-1}) > 0$ ; see Figure 4.

Thus,  $U(\tau) \in S_c^N$  as long as it is defined. The strict inequality (2.20) can be removed by continuity.

For (b) define the linear operators  $\delta$  and  $\delta^2$  on  $R^{2N}$  by

$$(\delta U)_k = U_{k-1} - U_{k+1} \quad \text{and} \quad (\delta^2 U)_k = U_{k+1} - 2U_k + U_{k-1}.$$

Then from its definition, (1.11),  $L$  can be written

$$L(U) = \frac{d}{\Delta x^2} \delta^2 U + \frac{1}{2\Delta x} \delta F(U),$$

where  $F(U)_k \equiv F(U_k)$ . Thus,

$$J_L(U) = \frac{d}{\Delta x^2} \delta^2 + \frac{1}{2\Delta x} \delta J_F(U)$$

and

$$I - \tau J_L(U) = \left( I - d \frac{\tau}{\Delta x^2} \delta^2 \right) \left[ I - \frac{\tau}{2\Delta x} \left( I - d \frac{\tau}{\Delta x^2} \delta^2 \right)^{-1} (\delta J_F(U)) \right].$$

From [4, p. 202] it follows that the symmetric operator  $I - d(\tau/\Delta x^2)\delta^2$  is not only invertible, but also that the norm of its inverse is strictly less than 1 for  $\tau > 0$ . It will follow then that  $I - \tau J_L(U)$  is invertible with the norm of its inverse bounded independently of  $U \in S_c^N$ , if we can show that

$$(2.22) \quad \frac{\tau}{2\Delta x} |\delta J_F(U)| \leq 1$$

for  $\tau \leq \Delta t$  and  $U \in S_c^N$ . Now,  $|J_F(U)|^2$  is the spectral radius of  $J_F(U)J_F(U)^t$ . Computing from (1.7), the latter matrix turns out to be a block diagonal matrix whose typical diagonal entry is

$$J_F(U_k)J_F(U_k)^t = \begin{bmatrix} p'(v_k)^2 & 0 \\ 0 & 1 \end{bmatrix}.$$

Thus,  $|J_F(U)| = \max(1, -p'(v_k)) \leq \max(1, -p'(c))$ , if  $U \in S_c^N$ . (2.22) then follows by our hypothesis on  $\alpha = \Delta t/\Delta x$  and the fact that  $|\delta| \leq 2$ , which is obvious from Gerschgorin's theorem.

Finally, (c) is precluded by observing that  $L(U)$  is of the order of  $U$ ; and since  $[I - \tau J_L(U)]^{-1}$  is uniformly bounded in  $S_c^N$ , it follows from (2.19) that  $U(\tau)$  could grow at most exponentially.

Thus, (2.19) is integrable for  $\tau$  up to  $\Delta t$ , and  $U^n \equiv U(\Delta t)$  provides a solution

of the implicit equation (1.10). If  $V^n$  is another solution in  $S_c^N$ , then subtracting, we obtain

$$\begin{aligned} 0 &= (U^n - V^n) - \Delta t[L(U^n) - L(V^n)] \\ &= \left[ I - \beta d\delta^2 - \frac{\alpha}{2} \delta J \right] (U^n - V^n) \\ &= (I - \beta d\delta^2) \left[ I - \frac{\alpha}{2} (I - \beta d\delta^2)^{-1} \delta J \right] (U^n - V^n), \end{aligned}$$

where  $J = J_F(U')$  for some  $U' \in S_c^N$  (note that  $F$  is nonlinear only in  $v$  so that the mean-value theorem applies). Above we argued that the matrix multiplying  $(U^n - V^n)$  is invertible, so that  $U^n - V^n = 0$ .  $\square$

We now consider the problem of implementing (1.10). That is, given  $U^{n-1} \in S_c^N$ , how do we compute the solution  $U^n$  of

$$(2.23) \quad U^n = U^{n-1} + \Delta t L(U^n)?$$

The simplest method would be the following fixed point iteration. Write  $\Delta t L = \beta d\delta^2 + \alpha \delta F/2$  and compute

$$U^{(m)} = U^{n-1} + \beta d\delta^2 U^{(m)} + \frac{\alpha}{2} \delta F(U^{(m-1)}),$$

or

$$(2.24) \quad (I - \beta d\delta^2)U^{(m)} = U^{n-1} + \frac{\alpha}{2} \delta F(U^{(m-1)}),$$

where presumably  $U^{(0)} = U^{n-1}$ . Notice that the computation of  $U^{(m)}$  involves only the solution of linear equations and that the relevant matrix is fixed. Furthermore, if  $\alpha$  is sufficiently small, the iteration function of (2.24) will be a contraction. In fact, the required condition on  $\alpha$  is precisely the hypothesis of Theorem 2.2, as a simple computation will show. We would, therefore, expect the  $U^{(m)}$  to converge to a vector  $U^n$  which by continuity solves (2.23).

There is a difficulty with all of this, however. The condition  $\alpha \leq [\max(1, -p'(c))]^{-1}$  does guarantee that the iteration function of (2.24) has Lipschitz constant less than 1, but only for arguments  $U$  which are in  $S_c^N$ . And we do not know that the iterants  $U^{(m)}$  remain in  $S_c^N$ .

