INVARIANT REGIONS FOR SYSTEMS OF CONSERVATION LAWS

BY

DAVID HOFF[1]

ABSTRACT. We describe necessary and sufficient conditions for a region in
$\mathbf{R}^n$ to be invariant for (Glimm) solutions of the system of n conservation laws
$u_t + f(u)_x = 0$. We also make some observations about the invariance of such
regions for certain finite difference approximations of solutions of systems of
conservation laws.

1. Introduction. In this paper we discuss invariant regions for systems of
conservation laws

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

$$(1.2) \qquad u(x,0) = u_0(x).$$

Here $u = u(x,t) \in \mathbf{R}^n$, f is a smooth function from an open set V of $\mathbf{R}^n$ into $\mathbf{R}^n$,
and u_0 has bounded total variation with values in V. Thus $S \subseteq V$ is invariant
if, whenever $u_0(x) \in S$ for all x, then the solution $u(x,t)$ of (1.1)–(1.2) is in S
for all (x,t). In addition, we make some preliminary observations concerning the
invariance of such sets for certain finite difference approximations of solutions of
(1.1)–(1.2).[1]

We remark that, when there exists an invariant set S for (1.1) in which the
Jacobian $f'(u)$ is bounded, then the results of Nishida-Smoller [9] and Hoff-Smoller
[5] become applicable to establish the existence of global smooth solutions of the
regularized problem

$$u_t + f(u)_x = \varepsilon u_{xx}, \qquad u(x,0) = u_0(x) \in S.$$

And when such a set S is also invariant for a given finite difference scheme, then
mesh-ratio conditions for that scheme can be achieved a priori, thus obviating the
need to calculate sound speeds locally and adjusting the time step repeatedly.

§1 contains some preliminary remarks concerning definitions and notation. Our
main results characterizing invariant regions are presented in §2. We first give
necessary and sufficient conditions for the invariance of a set S in terms of the
geometry of ∂S and the behavior of the shock curves for (1.1). In the case that (1.1)
is genuinely nonlinear, and either $n = 2$ or all states are sufficiently close, we refine
these conditions to the simple requirements that S be convex and that the normal
to ∂S be a left eigenvector of f'. We also give several examples, including one which
shows that, in general, invariant sets need not be convex. In §4 we consider the
question of whether such sets are also invariant for various finite difference schemes.
We show that, as a consequence of our results in §3, a convex invariant set is also

invariant for the Lax-Friedrichs and Godunov schemes. For Osher's scheme (see [10]) we give an example in which an invariant set for (1.1) is in fact not invariant for the difference approximations, and we propose a modification of the scheme which remedies this difficulty.

Our work is motivated by, and to some degree an extension of, earlier work of Chueh, Conley and Smoller [2]. These authors consider regions which are a priori invariant for smooth solutions of systems similar to (1.1), but with reaction and parabolic diffusion terms included. By contrast, our results characterize regions which are a posteriori invariant for certain constructed weak solutions of (1.1)–(1.2). Not surprisingly, the requirements for a set to be invariant are somewhat similar in the two cases. But there are notable differences, both in the results and in the methods of proof.

2. Preliminary remarks and notation. It is well known that global smooth solutions of (1.1) do not exist in general, no matter how smooth the initial data u_0 is. But weak solutions, which are defined in the obvious way, are known to be nonunique. It is therefore necessary to append to (1.1)–(1.2) extra conditions ("entropy conditions") which presumably select the correct, physical weak solutions. It has been shown by Glimm [3] that, for the systems of interest here, global solutions of (1.1)–(1.2) do exist, provided that the initial data u_0 has sufficiently small total variation. And as observed by Lax in [8], Glimm's solutions do satisfy all relevant entropy conditions. For our purposes, therefore, a set S will be invariant if it is invariant for the weak solutions constructed by Glimm.

The Glimm solutions are obtained as limits in L^1_{loc} of approximate solutions constructed by solving Riemann problems locally in $x - t$ space. By a Riemann problem we mean the system (1.1) together with initial data of the form

$$(2.1) \qquad u_0(x) = \begin{cases} u_l, & x < 0, \\ u_r, & x > 0. \end{cases}$$

It follows that a set S is invariant for the Glimm solutions of (1.1)–(1.2) if and only if it is invariant for Riemann problems (1.1)–(2.1). The solution of the Riemann problem was constructed by Lax in [7], and it will therefore be necessary for us to review this solution in some detail.

We now describe our assumptions about the flux f appearing in (1.1). First, we assume there is an open set V in $\mathbf{R}^n$ such that, for u in V, the matrix $f'(u)$ has distinct real eigenvalues $\lambda_1(u) < \cdots < \lambda_n(u)$ with corresponding right and left (column) eigenvectors $r_i(u)$ and $l_i(u)$. In addition, we assume that each characteristic field is either linearly degenerate, $\nabla \lambda_i^t r_i \equiv 0$, or genuinely nonlinear, $\nabla \lambda_i^t r_i \neq 0$. In the latter case we may normalize r_i by requiring that $\nabla \lambda_i^t r_i \equiv 1$, and in every case we choose the l_i so that

$$(2.2) \qquad r_i^t l_j = \delta_{ij}.$$

The precise result of Lax, then, is that, when u_l and u_r are points of V which are sufficiently close, there is a unique solution of the Riemann problem (1.1)–(2.1) of a certain type, which we shall describe below. Finally, as a matter of convenience, we always take V to be convex.

In addition to the above assumptions and normalizations, we assume throughout that S is closed and connected in V and ∂S is smooth and connected. We then

say that S is *invariant* if, whenever u_l and u_r are in S and the Lax solution $u(x,t)$ of the corresponding Riemann problem (1.1)–(2.1) exists, then $u(x,t) \in S$ for all (x,t). And S will be *locally invariant* if the above condition holds for u_l and u_r sufficiently close (which may be closer than required for the existence of a solution).

Given sufficiently close points u_l and u_r of V, the solution $u(x,t)$ of the corresponding Riemann problem may be described as follows. u has the constant value v_i in a sector $a_i \le x/t \le b_i$, where $u_l = v_0$ and $u_r = v_n$; and the sector $b_{i-1} \le x/t \le a_i$ is an "i-wave" of one of three types. First, an "i-shock" occurs when $b_{i-1} = a_i = s$ and v_{i-1} and v_i satisfy the Rankine-Hugoniot condition

$$(2.3) \qquad f(v_i) - f(v_{i-1}) = s(v_i - v_{i-1})$$

and the entropy condition

$$(2.4) \qquad \lambda_i(v_{i-1}) > s > \lambda_i(v_i).$$

Second, an "i-rarefaction wave" occurs when $b_{i-1} = \lambda_i(v_{i-1}) \le \lambda(v_i) = a_i$, and for $b_{i-1} \le x/t \le a_i$, $u(x,t) = w(x/t)$, where $\dot{w}$ is parallel to $r_i(w)$. i-shocks and i-rarefaction waves occur only when the ith characteristic field is genuinely nonlinear. In the linearly degenerate case the i-wave is a contact discontinuity, characterized by (2.3) and the condition $\lambda_i(v_{i-1}) = \lambda_i(v_i)$.