One way of circumventing this difficulty is the following. Define

$$\tilde{p}(v) = \begin{cases} p(v), & c \leq v, \\ p(c) + p'(c)(v - c), & v \leq c. \end{cases}$$

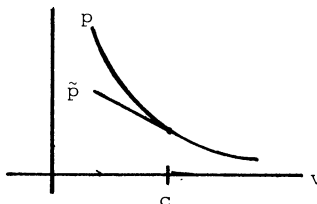


FIGURE 5

Now let  $\tilde{F}$ ,  $\tilde{L}$ , and  $\tilde{S}_c^N$  be just as in (1.7), (1.6), and (1.11), but with  $p$  replaced by  $\tilde{p}$ . If  $\alpha \leq [\max(1, -p'(c))]^{-1}$ , then  $\alpha \leq [\max(1, -\tilde{p}'(v))]^{-1}$  for all  $v$ , so that the iteration function of (2.24), with  $L$  replaced by  $\tilde{L}$ , is a contraction for *all* arguments. If we compute the  $U^{(m)}$  with  $\tilde{L}$  instead of  $L$ , then, we are guaranteed that  $U^{(m)} \rightarrow U^n$ , where  $U^n = U^{n-1} + \Delta t \tilde{L}(U^n)$ . Now apply Theorem 2.2 to conclude that  $U^n \in \tilde{S}_c^N$  (the requirement that  $p \in C^2$  in Theorem 2.2 can be removed by a limiting process). But  $p$  and  $\tilde{p}$  agree for  $c \leq v$  so that  $S_c^N = \tilde{S}_c^N$  and, in fact,  $U^n \in S_c^N$ . But then  $L(U^n) = \tilde{L}(U^n)$  so that  $U^n$  satisfies the desired equation  $U^n = U^{n-1} + \Delta t L(U^n)$ .

**3. Energy Estimates.** In this section we show that, under the stability conditions

$$(3.1) \quad \alpha \leq [\max(1, -p'(c))]^{-1}, \quad \Delta x \leq \frac{2d}{\sqrt{-p'(c)}},$$

the solution of the scheme (1.10) remains bounded and, in fact, decays to the boundary values exponentially as  $t \rightarrow \infty$ . This knowledge will be important for the application of the error estimates of Section 4. Our method of proof is essentially the discrete version of the technique employed in [2].

It will be convenient to rewrite the difference equation (1.10) in the following way. Let

$$u^n = \begin{bmatrix} u_1^n \\ \vdots \\ u_N^n \end{bmatrix} \quad \text{and} \quad v^n = \begin{bmatrix} v_1^n \\ \vdots \\ v_N^n \end{bmatrix}$$

so that the  $U^n$  of Section 1 and Section 2 can be written

$$U^n = [u_1^n, v_1^n, \dots, u_N^n, v_N^n]^t.$$

(1.10) then becomes

$$(3.2) \quad u^n = u^{n-1} + \beta d A u^n + \frac{\alpha}{2} D(p_b - p^n),$$

$$(3.3) \quad v^n = v^{n-1} + \beta d A(v^n - v_b) + \frac{\alpha}{2} D u^n.$$

Here  $A$  and  $D$  are the  $N \times N$  matrices

$$A = \begin{bmatrix} -2 & 1 & \cdot & \cdot & 0 \\ 1 & \cdot & \cdot & \cdot & 1 \\ 0 & \cdot & \cdot & 1 & -2 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 0 & 1 & \cdot & \cdot & 0 \\ -1 & \cdot & \cdot & \cdot & 1 \\ 0 & \cdot & \cdot & -1 & 0 \end{bmatrix}$$

and  $v_b, p_b$  and  $p^n$  are the  $N$ -vectors

$$v_b = \begin{bmatrix} v_b \\ \vdots \\ v_b \end{bmatrix}, \quad p_b = \begin{bmatrix} p(v_b) \\ \vdots \\ p(v_b) \end{bmatrix}, \quad \text{and} \quad p^n = \begin{bmatrix} p(v_1^n) \\ \vdots \\ p(v_N^n) \end{bmatrix},$$

(It will be clear from the context whether  $v_b$  is in  $R$  or  $R^N$ .)

Notice that  $A$  and  $D$  are  $O(\Delta x^2)$  approximations to  $d^2/dx^2$  and  $d/dx$ , respectively, only for functions satisfying zero Dirichlet boundary conditions. This accounts for the presence of  $v_b$  and  $p_b$  in (3.2) and (3.3).

Henceforth,  $\langle \cdot, \cdot \rangle$  and  $|\cdot|$  will denote the usual inner product and  $l^2$ -norm on  $R^N$ , respectively. Also, the letter  $K$  with or without a subscript will denote a positive constant which depends only on the parameters and data appearing in (1.1)–(1.3), but not on  $x, t$ , or the mesh parameters.

We shall require the following technical facts.

LEMMA. *Let  $w$  and  $z$  be in  $R^N$  with  $w_i = z_i = 0$  for  $i = 0$  and  $N + 1$ . Then*

$$(3.4) \quad \langle Aw, z \rangle = - \sum_{k=0}^N (w_{k+1} - w_k) (z_{k+1} - z_k),$$

$$(3.5) \quad |Dw|^2 \leq 4 \sum_{k=0}^N (w_{k+1} - w_k)^2 = -4 \langle Aw, w \rangle,$$

$$(3.6) \quad |w| \leq L \left[ \sum_{k=0}^N \left( \frac{w_{k+1} - w_k}{\Delta x} \right)^2 \right]^{1/2},$$

$$(3.7) \quad \langle Aw, w \rangle \leq - \frac{1}{L^2} |w|^2 \Delta x^2,$$

$$(3.8) \quad - \langle Aw, w \rangle \Delta x^2 \leq L^2 |Aw|^2.$$

$$(3.9) \quad w_j^2 \leq 2 \left[ \sum_{k=1}^N w_k^2 \Delta x \right]^{1/2} \left[ \sum_{k=0}^N \left( \frac{w_{k+1} - w_k}{\Delta x} \right)^2 \Delta x \right]^{1/2}.$$

The proofs are elementary and so are omitted.