In addition, there is the following geometrical structure in the relationship between v_{i-1} and v_i. If we fix $v_{i-1} \in V$, then there is a half curve $v_i(\varepsilon)$, called the i-shock curve, parameterized by $\varepsilon = \lambda_i(v_i) - \lambda_i(v_{i-1}) \le 0$, of states v_i for which (2.3) and (2.4) hold. The resulting shock speeds s then satisfy

$$(2.5) \qquad s = (\lambda_i(v_{i-1}) + \lambda_i(v_i))/2 + O(\varepsilon^2).$$

The states v_i which are connected to v_{i-1} by an i-rarefaction curve are precisely the points on the integral curve of r_i through v_{i-1} parameterized by the same ε, for $\varepsilon \ge 0$. The joined i-shock and i-rarefaction curves then comprise a curve which is globally C^2, called the "i-wave curve," and which we denote by $\phi_i^l(v_{i-1}, \varepsilon)$. In the linearly degenerate case, $\phi_i^l(v_{i-1}, \varepsilon)$ is simply the integral curve of r_i through v_{i-1}, parameterized as before. Finally, $\phi_i^r(v_i, \varepsilon)$ consists of those states v_{i-1} which are connected to v_i on the left by a wave of the appropriate type.

This structure implies the following technical fact, which will be used in §3.

LEMMA 2.1. *Let* $u_l = v_0, \ldots, v_n = u_r$ *be the states in the solution of the Riemann problem* (1.1)–(2.1). *Then*

$$\left. \frac{dv_i}{du_r} \right|_{u_r = u_l} = \sum_{k=1}^{i} r_k(u_l) l_k(u_l)^{\mathrm{t}}.$$

PROOF. From the above discussion, we have

$$(\partial \phi_i^l / \partial \varepsilon)(v_{i-1}, \varepsilon)\big|_{\varepsilon=0} = r_i(v_{i-1}),$$

and, since $\phi_i^l(v_{i-1}, 0) = v_{i-1}$,

$$\partial \phi_i^l / \partial v_{i-1}\big|_{\varepsilon=0} = I.$$

Now fix $u_l = v_0$, let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$, and define states $v_i(\varepsilon)$ by $v_i = \phi_i^l(v_{i-1}, \varepsilon_i)$. Then since $v_i|_{\varepsilon=0} = u_l$, we have

$$\left. \frac{\partial v_i}{\partial \varepsilon_j} \right|_{\varepsilon=0} = \begin{cases} r_j(u_l), & j \le i, \\ 0, & j > i. \end{cases}$$

That is,

$$(2.6) \qquad dv_i/d\varepsilon|_{\varepsilon=0} = [r_1, \ldots, r_i, 0, \ldots 0].$$

In particular, $dv_n/d\varepsilon|_{\varepsilon=0}$ is an invertible matrix, so that the map $\varepsilon \to v_n(\varepsilon)$ is invertible near $\varepsilon = 0$. (This proves that (1.1)–(2.1) is solvable when $v_n = u_r$ is close to u_l.) In addition, the normalization (2.2) implies that

$$(2.7) \qquad \frac{d\varepsilon}{dv_n}\bigg|_{v_n=u_l} = \begin{bmatrix} l_1^t \\ \vdots \\ l_n^t \end{bmatrix}.$$

Combining (2.6) and (2.7) we obtain

$$\frac{dv_i}{du_r}\bigg|_{u_r=u_l} = \frac{dv_i}{d\varepsilon}\frac{d\varepsilon}{dv_n} = \sum_{k=1}^{i} r_k l_k^t,$$

as required. $\quad\square$

3. Invariant regions.

In this section we present our main results characterizing invariant regions. We assume that V is an open convex set in $\mathbf{R}^n$ in which the hypotheses of §2 hold, and the set S, which is closed and connected in V, has a smooth connected boundary ∂S relative to V.

The following lemma gives a necessary condition that S be invariant.

LEMMA 3.1. *Assume that S is invariant. Then there is an index p such that, at each point of ∂S, the normal vector to ∂S is parallel to the eigenvector l_p.*

PROOF. If $u_l \in \partial S$ there is a neighborhood B of u_l and a C^2 function $g\colon B \to \mathbf{R}$ such that $B \cap S = \{u \in B\colon g(u) \leq 0\}$. Let $u_r(s)$ be an arbitrary smooth curve on ∂S with $u_r(0) = u_l$, and let $u_l = v_0(s), \ldots, v_n(s) = u_r(s)$ be the states in the solution of the Riemann problem with data $(u_l, u_r(s))$. Our hypothesis is that $v_i(s) \in S$ for each i and s, so that $g(v_i(s)) \leq 0$. That is,

$$g(u_l) + s\nabla g(u_l) + \left[dv_i/du_r|_{u_r=u_l}\, \dot{u}_r(0)\right] + O(s^2) \leq 0.$$

But since $g(u_l) = 0$, the coefficient of s must therefore also be zero. And because $\dot{u}_r(0)$ is an arbitrary vector in the tangent space T of ∂S at u_l, we may conclude that $dv_i/du_r|_{u_r=u_l}$ maps T into itself. Therefore $(dv_i/du_r)^t$ maps $T^\perp$ into $T^\perp$. But $T^\perp = \mathrm{span}(\nabla g)$, so that

$$(dv_i/du_r)^t\nabla g = c_i\nabla g$$

for some scalar c_i. Applying Lemma 2.1, we may rewrite this as

$$\sum_{k=1}^{i} (r_k^t\nabla g)l_k = c_i\nabla g.$$

Now choose p minimal such that $c_p \neq 0$. Then

$$c_p\nabla g = (c_p - c_{p-1})\nabla g = (r_p^t\nabla g)l_p.$$

That is, the normal vector ∇g to ∂S is parallel to l_p at u_l. p is constant throughout ∂S because ∂S is smooth and connected.
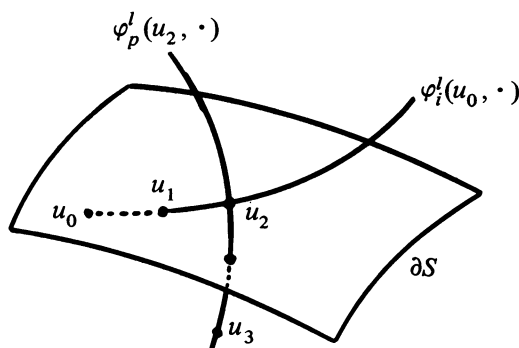
FIGURE 1

We remark that, when S satisfies the conclusion of Lemma 2.1 above, then, in view of (2.2), integral curves of r_i for $i \neq p$ which intersect ∂S must in fact be contained in ∂S. Therefore, i-rarefaction curves or i-contact discontinuity curves through points $u_l \in \partial S$ (S) remain in ∂S (S).

We can now give a complete characterization of invariant sets in terms of the behavior of the shock curves $\phi_i^l(u_0, \varepsilon)|_{\varepsilon \leq 0}$.

THEOREM 3.2. S is invariant if and only if the following two conditions hold:

(a) There is a p such that, at each point u_0 of ∂S, $l_p(u_0)$ is parallel to the normal to ∂S.

(b) If $i < p$ $(i > p)$ and if the ith characteristic field is genuinely nonlinear, then $\phi_i^l(u_0, \varepsilon) \in S$ $(\phi_i^r(u_0, \varepsilon) \in S)$ for all $u_0 \in \partial S$ and all $\varepsilon < 0$ $(\varepsilon > 0)$.

PROOF. If S is invariant, then (a) holds by Lemma 3.1. Assume that (b) fails for some $i < p$. Then there is a point $u_0 \in \partial S$ such that the curve $\phi_i^l(u_0, \varepsilon)|_{\varepsilon < 0}$ leaves S, say at $u_1 = \phi_i^l(u_0, \varepsilon_1)$. Now $(d/d\varepsilon)\phi_p^l(u_1, \varepsilon)|_{\varepsilon=0} = r_p(u_1)$, and $r_p(u_1)^t l_p(u_1) = 1$, so that the curve $\phi_p^l(u_1, \varepsilon)$ is transverse to ∂S at $\varepsilon = 0$. By perturbing u_1 to u_2 along the curve $\phi_i^l(u_0, \cdot)$ (Figure 1), we obtain a point $u_2 = \phi_i^l(u_0, \varepsilon_2) \notin S$ such that $\phi_p^l(u_2, \cdot)$ contains a point $u_3 \in S$. The Riemann problem with data $u_l = u_0$ and $u_r = u_3$ is therefore solved with an i-shock connecting u_0 and u_2, and a p-wave connecting u_2 and u_3. However, both u_0 and u_3 are in S but u_2 is not, thus violating invariance. The proof of (b) for $i > p$ is similar.

Now assume that (a) and (b) hold. We show first that (b) is satisfied by all points u_0 of S. If not, then there are points $u_0 \in \text{int}(S)$ and $u_1 \in \phi_i^l(u_0, \cdot) - S$. But this shows that the curve $\phi_i^r(u_1, \cdot)$ enters $\text{int}(S)$, so that there is a point $u_2 \in \phi_i^r(u_1, \cdot) \cap \partial S$. But then $u_1 \in \phi_i^l(u_2, \cdot) - S$, violating (b), since $u_2 \in \partial S$.

Now let u_r and u_l be points of S for which the solution of the corresponding Riemann problem contains states $u_l = v_0, \ldots, v_n = u_r$. We claim that if $v_{k-1} \in S$ and $k < p$, then $v_k \in S$. To prove this we simply observe, by the remarks following Lemma 3.1 and by hypothesis (b), that the entire curve $\phi_k^l(v_{k-1}, \cdot)$ is contained in S. Since v_k is a point on this curve, v_k is in S. We therefore have by induction that the points $v_0, \ldots, v_{p-1}$ are in S. Proceeding similarly from the fact that $v_n = u_r$ is in S and using the other half of (b), we find that the points $v_n, \ldots, v_p$ are in S.

To complete the proof we need to show that, if v_{p-1} and v_p are connected by a p-rarefaction wave, then all the intermediate states in that wave are points of

S. Suppose then that the p-rarefaction curve $\phi_p^l(v_{p-1}, \cdot)|_{\varepsilon \geq 0}$ leaves S, say at u_1, and then reenters S at u_2. Then since $(d/d\varepsilon)\phi_p^l(v_{p-1}, \varepsilon) = r_p(\phi_p^l)$ for $\varepsilon \geq 0$, the numbers $r_p(u_1)^t l_p(u_1)$ and $r_p(u_2)^t l_p(u_2)$ have opposite signs. But this implies the vanishing of $r_p^t l_p$ at some point of V, which violates (2.2).

The simplest situation is that in which S is a half space (intersected with V). In this case a certain uncoupling of the shock relation (2.3) allows us to simplify the criteria for S to be an invariant set.

COROLLARY 3.3. *Suppose that S is the intersection of V with a half space. Then Theorem 3.2(a) is both necessary and sufficient for S to be an invariant set.*

PROOF. We need to show that, under these hypotheses, Theorem 3.2(b) is a consequence of (a). We assume in this proof that the normal to ∂S is the *constant* left eigenvector l_p (so that the normalization (2.2) may no longer be satisfied).

The proof consists of showing that, for $u_0 \in \partial S$ and $i \neq p$, the i-shock curve through u_0 in fact lies on the plane ∂S. To see this, we map the plane $\{v_n = 0\}$ in $\mathbf{R}^n$ onto (the global extension of) ∂S by an affine map $v \mapsto Av + u_0$, and we let

$$g(v) = A^{-1}[f(Av + u_0) - f(u_0)].$$

Then the Rankine-Hugoniot condition

$$f(u) - f(u_0) = s(u - u_0)$$

is equivalent to the relation

$$(3.1) \qquad\qquad g(v) = sv.$$

Now, since A maps the plane $\{v_n = 0\}$ onto a plane whose normal is l_p, A^t must map l_p to $e_n = (0, \ldots, 0, 1)$. And since

$$dg/dv = A^{-1}(df/du)A,$$

we see that e_n must be a left eigenvector of $dg/dv|_{v_n=0}$. Therefore, on $\{v_n = 0\}$, dg/dv has the structure

$$(3.2) \qquad\qquad \frac{dg}{dv} = \left[\begin{array}{c|c} B & * \\ \hline 0 & \lambda_p \end{array}\right].$$

We let $v = (\bar{v}, v_n)$ and $g = (\bar{g}, g_n)$, where $\bar{v}, \bar{g} \in \mathbf{R}^{n-1}$. Evidently $B = (d\bar{g}/d\bar{v})(\bar{v}, 0)$ has $n-1$ distinct eigenvalues, so that the reduced Rankine-Hugoniot condition

$$(3.3) \qquad\qquad \bar{g}(\bar{v}, 0) = s\bar{v}$$

has $n-1$ curves of solutions through $\bar{v} = 0$. However, (3.2) shows that $(\partial g_n/\partial \bar{v})(\bar{v}, 0) = 0$, so $g_n(\bar{v}, 0) = g_n(0, 0) = 0 = s \cdot 0$. Thus if $\bar{v}$ and s solve (3.3), then in fact $g(\bar{v}, 0) = s(\bar{v}, 0)$, which is (3.1). We have therefore shown that there are $n-1$ curves of solutions of (3.1) all lying in the plane $\{v_n = 0\}$. As remarked earlier, the images of these curves under the map $v \mapsto Av + u_0$ are the $n-1$ i-shock or i-contact discontinuity curves for $i \neq p$, which are therefore seen to lie in ∂S.