Define the discrete energies

$$E^n = \frac{1}{2} |u^n|^2 \Delta x + \sum_{k=1}^N \psi(v_k^n) \Delta x,$$

where

$$\psi(v) = \int_{v_b}^v [p(v_b) - p(s)] ds \geq 0;$$

and

$$F^n = \frac{1}{2} \sum_{k=0}^N \left[ \left( \frac{u_{k+1}^n - u_k^n}{\Delta x} \right)^2 + \left( \frac{v_{k+1}^n - v_k^n}{\Delta x} \right)^2 \right] \Delta x.$$

(We tacitly assume that  $u_0^n = u_{N+1}^n = 0$  and  $v_0^n = v_{N+1}^n = v_b$ .)

Our immediate goal is to estimate  $E^n$  and  $F^n$ .

LEMMA 1. *Assume that the stability conditions (3.1) are satisfied. Let  $M = \max_k v_k^n$  and  $q(M) = \min(1, -p'(M))$ . Then there is a  $K$  such that*

$$E^n \leq E^{n-1} - Kq(M)\Delta tF^n.$$

*Proof.* Since  $p(v_b) - p(v)$  is monotone, we have for some  $s \in (0, 1)$  that

$$\begin{aligned} \psi(v_k^n) - \psi(v_k^{n-1}) &= \int_{v_k^{n-1}}^{v_k^n} [p(v_b) - p(s)] ds \\ &= [p(v_b) - sp_k^{n-1} - (1-s)p_k^n] (v_k^n - v_k^{n-1}) \\ &= [p(v_b) - p_k^n] (v_k^n - v_k^{n-1}) + s(p_k^n - p_k^{n-1})(v_k^n - v_k^{n-1}), \\ &\leq [p(v_b) - p_k^n] (v_k^n - v_k^{n-1}), \end{aligned}$$

since  $p' < 0$ . Summing over  $k$  and using (3.3), we obtain

$$(3.10) \quad \begin{aligned} \sum_k \psi_k^n &\leq \sum_k \psi_k^{n-1} + \beta d \langle A(v^n - v_b), (p_b - p^n) \rangle \\ &\quad + \frac{\alpha}{2} \langle Du^n, (p_b - p^n) \rangle. \end{aligned}$$

If we inner product (3.2) with  $u^n$  and bound  $\langle u^n, u^{n-1} \rangle$  by  $\frac{1}{2}(|u^n|^2 + |u^{n-1}|^2)$ , we obtain

$$(3.11) \quad \frac{1}{2} |u^n|^2 \leq \frac{1}{2} |u^{n-1}|^2 + \beta d \langle Au^n, u^n \rangle - \frac{\alpha}{2} \langle Du^n, p_b - p^n \rangle.$$

Add (3.10) and (3.11) and multiply by  $\Delta x$ . The result is  $E^n \leq E^{n-1} + \beta d[\langle Au^n, u^n \rangle + \langle A(v^n - v_b), (p_b - p^n) \rangle] \Delta x$ . The brackets may be estimated by (3.4) to give

$$E^n \leq E^{n-1} - d\Delta t \sum_{k=0}^N \left[ \left( \frac{u_{k+1}^n - u_k^n}{\Delta x} \right)^2 + \left( \frac{v_{k+1}^n - v_k^n}{\Delta x} \right) \left( \frac{p_k^n - p_{k+1}^n}{\Delta x} \right) \right] \Delta x.$$

Since

$$(3.12) \quad (v_{k+1}^n - v_k^n)(p_k^n - p_{k+1}^n) = -p'(\xi) (v_{k+1}^n - v_k^n)^2 \geq -p'(M) (v_{k+1}^n - v_k^n)^2,$$

we obtain finally

$$E^n \leq E^{n-1} - 2d\Delta t \min(1, -p'(M))F^n. \quad \square$$

LEMMA 2. Assume (3.1). Then there is a constant  $K$  such that

$$F^n \leq F^{n-1} + K \Delta t F^n.$$

*Proof.* From (3.4) and the difference equation (3.2) we have

$$(3.13) \quad \begin{aligned} \sum_{k=0}^N (u_{k+1}^n - u_k^n)^2 &= -\langle Au^n, u^n \rangle \\ &= -\langle Au^n, u^{n-1} \rangle - \beta d |Au^n|^2 + \frac{\alpha}{2} \langle D(p_b - p^n), Au^n \rangle. \end{aligned}$$

Again by (3.4),

$$\begin{aligned}
 -\langle Au^n, u^{n-1} \rangle &= \sum_0^N (u_{k+1}^n - u_k^n) (u_{k+1}^{n-1} - u_k^{n-1}) \\
 &\leq \frac{1}{2} \sum_0^N (u_{k+1}^n - u_k^n)^2 + \frac{1}{2} \sum_0^N (u_{k+1}^{n-1} - u_k^{n-1})^2
 \end{aligned}$$

so that (3.13) becomes

$$\begin{aligned}
 \frac{1}{2} \sum_0^N (u_{k+1}^n - u_k^n)^2 &\leq \frac{1}{2} \sum_0^N (u_{k+1}^{n-1} - u_k^{n-1})^2 - \beta d |Au^n|^2 \\
 &\quad + \frac{\alpha}{2} \left[ \frac{1}{2\epsilon} |D(p_b - p^n)|^2 + \frac{\epsilon}{2} |Au^n|^2 \right].
 \end{aligned}$$