EXAMPLE 3.4. We consider the following system of conservation laws, which arises in chemical chromatography (see [1, p. 36]):

$$(3.4) \qquad\qquad \frac{\partial}{\partial t}(\delta n_i + c_i) + A\frac{\partial c_i}{\partial x} = 0, \qquad i = 1, \ldots, n.$$

Here n_i and c_i are concentrations which depend on x and t, and δ and A are positive constants. The system becomes closed under an additional hypothesis of the form

$$n_i = \frac{NK_ic_i}{1 + \sum_j K_jc_j},$$

where N and K_i are positive constants. The conserved quantities are then

$$(3.5) \qquad u_i = \delta n_i + c_i = c_i\left[1 + \frac{\delta NK_i}{1 + \sum_j K_jc_j}\right].$$

We first observe that $u = (u_1, \ldots, u_n)$ is a locally invertible function of $c = (c_1, \ldots, c_n)$. To see this, we compute

$$\frac{\partial u_i}{\partial c_i} = 1 + \frac{\delta NK_i(1 + \sum_{j \neq i} K_jc_j)}{(1 + \sum_j K_jc_j)^2}$$

and

$$(3.6) \qquad \frac{\partial u_i}{\partial c_j} = -\frac{\delta NK_iK_jc_i}{(1 + \sum_j K_jc_j)^2}.$$

Thus

$$\frac{\partial u_i}{\partial c_i} - \sum_{j \neq i}\left|\frac{\partial u_j}{\partial c_i}\right| = 1 + \frac{\delta NK_i}{(1 + \sum_l K_lc_l)^2},$$

which is positive in the nonnegative octant of c-space. Thus $(du/dc)^t$ is strictly diagonally dominant, so that dc/du is an invertible matrix in a neighborhood V of the nonnegative octant of c-space.

Scaling out the constant A, we rewrite (3.4) in the form

$$\partial u/\partial t + \partial c(u)/\partial x = 0.$$

Now, c_i is a concentration, so the set $\{c_i \geq 0\}$ should be invariant. By (3.5) the invariance of $\{c_i \geq 0\}$ is equivalent to that of $\{u_i \geq 0\}$. Applying Corollary 3.3, we have only to check that, when $u_i = 0$, the standard basis vector e_i is a left eigenvector of dc/du or equivalently, du/dc. But this follows from (3.6), for when $u_i = 0$ $(c_i = 0)$, $\partial u_i/\partial c_j = 0$ for $j \neq i$, so that the ith row of du/dc is a multiple of e_i, as required. And since the intersection of invariant sets is invariant, we thus have that the nonnegative octant of u-space is invariant for (3.4).

In general, when S is not a half space, the shock curves may be difficult or impossible to compute, so that condition (b) of Theorem 3.2 cannot be checked. As we shall see, however, (b) is closely related to the convexity of S. To motivate this result, we shall apply Theorem 3.2 to compute (well-known) invariant sets for the equations of isentropic gas dynamics.

EXAMPLE 3.5. Consider the p-system

$$v_t - u_x = 0, \qquad u_t + p(v)_x = 0$$

(v and u are now scalars), where p satisfies $p'(v) < 0 < p''(v)$ for $v > 0$. One easily computes that in the set $V = \{(v, u): v > 0\}$, the eigenvalues of $\partial(-u, p(v))/\partial(u, v)$ are $\pm\lambda(v) = \pm\sqrt{-p'(v)}$, with corresponding (unnormalized) right and left eigenvectors

$$r_\pm = \begin{bmatrix} \mp 1 \\ \lambda \end{bmatrix} \quad \text{and} \quad l_\pm = \begin{bmatrix} \lambda \\ \mp 1 \end{bmatrix}.$$

Now, if S is to be an invariant set, then by Theorem 3.2(a), ∂S must be an integral curve of r_- or r_+. The integral curves of r_- are given by

$$(3.7) \qquad\qquad u = u_0 + \int_{v_0}^{v} \lambda(s)\, ds.$$

There are thus two candidates for a corresponding invariant set: namely, the regions above and below the curve (3.7). To check condition (b) of Theorem 3.2, we compute the 1-shock curve through (v_0, u_0) from the Rankine-Hugoniot condition (2.3):

$$(3.8) \qquad\qquad u = u_0 - \sqrt{-[p(v) - p(v_0)](v - v_0)}, \qquad v < v_0.$$

(The requirement that $v < v_0$ is dictated by the entropy condition (2.4).) It is easy to see that the 1-shock curve (3.8) lies below the integral curve (3.7), for when $v < v_0$,

$$\int_{v}^{v_0} \sqrt{-p'(s)}\, ds \le \left[\left(\int_{v}^{v_0} -p'(s)\, ds \right) \left(\int_{v}^{v_0} ds \right) \right]^{1/2}$$
$$= \sqrt{-[p(v) - p(v_0)](v - v_0)}.$$

This shows that the *lower* region

$$S = \left\{ (v, u) \colon\; u \le u_0 + \int_{v_0}^{v} \lambda(s)\, ds \right\}$$

is in fact an invariant region. The condition $p'' > 0$ implies that S is the *convex* component of $V - \partial S$. (If p were to satisfy p', $p'' < 0$ instead, then the 1-shock curve would still be given by (3.8), but for $v > v_0$, and the result would be the same: the invariant set S would be convex.)

In a similar way, starting from an integral curve of r_+, we can show that the sets

$$S = \left\{ (v, u) \colon\; u \ge u_0 - \int_{v_0}^{v} \lambda(s)\, ds \right\}$$

are also invariant. Finally, by taking the intersection of two invariant sets, one from each family, we obtain an invariant set in which v, the reciprocal of density, is bounded below, and in which, therefore, the matrix f' is bounded.

It was no accident that the i-shock emanating from $(v_0, u_0) \in \partial S$ broke into the *convex* component of $V - \partial S$. In fact, this phenomenon generalizes, and allows us to replace condition (b) of Theorem 3.2, in many cases of interest, by the simple requirement that S be convex.

THEOREM 3.6. *Assume that all fields are genuinely nonlinear.*
(a) *If S is invariant, then S must be convex.*
(b) *If S satisfies Theorem 3.2(a) and is convex, then S is locally invariant.*

The proof of Theorem 3.6 is rather long and so will be deferred to the end of this section. It is easy to see that the genuine nonlinearity hypothesis is necessary here. For suppose that, for a system of two conservation laws, the first characteristic field is linearly degenerate. Take C to be any integral curve of r_1. (C is therefore a contact discontinuity curve.) Then for either side S_- or S_+ of C, of Theorem 2.2(a) is satisfied by construction, and (b) is vacuous. Thus both sides S_- and S_+

are invariant sets; but unless C is a line, they cannot both be convex. A specific example is provided by the following system:

$$u_t + (e^u + e^v)_x = 0, \qquad v_t = 0.$$

One can easily apply Theorem 3.2 to this system to show that the set

$$S = \{(u, v): u < 0 \text{ and } v \geq \ln(1 - e^u)\}$$

is a nonconvex invariant set in $V = \mathbf{R}^2$.

When $n = 2$ the result of Theorem 3.6 can be sharpened as follows:

COROLLARY 3.7. *Assume $n = 2$ and both characteristic fields are genuinely nonlinear. Then S is invariant if and only if S is convex and satisfies Theorem 3.2(a).*

PROOF. In light of Theorems 3.2 and 3.6, we have only to show that, if S is convex and satisfies Theorem 3.2(a), then the shock curves satisfy Theorem 3.2(b) globally. Thus let $u_0 \in \partial S$ and let ∂S be the curve $u(\varepsilon)$ parameterized by $\varepsilon = \lambda(u(\varepsilon)) - \lambda(u_0)$ and such that $\dot{u}(\varepsilon) = r(u(\varepsilon))$. ($r$ is the right eigenvector of f' corresponding to the eigenvalue λ.) Now, by Theorems 3.2 and 3.6, the shock curve $\phi^l(u_0, \cdot)|_{\varepsilon < 0}$ of the same family remains in S near u_0. We claim that if this shock curve intersects ∂S at a point u_1 other than u_0, then the shock curve and ∂S are coinciding straight lines between u_0 and u_1. To see this, we have from the Rankine-Hugoniot condition that

$$0 = f(u_1) - f(u_0) - s(u_1 - u_0) = \int_0^{\lambda(u_1) - \lambda(u_0)} [f'(u(\varepsilon)) - s]\dot{u}(\varepsilon)\, d\varepsilon$$

$$= \int [\lambda(u(\varepsilon)) - s]\dot{u}(\varepsilon)\, d\varepsilon = \int [\varepsilon + \lambda(u_0) - s]\dot{u}(\varepsilon)\, d\varepsilon$$

$$= [\lambda(u_1) - s]u_1 - [\lambda(u_0) - s]u_0 - \int u(\varepsilon)\, d\varepsilon.$$

Thus

$$\frac{1}{\lambda(u_0) - \lambda(u_1)} \int_0^{\lambda(u_0) - \lambda(u_1)} u(\varepsilon)\, d\varepsilon$$

$$= \frac{s - \lambda(u_1)}{\lambda(u_0) - \lambda(u_1)} u_1 + \frac{\lambda(u_0) - s}{\lambda(u_0) - \lambda(u_1)} u_0.$$

Since the entropy condition requires that $\lambda(u_0) > s > \lambda(u_1)$, we thus have that an integral average of points on ∂S is a point on the line segment $[u_0, u_1]$ joining u_0 and u_1. Thus $u(\varepsilon) = \partial S$ is in fact a straight line between u_0 and u_1, and we may take $\dot{u}(\varepsilon) = r$ to be constant. But then

$$(3.9) \qquad f(u(\varepsilon)) - f(u_0) = \int_0^\varepsilon f'(u(\varepsilon))\dot{u}(\varepsilon)\, d\varepsilon$$

$$= \left(\int_0^\varepsilon \lambda(u(\varepsilon))\, d\varepsilon \right) r = \left(\frac{1}{\varepsilon} \int_0^\varepsilon \lambda \right) [u(\varepsilon) - u_0],$$

so that this line segment coincides with the shock curve, as claimed.

To prove the corollary, we need to show that the shock curve $\phi^l(u_0, \cdot)|_{\varepsilon < 0}$ remains in S as long as it is defined. If not, then by Theorem 3.6 and the above claim, there is a minimal (negative) ε_1 such that $\phi^l(u_0, \cdot)|_{\varepsilon < 0}$ coincides with ∂S up to $u_1 = u(\varepsilon_1)$
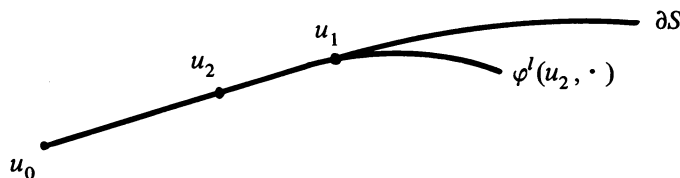
FIGURE 2

(Figure 2). Evidently $u(\varepsilon)$ is not linear on any interval $(\varepsilon, \varepsilon_1]$. Let $\mathcal{A}$ be the set of all u on the line segment $[u_0, u_1]$ such that $\phi^l(u, \cdot)|_{\varepsilon<0}$ is globally contained in S. Then $\mathcal{A}$ is clearly closed, and $u_1 \in \mathcal{A}$ by Theorem 3.6 and our claim above. But $\mathcal{A}$ is also open in $[u_0, u_1]$, for if $u_2 \in \mathcal{A}$, then (3.9) shows that the line segment $[u_2, u_1]$ is contained in $\phi^l(u_2, \cdot)|_{\varepsilon<0}$, say $u_1 = \phi^l(u_2, \varepsilon_1')$. Then since $u_2 \in \mathcal{A}$, we must have that, for $\varepsilon < \varepsilon_1'$, $\phi^l(u_2, \varepsilon) \in \mathrm{int}(S)$ (since ∂S is not a line past u_1). By continuity and our claim above, the shock curves $\phi^l(u, \cdot)|_{\varepsilon<0}$ for u near u_2 also enter $\mathrm{int}(S)$, and therefore will have precisely the same behavior. That is, such nearby points are also in $\mathcal{A}$. Thus $\mathcal{A} = [u_0, u_1]$, and the shock curve $\phi^l(u_0, \cdot)$ remains in S globally.