Choose  $\alpha\epsilon/4 = \beta d$  so that the  $|Au^n|^2$  terms cancel. Then bounding  $|D(p_b - p^n)|^2$  by  $p'(c)^2 |D(v^n - v_b)|^2$ , we obtain

$$(3.14) \quad \frac{1}{2} \sum_0^N (u_{k+1}^n - u_k^n)^2 \leq \frac{1}{2} \sum_0^N (u_{k+1}^{n-1} - u_k^{n-1})^2 + \frac{\Delta t}{16d} p'(c)^2 |D(v^n - v_b)|^2.$$

Similarly,

$$(3.15) \quad \frac{1}{2} \sum_0^N (v_{k+1}^n - v_k^n)^2 \leq \frac{1}{2} \sum_0^N (v_{k+1}^{n-1} - v_k^{n-1})^2 + \frac{\Delta t}{16d} |Du^n|^2.$$

Add (3.14) and (3.15) and divide by  $\Delta x$  to obtain

$$F^n \leq F^{n-1} + \frac{\Delta t}{16d} \max(1, p'(c)^2) \left( \left| \frac{Du^n}{\Delta x} \right|^2 + \left| \frac{D(v^n - v_b)}{\Delta x} \right|^2 \right) \Delta x.$$

The result then follows from (3.5).  $\square$

**THEOREM 3.1.** *Assume (3.1). Then there is a constant  $M$  depending only on the parameters in (1.1)–(1.3) such that  $v_k^n \leq M$  for all  $n$  and  $k$ .*

*Proof.* Fix  $n$  and let  $M = \max_{k,m;n} v_k^m$ . From Lemmas 1 and 2 we have

$$(3.16) \quad F^n \leq F^{n-1} + \frac{K}{q(M)} (E^{n-1} - E^n) \leq \dots \leq F^0 + \frac{K}{q(M)} (E^0 - E^n) \leq \frac{K}{q(M)},$$

where  $K$  is a constant depending on the data. The idea is that  $M$  can be estimated by  $F^n$  and, as above,  $F^n$  by  $M$ . The result will then follow from the assumption (1.4).

$$(3.17) \quad \lim_{v \rightarrow \infty} v^3 |p'(v)| = \infty.$$

Let  $f(v) = \int_{v_b}^v \sqrt{\psi(s)} ds$ . Then

$$f(v_j^n) = \sum_0^{j-1} [f(v_{k+1}^n) - f(v_k^n)] = \sum_0^{j-1} \sqrt{\psi(\xi_k)} (v_{k+1}^n - v_k^n),$$

where, since  $\sqrt{\psi}$  is monotone,  $\psi(\xi_k)$  is a convex combination of  $\psi(v_{k+1}^n)$  and  $\psi(v_k^n)$ .

Hence

$$\begin{aligned}
 (3.18) \quad f(v_j^n)^2 &\leq \sum_0^N \psi(\xi_k) \sum_0^N (v_{k+1}^n - v_k^n)^2 \\
 &\leq 2 \left[ \sum_0^N \psi(v_k^n) \Delta x \right] \left[ \sum_0^N \left( \frac{v_{k+1}^n - v_k^n}{\Delta x} \right)^2 \Delta x \right] \\
 &\leq 2E^n F^n \leq 2E^0 F^n = KF^n.
 \end{aligned}$$

Combining (3.16) and (3.18), we have

$$(3.19) \quad f(M)^2 \leq \frac{K}{q(M)}.$$

Now if  $v \geq 3v_b$ ,

$$\begin{aligned}
 f(v) &> \int_{2v_b}^v \left[ \int_{2v_b}^s (p(v_b) - p(y)) dy \right]^{1/2} ds \\
 &> \int_{2v_b}^v \left[ \int_{2v_b}^s (p(v_b) - p(2v_b)) dy \right]^{1/2} ds = K(v - v_b)^{3/2},
 \end{aligned}$$

and by changing  $K$ ,

$$(3.20) \quad f(v) > Kv^{3/2}.$$

From (1.4) it is easy to check that either  $-p'(v) < 1$  or there is a unique  $M_0$  with  $-p(M_0) = 1$ . If  $M$  is smaller than neither  $M_0$  nor  $3v_b$ , then  $-p'(M) < 1$  so that  $q(M) = -p'(M)$ . And from (3.20) and (3.19),  $M^3 < K/-p'(M)$ . It then follows from (3.17) that  $M$  cannot be arbitrarily large.  $\square$

COROLLARY 1.  $F^n \leq K$  for all  $n$ .