We can now work out explicitly the invariant sets S for a system of two genuinely nonlinear conservation laws. First, by Theorem 3.2(a), the boundary ∂S must be an integral curve of a right eigenvector field, say of r_j. Then ∂S is locally the solution set of an equation $g(u) = 0$, where $g: \mathbf{R}^2 \to \mathbf{R}^2$, $\nabla g = \beta l_p$, β is a scalar, and $j \neq p$. By the above corollary, S is then invariant if only if S is convex. And since ∂S is smooth, one of the components S of $V - \partial S$ will be convex provided that $r_j^t g'' r_j$ is of one sign on ∂S (see the remarks in the proof of Theorem 3.6(a) below). This latter condition can be computed explicitly in terms of f as follows. Since ∇g is a left eigenvector of

$$f' = \begin{bmatrix} \nabla f_1^t \\ \nabla f_2^t \end{bmatrix},$$

we have

$$\lambda_p \nabla g^t = \nabla g^t \begin{bmatrix} \nabla f_1^t \\ \nabla f_2^t \end{bmatrix}.$$

Differentiating in the direction r_j, we then obtain

$$(\nabla \lambda_p \cdot r_j) \nabla g^t + \lambda_p r_j^t g'' = r_j^t g'' f' + \nabla g^t \begin{bmatrix} r_j^t f_1'' \\ r_j^t f_2'' \end{bmatrix}.$$

Now take the dot product of both sides with r_j:

$$\lambda_p r_j^t g'' r_j = \lambda_j r_j^t g'' r_j + \nabla g^t \begin{bmatrix} r_j^t f_1'' r_j \\ r_j^t f_2'' r_j \end{bmatrix},$$

so that

$$r_j^t g'' r_j = \frac{\beta}{\lambda_p - \lambda_j} l_p^t \begin{bmatrix} r_j^t f_1'' r_j \\ r_j^t f_2'' r_j \end{bmatrix}.$$

Since β and $\lambda_p - \lambda_j$ are of one sign, we conclude that one of the sides S of an integral curve of r_i will be an invariant region provided that, along this integral curve, the quantity

$$l_p^t \begin{bmatrix} r_j^t f_1'' r_j \\ r_j^t f_2'' r_j \end{bmatrix}$$

is of one sign. Systems of two conservation laws which satisfy this condition globally are precisely those in the Johnson-Smoller class (see [6]).

We do not know whether Theorem 3.2(a) and convexity are sufficient to insure invariance in the large when $n \geq 3$. It is worth observing that a counterexample here would then be an example for which parabolic regularization fails. For suppose that, for some system of three or more conservation laws, there is a convex set S satisfying Theorem 3.2(a), but which is not invariant for Riemann problems. If then u_l and u_r are points of S for which the solution $u(x,t)$ of the corresponding Riemann problem contains a state not in S, then $u(x,t)$ cannot be the limit (in L_{loc}^1, say) as $\varepsilon \to 0$ of solutions $u^\varepsilon(x,t)$ of

$$(3.10) \qquad\qquad u_t + f(u)_x = \varepsilon u_{xx}.$$

This follows because, by the results of [2], S is in fact invariant for solutions of (3.10), so that $u^\varepsilon(x,t) \in S$ for all ε, x, t.

The question of the global invariance of such sets S (i.e., convex sets satisfying Theorem 3.2(a)) reduces by Theorems 3.2 and 3.6 to the question of whether or not i-shock curves originating on ∂S remain in S globally. The following result gives sufficient conditions for an affirmative answer.

THEOREM 3.8. *Suppose S satisfies Theorem 3.2(a). Let $i < p$ and assume that the ith characteristic field is genuinely nonlinear. Then if $u(\varepsilon)$ is the i-shock curve through $u_l \in \partial S$, $u(\varepsilon)$ will be contained in S globally provided that the following conditions hold:*

(a) *S is strictly convex in the sense that, when $u_1, u_2 \in \partial S$, $(u_2 - u_1)^t n(u_2) > 0$ (n is the outer normal to ∂S).*

(b) *$s(u_l, u(\varepsilon)) < \lambda_p(u(\varepsilon))$ (s is the shock speed).*

(c) *$(d/d\varepsilon)s(u_l, u(\varepsilon)) > 0$.*

(*The parameterization of $u(\varepsilon)$ is taken so that ε decreases away from u_l.*)

PROOF. Suppose that the i-shock $u(\varepsilon)$ intersects ∂S at $u_r = u(\varepsilon_0)$, and let S be given by $\{u: g(u) \leq 0\}$ near u_r. Then setting $s(\varepsilon) = s(u_l, u(\varepsilon))$, we have, from (2.3),

$$[f'(u(\varepsilon)) - s(\varepsilon)]\dot{u}(\varepsilon) = \dot{s}(\varepsilon)[u(\varepsilon) - u_l].$$

Letting $\varepsilon = \varepsilon_0$ and multiplying by $\nabla g(u_r)$, we obtain

$$(\lambda_p - s)\nabla g^t \dot{u} = \dot{s}\nabla g^t(u_r - u_l);$$

that is,

$$\frac{d}{d\varepsilon}g(u(\varepsilon))|_{\varepsilon=\varepsilon_0} = \frac{\dot{s}}{\lambda_p - s}\nabla g^t(u_r - u_l).$$

Our hypotheses (a)–(c) together with the fact ∇g is a positive multiple of n show that $\dot{g}$ is strictly positive. Since ε decreases along the shock curve, we conclude that $u(\varepsilon)$ returns to int(S) past u_r. $\square$
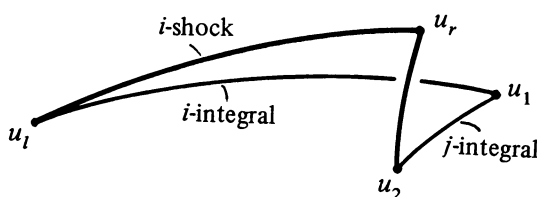
FIGURE 3

PROOF OF THEOREM 3.6. Suppose S satisfies Theorem 3.2(a). We fix a point $u_l \in \partial S$ and let u_r be a nearby point on the i-shock curve through u_l, where $i < p$. If u_r is sufficiently close to u_l, the implicit function theorem applies to show that u_l can be connected to a point u_1 by an integral curve of r_i and u_r to a point u_2 by an integral curve of r_p, in such a way that u_1 and u_2 are connected by a sequence of integral curves of the fields r_j for $j \neq i, p$ (Figure 3). Moreover, since the i-shock curve and the integral curve of r_i differ by $O(\varepsilon^3)$ when the curves are parameterized by $\varepsilon = \lambda_i(u) - \lambda_i(u_l)$, it is easy to see that

$$(3.11) \qquad u_r - u_2, u_2 - u_1 = O(\Delta\lambda_i^3),$$

where $\Delta\lambda_i = \lambda_i(u_1) - \lambda_i(u_l)$.

Now let ∂S be given locally as the solution set of the equation $g(u) = 0$, where $\nabla g(u) = \beta(u)l_p(u)$ on ∂S and, without loss of generality, $\beta > 0$. We shall show below by Taylor series that, for the above configuration,

$$(3.12) \qquad \underbrace{[\lambda_p(u_2) - s]}_{+}\Delta\lambda_p = \frac{\Delta\lambda_i^3}{12\beta}\underbrace{(r_i^t g'' r_i)|_{u_2}}_{-} + O(\Delta\lambda_i^4),$$

where s is the speed of the shock (u_l, u_r) and $\Delta\lambda_p = \lambda_p(u_r) - \lambda_p(u_2)$. Observe that, in the small, $\lambda_p(u_2) - s > 0$ by the ordering of the characteristic speeds, and $\Delta\lambda_i < 0$ if the shock (u_l, u_r) is entropy-satisfying. Now, since $l_p^t r_p = 1$ and $\beta > 0$, r_p points out of S. And since $\nabla\lambda_p^t r_p = 1$, it follows that $u_r \in S \Leftrightarrow \Delta\lambda_p \leq 0$. Now suppose S is invariant. Then all such points u_r are in S, so the corresponding $\Delta\lambda_p$ are nonpositive. Letting $\Delta\lambda_i \to 0$ in (3.12), we then have that $r_i^t g'' r_i \geq 0$ at u_l for any $u_l \in \partial S$. An easy local calculus argument then shows that each point u_l of ∂S is the center of a ball B such that $B \cap S$ is convex. It then follows that S itself is convex (see [**12**, pp. 48–50]).

To prove the converse, suppose S satisfies Theorem 3.2(a) but is not a locally invariant set. Then by (b) of that theorem there must be a point $u_l \in \partial S$ whose i-shock curve (for $i < p$, say) includes points $u_r \notin S$ arbitrarily close to u_l. For such a point u_r we then have that $\Delta\lambda_p > 0$, so that by (3.12), $r_i^t g'' r_i < 0$ at some point u_2 of ∂S. But this implies that S cannot be convex. Thus convex sets satisfying Theorem 3.2(a) are at least locally invariant.

To prove (3.12) we expand each term in the shock relation

$$(3.13) \qquad l_p(u_2)^t[f(u_r) - f(u_l)] = sl_p(u_2)^t(u_r - u_l)$$

about u_2, making repeated use of (3.10). In the following computations, functions without arguments are understood to be evaluated at u_2. First,

$$u_l = u_1 - \Delta\lambda_i r_i(u_1) + \frac{\Delta\lambda_i^2}{2}(r_i' r_i)|_{u_1} - \frac{\Delta\lambda_i^3}{6}\frac{d}{dr_i}(r_i' r_i)|_{u_1} + O(\Delta\lambda_i^4),$$

so that, by (3.11),

$$(3.14) \qquad u_1 - u_l = \Delta\lambda_i r_i - \frac{\Delta\lambda_i^2}{2} r_i' r_i + \frac{\Delta\lambda_i^3}{6} \frac{d}{dr_i}(r_i' r_i) + O(\Delta\lambda_i^4).$$

Thus

$$(3.15) \qquad l_p^t(u_1 - u_l) = -\frac{\Delta\lambda_i^2}{2} l_p^t r_i' r_i + \frac{\Delta\lambda_i^3}{6} l_p^t \left(\frac{d}{dr_i} r_i' r_i\right) + O(\Delta\lambda_i^4).$$

Next, if v_j are the intermediate states in the curve joining u_2 with u_1 and $\Delta\lambda_j = \lambda_j(v_j) - \lambda_j(v_{j-1})$, then

$$(3.16) \qquad u_2 = u_1 + \sum_{j\neq i,p} \Delta\lambda_j r_j + O(\Delta\lambda_i^4),$$

so that

$$(3.17) \qquad l_p^t(u_2 - u_1) = O(\Delta\lambda_i^4).$$

Finally, if

$$\Delta\lambda_p = \lambda_p(u_r) - \lambda_p(u_2),$$

then

$$(3.18) \qquad u_r = u_2 + \Delta\lambda_p r_p + O(\Delta\lambda_i^4),$$

so that

$$(3.19) \qquad l_p^t(u_r - u_2) = \Delta\lambda_p + O(\Delta\lambda_i^4).$$

Adding (3.15), (3.17), and (3.19), we therefore obtain

$$(3.20) \qquad l_p^t(u_r - u_l) = \Delta\lambda_p - \frac{\Delta\lambda_i^2}{2} l_p^t r_i' r_i + \frac{\Delta\lambda_i 3}{6} l_p^t \left(\frac{d}{dr_i} r_i' r_i\right) + O(\Delta\lambda_i^4).$$

Next, we expand the left side of (3.13). In this computation the vectors $f''(v,v)$ and $f'''(v,v,v)$ will be denoted by $f'' v^2$ and $f''' v^3$. First, from (3.11) and (3.14),

$$f(u_l) = f(u_1) + f'(u_l - u_1) + \tfrac{1}{2} f''(u_l - u_1)^2 + \tfrac{1}{6} f'''(u_l - u_1)^3 + O(\Delta\lambda_i^4)$$

$$= f(u_1) - \lambda_i \Delta\lambda_i r_i + \frac{\Delta\lambda_i^2}{2}(f' r_i' r_i + f'' r_i^2)$$

$$+ \frac{\Delta\lambda_i^3}{6}\left[-f'\frac{d}{dr_i}(r_i' r_i) - 3f''(r_i, r_i' r_i) - f''' r_i^3\right] + O(\Delta\lambda_i^4),$$

so that

$$(3.21) \quad l_p^t[f(u_1) - f(u_l)] = -\frac{\Delta\lambda_i^2}{2}(\lambda_p l_p^t r_i' r_i + l_p^t f'' r_i^2)$$

$$+ \frac{\Delta\lambda_i^3}{6}\left[\lambda_p l_p^t \frac{d}{dr_i}(r_i' r_i) + 3 l_p^t f''(r_i, r_i' r_i) + l_p^t f''' r_i^3\right]$$

$$+ O(\Delta\lambda_i^4).$$

Next, by (3.16),

$$f(u_1) = f(u_2) + f'(u_1 - u_2) + O(\Delta\lambda_i^4)$$

$$= f(u_2) = \sum_{j\neq i,p} \lambda_j \Delta\lambda_i r_j + O(\Delta\lambda_i^4),$$

so that

(3.22) $$l_p^t[f(u_2) - f(u_1)] = O(\Delta\lambda_i^4).$$

Finally, by (3.18),

$$f(u_r) - f(u_2) = f'(u_r - u_2) + O(\Delta\lambda_i^4)$$
$$= \lambda_p \Delta\lambda_p r_p + O(\Delta\lambda_i^4),$$

so that

(3.23) $$l_p^t[f(u_r) - f(u_2)] = \lambda_p \Delta\lambda_p + O(\Delta\lambda_i^4).$$

Now add (3.21)–(3.23) to obtain the expansion of $l_p^t[f(u_r) - f(u_l)]$, and subtract the expansion of $sl_p^t(u_r - u_l)$ (3.15). The result is