*Proof.* This is (3.16) with  $M = \sup_{k,n} v_k^n$ .  $\square$

COROLLARY 2. There is a  $K$  such that

$$E^n \leq (1 - K \Delta t)^n E^0.$$

*Proof.* From (3.6),

$$(3.21) \quad \frac{1}{2} |u^n|^2 \Delta x \leq K \sum_0^N \left( \frac{u_{k+1}^n - u_k^n}{\Delta x} \right)^2 \Delta x.$$

Also from Taylor's theorem,

$$\begin{aligned}
 (3.22) \quad \psi(v_k^n) &= \psi(v_b) + \psi'(v_b)(v_k^n - v_b) + \frac{1}{2} \psi''(\xi)(v_k^n - v_b)^2 \\
 &= -\frac{1}{2} p'(\xi)(v_k^n - v_b)^2 \leq K(v_k^n - v_b)^2.
 \end{aligned}$$

Hence, using (3.6) again, we have

$$(3.23) \quad \sum_1^N \psi(v_k^n) \Delta x \leq K \sum_0^N (v_{k+1}^n - v_b)^2 \Delta x \leq K \sum_0^N \left( \frac{v_{k+1}^n - v_k^n}{\Delta x} \right)^2 \Delta x.$$



Adding (3.21) and (3.23), we see that  $E^n \leq KF^n$ . Combining this with Lemma 1, we therefore have

$$(1 + K \Delta t)E^n \leq E^{n-1} \quad \text{or} \quad E^n \leq (1 - K \Delta t)E^{n-1}.$$

The result then follows by induction.  $\square$

**THEOREM 3.2.** *Assume (3.1). Then there are constants  $K$  and  $K_1$  such that*

$$|u_k^n| + |v_k^n - v_b| \leq K_1 \left(1 - \frac{K t_n}{n}\right)^n.$$

*Proof.* From (3.9),

$$(3.24) \quad (u_k^n)^2 \leq K(E^n F^n)^{1/2}.$$

Also, from (3.22) and the fact that  $v_k^n$  is bounded,

$$|v_k^n - v_b|^2 \leq K \psi(v_k^n)$$

so that

$$|v^n - v_b|^2 \Delta x \leq K E^n.$$

Applying (3.9), we therefore have

$$(3.25) \quad |v_k^n - v_b|^2 \leq K(E^n F^n)^{1/2}.$$

The result then follows from (3.24), (3.25), and Corollaries 1 and 2.  $\square$

**4. Error Estimates.** Let  $e^n$  be the difference between  $u^n$  and  $[u(x_1, t_n), \dots, u(x_N, t_n)]^t$  and  $f^n$  between  $v^n$  and  $[v(x_1, t_n), \dots, v(x_N, t_n)]^t$ . From (3.2) and (3.3) we have

$$(4.1) \quad e^n = e^{n-1} + \beta d A e^n - \frac{\alpha}{2} D Q^n f^n + \tau^n,$$

$$(4.2) \quad f^n = f^{n-1} + \beta d A f^n + \frac{\alpha}{2} D e^n + \sigma^n,$$

where  $Q^n$  is the diagonal matrix

$$(4.3) \quad Q^n = \begin{vmatrix} p'(\xi_1^n) & 0 \\ & \cdot \\ 0 & p'(\xi_N^n) \end{vmatrix}$$

with

$$(4.4) \quad p'(\xi_k^n) = \frac{p(v_k^n) - p(v(x_k, t_n))}{v_k^n - v(x_k, t_n)};$$

and  $\tau^n$  and  $\sigma^n$  are the local truncation errors. Since  $u(x, t)$  and  $v(x, t)$  are smooth with derivatives bounded independently of  $t$ , (see [2]), and because we used symmetric centered space differences,  $\tau_k^n$  and  $\sigma_k^n$  are of the order  $\Delta t h$ , where  $h = \Delta t + \Delta x^2$ . Hence

$$(4.5) \quad |\tau^n|^2 \Delta x, |\sigma^n|^2 \Delta x \leq K \Delta t^2 h^2.$$

We shall estimate the  $L^2$  errors

$$E^n = \frac{1}{2}(|e^n|^2 + |f^n|^2) \Delta x$$

and

$$F^n = \frac{1}{2}(|e^n|^2 - p'_b |f^n|^2) \Delta x,$$

where  $p'_b = p'(v_b) < 0$ .

**THEOREM 4.1.** *Assume the stability conditions (3.1), let  $h = \Delta t + \Delta x^2$ , and let  $E^n$  and  $F^n$  be as above.*

(a) *Then*

$$\sqrt{E^n} \leq e^{Kt_n}(\sqrt{E^0} + K_1 h).$$

(b) *If in addition  $d/L > (1 - p'(c))/2$ , then*

$$\sqrt{E^n} \leq \left(1 - \frac{Kt_n}{n}\right)^n (\sqrt{E^0} + K_1 h).$$

(c) *Suppose that  $a \leq v_k^n, v(x_k, t_n) \leq b$  for all  $n$  and  $k$ . Define*

$$\gamma = \max_{a \leq v_1, v_2 \leq b} |p'(v_1) - p'(v_2)|.$$

*If  $d/L > \gamma/2\sqrt{-p'_b}$ , then*

$$\sqrt{F^n} \leq \left(1 - \frac{Kt_n}{n}\right)^n (\sqrt{F^0} + K_1 h).$$

In each case  $K$  and  $K_1$  are positive constants independent of  $x, t, \Delta x$  and  $\Delta t$ .

*Proof.* Inner product (4.1) with  $e^n$  to obtain

$$|e^n|^2 \leq \langle e^n, e^{n-1} \rangle + \beta d \langle Ae^n, e^n \rangle + \frac{\alpha}{2} \langle Q^n f^n, De^n \rangle + |\tau^n| |e^n|$$

so that

$$(4.6) \quad \frac{1}{2} |e^n|^2 \leq \frac{1}{2} |e^{n-1}|^2 + \beta d \langle Ae^n, e^n \rangle + \frac{\alpha}{2} \langle Q^n f^n, De^n \rangle + \frac{\eta}{2} |e^n|^2 + \frac{1}{2\eta} |\tau^n|^2,$$

where  $\eta > 0$  will be chosen. Similarly,

$$(4.7) \quad \frac{1}{2} |f^n|^2 \leq \frac{1}{2} |f^{n-1}|^2 + \beta d \langle Af^n, f^n \rangle + \frac{\alpha}{2} \langle f^n, De^n \rangle + \frac{\eta}{2} |f^n|^2 + \frac{1}{2\eta} |\sigma^n|^2.$$

Add (4.6) and (4.7) and multiply by  $\Delta x$  to obtain

$$(4.8) \quad \begin{aligned} E^n &\leq E^{n-1} + \beta d [\langle Ae^n, e^n \rangle + \langle Af^n, f^n \rangle] \Delta x \\ &\quad + \frac{\alpha}{2} (1 - p'(c)) \left[ \frac{\epsilon}{2} |f^n|^2 + \frac{1}{2\epsilon} |De^n|^2 \right] \Delta x \\ &\quad + \eta E^n + \frac{1}{2\eta} [|\tau^n|^2 + |\sigma^n|^2] \Delta x. \end{aligned}$$

*Proof of (a).* Since  $|De^n|^2 \leq -4 \langle Ae^n, e^n \rangle$  by (3.5), the terms  $|De^n|^2$  and  $\langle Ae^n, e^n \rangle$  will cancel if we choose  $\epsilon = O(\Delta x)$  (recall the definitions of  $\alpha$  and  $\beta$ , (2.1)). Also,  $\langle Af^n, f^n \rangle \leq 0$  and  $\alpha\epsilon |f^n|^2 \Delta x \leq K\Delta t E^n$ . Thus, if we choose  $\eta = O(\Delta t)$ , it follows that

$$E^n \leq E^{n-1} + K\Delta t E^n + K_1 \Delta t h^2,$$

where we used (4.5) to estimate the  $\tau$  and  $\sigma$  terms. Thus, for different  $K$  and  $K_1$ ,

$$E^n \leq (1 + K\Delta t) E^{n-1} + K_1 \Delta t h^2,$$

and (a) follows by induction.

*Proof of (b).* Alternatively, the second brackets on the right of (4.8) is bounded by

$$-\frac{\epsilon L^2}{2\Delta x^2} \langle Af^n, f^n \rangle - \frac{2}{\epsilon} \langle Ae^n, e^n \rangle$$

by (3.7) and (3.5). Thus, (4.8) becomes

$$\begin{aligned} E^n &\leq E^{n-1} + \left[ \beta d - \frac{\alpha(1-p'(c))}{\epsilon} \right] \langle Ae^n, e^n \rangle \Delta x \\ (4.9) \quad &+ \left[ \beta d - \frac{\alpha\epsilon(1-p'(c))L^2}{4\Delta x^2} \right] \langle Af^n, f^n \rangle \Delta x \\ &+ \eta E^n + \frac{1}{2\eta} (|\tau^n|^2 + |\sigma^n|^2) \Delta x. \end{aligned}$$

If we choose  $\epsilon = 2\Delta x/L$ , then both brackets equal

$$\beta d - \frac{\alpha L(1-p'(c))}{2\Delta x} = \beta \left[ d - \frac{L(1-p'(c))}{2} \right] = K\beta,$$

where by hypothesis  $K$  is positive. Now, from (3.7),

$$K\beta [\langle Ae^n, e^n \rangle + \langle Af^n, f^n \rangle] \Delta x \leq -K\Delta t (|e^n|^2 + |f^n|^2) \Delta x = -K\Delta t E^n$$

so that (4.9) becomes

$$(4.10) \quad E^n \leq E^{n-1} - (K\Delta t - \eta) E^n + \frac{1}{2\eta} (|\tau^n|^2 + |\sigma^n|^2) \Delta x.$$

Choosing  $\eta = O(\Delta t)$  and using (4.5), we finally have, for some positive  $K$  and  $K_1$ ,

$$E^n \leq (1 - K\Delta t) E^{n-1} + K_1 \Delta t h^2.$$

(b) then follows by induction.

*Proof of (c).* Add (4.6) to  $-p'_b$  times (4.7) and multiply by  $\Delta x$  to obtain

$$\begin{aligned} F^n &\leq F^{n-1} + \beta d [\langle Ae^n, e^n \rangle - p'_b \langle Af^n, f^n \rangle] \Delta x \\ (4.11) \quad &+ \frac{\alpha}{2} \langle (Q^n - Q_b) f^n, De^n \rangle \Delta x + \eta F^n + \frac{1}{2\eta} (|\tau^n|^2 - p'_b |\sigma^n|^2) \Delta x, \end{aligned}$$

where  $Q_b = p'_b I$ . From the definition of  $Q^n$ , (4.3)–(4.4), the inner product on the right of (4.11) is bounded by

$$\begin{aligned} & \sup |p'(\xi_k^n) - p'_b| \left( \frac{\epsilon}{2} |f^n|^2 + \frac{1}{2\epsilon} |De^n|^2 \right) \Delta x \\ & \leq -\gamma \left[ \frac{\epsilon}{2} \frac{L^2}{\Delta x^2} \langle A f^n, f^n \rangle + \frac{2}{\epsilon} \langle A e^n, e^n \rangle \right] \Delta x, \end{aligned}$$

where again we used (3.7) and (3.5). Substituting into (4.11), we obtain

$$\begin{aligned} (4.12) \quad \bar{F}^n & \leq \bar{F}^{n-1} + \left[ \beta d - \frac{\alpha \gamma}{\epsilon} \right] \langle A e^n, e^n \rangle \\ & + \left[ \beta d + \frac{\alpha \gamma \epsilon L^2}{4 p'_b \Delta x^2} \right] (-p'_b) \langle A f^n, f^n \rangle + \eta \bar{F}^n + \frac{1}{2\eta} (|\tau^n|^2 - p'_b |\sigma^n|^2) \Delta x. \end{aligned}$$

If we choose  $\epsilon = 2\sqrt{-p'_b} \Delta x / L$ , then both brackets equal

$$\beta d - \frac{\alpha \gamma L}{2\sqrt{-p'_b} \Delta x} = \beta \left[ d - \frac{\gamma L}{2\sqrt{-p'_b}} \right] = K\beta,$$

where by hypothesis  $K$  is positive. Just as in the proof of (b),

$$K\beta [\langle A e^n, e^n \rangle - p'_b \langle A f^n, f^n \rangle] \Delta x \leq -K\Delta t \bar{F}^n,$$

so that (4.12) becomes

$$\bar{F}^n \leq \bar{F}^{n-1} - (K\Delta t - \eta) \bar{F}^n + \frac{1}{2\eta} (|\tau^n|^2 - p'_b |\sigma^n|^2) \Delta x.$$

This is identical in form to (4.10). The proof then proceeds just like the proof of (b).  $\square$

These results may be interpreted as follows. In all cases the approximants converge like  $O(h)$  in finite time as  $h \rightarrow 0$  (part (a) of the theorem). On the other hand, for  $h$  fixed, the error at first grows exponentially as  $t$  increases (unless the diffusion term  $d$  is large, by part (b)). But eventually, as soon as  $v_k^n$  and  $v(x, t)$  get close enough to  $v_b$  (which is exponentially soon by Theorem 3.2 and the corollary below), part (c) may be applied at a new initial time to show that the error remains  $O(h)$  for all subsequent times.

**COROLLARY.** *There are positive constants  $K$  and  $K_1$  so that the solution  $[u(x, t), v(x, t)]$  of (1.1)–(1.3) satisfies*

$$\max_{0 \leq x \leq L} (|u(x, t)| + |v(x, t) - v_b|) \leq K_1 e^{-Kt}.$$

*Proof.* Fix  $t$  and solve the difference equations (3.2)–(3.3) with  $\Delta t = 4^{-j}t$  and  $\Delta x = 2^{-j}L$ , committing no initial error. Then for  $j$  sufficiently large the stability conditions (3.1) are satisfied so that, if  $n = 4^j$ , we obtain from Theorem 3.2

$$(4.13) \quad |u_k^n| \leq K_1 \left( 1 - \frac{Kt}{n} \right)^n.$$

On the other hand, from Theorem 4.1(a).

$$(4.14) \quad |e_k^n|^2 \leq \frac{1}{\Delta x} E^n \leq K_2 \frac{h^2}{\Delta x} = K_2 \Delta x^3,$$

where  $K_2$  may depend on  $t$ . Combining (4.13) and (4.14), we have

$$|u(x, t)| \leq K_1 \left(1 - \frac{Kt}{n}\right)^n + K_2 \Delta x^{3/2}$$

for any  $x$  of the form  $K2^{-j}L$ . Now let  $j \rightarrow \infty$  and  $n \rightarrow \infty$  with  $n\Delta t = t$  fixed. The result is that  $|u(x, t)| \leq K_1 e^{-Kt}$  holds for a dense set of  $x$ , hence for all  $x$ . The estimate for  $v - v_b$  is similar.  $\square$

As the above proof shows, the  $O(\Delta x^2)$  error estimates in  $L^2$  of Theorem 3.1 easily translate into  $O(\Delta x^{3/2})$  estimates in sup norm. Actually,  $O(\Delta x^2)$  estimates are valid in sup norm. We prove this below for case (c) of Theorem 4.1 by way of the inequality (3.9).

**THEOREM 4.2.** *In addition to the hypotheses of Theorem 4.1(c), assume that  $\sqrt{F^0} \leq Kh$  and that*

$$\sum_{k=0}^N [(e_{k+1}^0 - e_k^0)^2 + (f_{k+1}^0 - f_k^0)^2] \leq Kh^2 \Delta x.$$

Then,  $|e_k^n| + |f_k^n| \leq Kh$  for some  $K$  independent of  $n$  and  $k$ .

*Proof.* By Theorem 3.1(c),  $\sqrt{F^n} \leq Kh$  for all  $n$  so that from (3.9),

$$|e_k^n| + |f_k^n| \leq Kh^{1/2} (G^n)^{1/4},$$

where

$$G^n = \frac{1}{2} \sum_{k=0}^N \left[ \left( \frac{e_{k+1}^n - e_k^n}{\Delta x} \right)^2 + \left( \frac{f_{k+1}^n - f_k^n}{\Delta x} \right)^2 \right] \Delta x.$$

The result then will follow from the estimate

$$(4.15) \quad G^n \leq Kh^2.$$

To prove (4.15), inner product (4.1) with  $-Ae^n$ , (4.2) with  $-Af^n$ , add, and divide by  $\Delta x$  to obtain

$$\begin{aligned} G^n &\leq G^{n-1} - \frac{\beta d}{\Delta x} (|Ae^n|^2 + |Af^n|^2) \\ &\quad + \frac{\beta}{2} [\langle DQ^n f^n, Ae^n \rangle - \langle De^n, Af^n \rangle] \\ &\quad + \frac{\eta}{2\Delta x} (|Ae^n|^2 + |Af^n|^2) + \frac{1}{2\eta\Delta x} (|\tau^n|^2 + |\sigma^n|^2). \end{aligned}$$

Choose  $\eta = \beta d$  so that the last term, via (4.5), is bounded by  $K_1 \Delta t h^2$ . Then

$$\begin{aligned}
 (4.16) \quad G^n - G^{n-1} &\leq -K \overbrace{\frac{\beta}{\Delta x} (|Ae^n|^2 + |Af^n|^2)}^{\text{I}} \\
 &\quad + \frac{\beta}{2} [\langle DQ^n f^n, Ae^n \rangle - \langle De^n, Af^n \rangle] + \underbrace{K_1 \Delta t h^2}_{\text{II}}.
 \end{aligned}$$

We shall show that, by altering  $K$  and  $K_1$ , the middle term on the right of (4.16) can be absorbed into the other two terms.

The second component of the middle term is

$$\beta \langle De^n, -Af^n \rangle \leq \beta \left( \frac{\epsilon}{2} |De^n|^2 + \frac{1}{2\epsilon} |Af^n|^2 \right).$$

If we choose  $\epsilon = O(\Delta x)$ , then  $(\beta/2\epsilon) |Af^n|^2$  is absorbed into I. What remains is

$$(4.17) \quad K\beta\Delta x |De^n|^2 \leq K\beta\Delta x |\langle Ae^n, e^n \rangle| \leq K\beta\Delta x \left( \frac{1}{2\epsilon} |Ae^n|^2 + \frac{\epsilon}{2} |e^n|^2 \right),$$

where we used (3.5). Now choose  $\epsilon = O(\Delta x^2)$  so that

$$\frac{K\beta\Delta x}{2\epsilon} |Ae^n|^2 = O\left(\frac{\beta}{\Delta x}\right) |Ae^n|^2$$

is absorbed into I. What remains is then

$$O(\Delta t |e^n|^2 \Delta x) = O(\Delta t h^2)$$

by Theorem 4.1. This term then is absorbed into II.

The first term in the brackets on the right side of (4.16) is bounded by

$$K\beta \left( \frac{1}{2\epsilon} |Ae^n|^2 + \frac{\epsilon}{2} |DQ^n f^n|^2 \right).$$

Again, the  $|Ae^n|^2$  is absorbed by I if  $\epsilon = O(\Delta x)$ . Letting  $p'(\xi_k^n) = p'_k$ , the other term then is

$$\begin{aligned}
 &K \frac{\Delta t}{\Delta x} \sum_{k=1}^N [p'_{k+1} f_{k+1}^n - p'_{k-1} f_{k-1}^n]^2 \\
 &= K \frac{\Delta t}{\Delta x} \sum [p'_{k+1} (f_{k+1}^n - f_{k-1}^n) + (p'_{k+1} - p'_{k-1}) f_{k-1}^n]^2.
 \end{aligned}$$

From the definition of  $\xi_k^n$ , (4.4), it follows that

$$\begin{aligned}
 p'_{k+1} - p'_{k-1} &\leq K(|f_{k-1}^n| + |f_{k+1}^n| + |v_x(y, t_n)| \Delta x) \\
 &\leq K(|f_{k-1}^n| + |f_{k+1}^n| + \Delta x) \leq K \Delta x
 \end{aligned}$$

by Theorem 4.1(c). Thus, the term in question is bounded by

$$K \frac{\Delta t}{\Delta x} (|Df^n|^2 + \Delta x^2 |f^n|^2).$$

The term  $|Df^n|^2$  is bounded just as was  $|De^n|^2$  in (4.17). And the other term, by Theorem 4.1, is bounded by  $K \Delta t \bar{F}^n \leq K \Delta t h^2$ , which may be absorbed into II.

(4.16) has thus been reduced to

$$(4.18) \quad G^n \leq G^{n-1} - K_1 \frac{\beta}{\Delta x} (|Ae^n|^2 + |Af^n|^2) + K_2 \Delta t h^2.$$

From (3.5), (3.8), and the definition of  $G^n$  we have

$$G^n \leq -\frac{K}{\Delta x} (\langle Ae^n, e^n \rangle + \langle Af^n, f^n \rangle) \leq \frac{K}{\Delta x^3} (|Ae^n|^2 + |Af^n|^2),$$

so that

$$-\frac{\beta}{\Delta x} (|Ae^n|^2 + |Af^n|^2) \leq -K \Delta t G^n.$$

Hence (4.18) becomes

$$G^n \leq (1 - K \Delta t) G^{n-1} + K_1 \Delta t h^2.$$

By induction we have

$$(4.19) \quad G^n \leq (1 - K \Delta t)^n G^0 + \frac{K_1}{K} h^2,$$

and the required estimate (4.15) follows from (4.19) and our hypothesis, which ensures that  $G^0 = O(h^2)$ .  $\square$

Department of Mathematics  
Indiana University  
Bloomington, Indiana 47405

1. K. N. CHUEH, C. C. CONLEY & J. A. SMOLLER, "Positively invariant regions for systems of nonlinear diffusion equations," *Indiana Univ. Math. J.*, v. 26, 1977, pp. 373-392.
2. YA. I. KANEL, "On some systems of quasilinear parabolic equations of the divergence type," *U.S.S.R. Computational Math. and Math. Phys.*, v. 6, 1966, pp. 74-88.
3. PETER LAX, "Shock waves and entropy," in *Contributions to Nonlinear Functional Analysis* (E. H. Zarantonello, Ed.), Academic Press, New York, 1971.
4. RICHARD S. VARGA, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1962.