(3.24) $$(\lambda_p - s)\Delta\lambda_p + \frac{\Delta\lambda_i^2}{2}[(s - \lambda_p)l_p^t r_i' r_i - l_p^t f'' r_i^2]$$
$$+ \frac{\Delta\lambda_i^3}{6}[(\lambda_p - s)l_p^t \frac{d}{dr_i}(r_i' r_i) + 3l_p^t f''(r_i, r_i' r_i) + l_p^t f''' r_i^2]$$
$$= O(\Delta\lambda_i^4).$$

However, if we differentiate the relation $f' r_i = \lambda_i r_i$ in the r_i direction, we obtain

(3.25) $$f'' r_i^2 + f' r_i' r_i = r_i + \lambda_i r_i' r_i,$$

so that

$$l_p^t f'' r_i^2 = (\lambda_i - \lambda_p)l_p^t r_i' r_i.$$

The coefficient of $\Delta\lambda_i^2/2$ in (3.24) is therefore $(s - \lambda_i)l_p^t r_i' r_i$. Next, differentiate (3.25) in the r_i direction and take l_p^t to obtain

$$l_p^t f''' r_i^3 + 3l_p^t f''(r_i' r_i, r_i) = (\lambda_i - \lambda_p)l_p^t \frac{d}{dr_i}(r_i' r_i) + 2l_p^t r_i' r_i.$$

The coefficient of $\Delta\lambda_i^3/6$ in (3.24) is therefore

$$2l_p^t r_i' r_i + (\lambda_i - s)l_p^t \frac{d}{dr_i}(r_i' r_i) = 2l_p^t r_i' r_i + O(\Delta\lambda_i).$$

(3.24) has thus been simplified to

(3.26) $$(\lambda_p - s)\Delta\lambda_p + \frac{\Delta\lambda_i^2}{2}(s - \lambda_i)l_p^t r_i' r_i + \frac{\Delta\lambda_i^3}{3}l_p^t r_i' r_i = O(\Delta\lambda_i^4).$$

However, by (2.5),

$$s = \frac{\lambda_i(u_l) + \lambda_i(u_r)}{2} + O(\Delta\lambda_i^2) = \frac{\lambda_i(u_l) + \lambda_i}{2} + O(\Delta\lambda_i^2),$$

so that

$$s - \lambda_i = -\Delta\lambda_i/2 + O(\Delta\lambda_i^2).$$

Substituting this into (3.26), we then obtain

(3.27) $$(\lambda_p - s)\Delta\lambda_p + (\Delta\lambda_i^3/12)l_p^t r_i' r_i = O(\Delta\lambda_i^4).$$

Finally, if we differentiate the relation $\nabla g^t r_i = 0$ in the r_i direction and use the fact that $\nabla g = \beta l_p$, we obtain

$$l_p^t r_i' r_i = -(1/\beta)r_i^t g'' r_i.$$

(3.12) then follows by substituting this into (3.27). $\quad\square$

4. Finite difference schemes and invariant regions. In this section we make some preliminary remarks concerning invariant regions and finite difference schemes for systems of conservation laws.

Let $x_k = k\Delta x$, $k = 0, \pm 1, \pm 2, \ldots$, and $t_m = m\Delta t$, $m = 0, 1, 2, \ldots$, and denote by u_k^m an approximation to $u(x_k, t_m)$. For example, the Lax-Friedrichs scheme is based upon taking integral averages of solutions of Riemann problems as follows. Given $u_{k\pm 1}^m$, let $u(x, t), t > t_m$, be the solution of the Riemann problem with "initial" data

$$u(x, t_m) \begin{cases} u_{k-1}^m, & x < x_k, \\ u_{k+1}^m, & x > x_k, \end{cases}$$

and assume that the CFL condition

$$(4.1) \qquad \frac{\Delta x}{\Delta t} > \max_{j, x, t} |\lambda_j(u(x, t))|$$

holds. The Lax-Friedrichs approximation is then defined by

$$(4.2) \qquad u_k^{m+1} = \frac{1}{2\Delta x} \int_{x_{k-1}}^{x_{k+1}} u(x, t_{m+1}) \, dx.$$

If we apply the divergence theorem to the vector field $(f(u), u)$ on the rectangle $[x_{k-1}, x_{k+1}] \times [t_m, t_{m+1}]$ and use (1.1) and (4.1), we obtain the more familiar, computational form of the Lax-Friedrichs scheme,

$$u_k^{m+1} = \frac{1}{2}(u_{k-1}^m + u_{k+1}^m) + \frac{\Delta t}{2\Delta x}[f(u_{k-1}^m) - f(u_{k+1}^m)].$$

The Godunov scheme is defined in a similar way: we let $u(x, t)$ solve the double Riemann problem with data

$$u(x, t_m) = \begin{cases} u_{k-1}^m, & x < x_{k-1/2}, \\ u_k^m, & x_{k-1/2} < x < x_{k+1/2}, \\ u_{k+1}^m & x_{k+1/2} < x, \end{cases}$$

and take

$$(4.3) \qquad u_k^{m+1} = \frac{1}{\Delta x} \int_{x_{k-1/2}}^{x_{k+1/2}} u(x, t_{m+1}) \, dx.$$

As above, this scheme is equivalent to

$$(4.4) \qquad u_k^{m+1} = u_k^m + \frac{\Delta t}{\Delta x}[f(w(u_{k-1}^n, u_k^n)) - f(w(u_k^n, u_{k+1}^n))],$$

where $w(u_l, u_r)$ is the solution of the Riemann problem with data (u_l, u_r), evaluated at $(0, t)$.

Now, if a convex set S is invariant for solutions of Riemann problems, then the forms (4.2) and (4.3) show that S remains invariant for the above two schemes. We thus have:

THEOREM 4.1. *Let S be a convex set in u-space which is invariant in the sense of §2. Assume that all states are sufficiently close and that the CFL condition (4.1) holds. Then S is also invariant for solutions of the Lax-Friedrichs and Godunov schemes.*

Of course, when all fields are genuinely nonlinear, invariant sets are necessarily convex, by Theorem 3.6. We also remark that there is a proof (Hoff [4]) that the
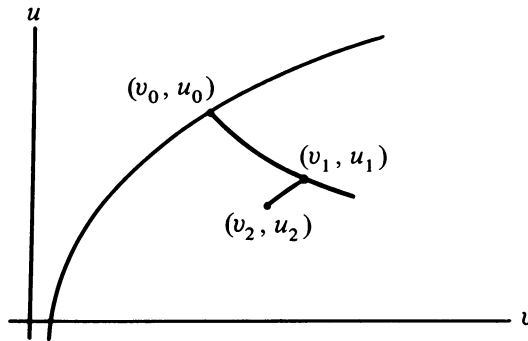
FIGURE 4

invariant regions found in Example 3.5 for the p-system remain invariant for the Lax-Friedrichs scheme, even when the states u_{k-1}^m and u_{k+1}^m are so far apart that the solution of the corresponding Riemann problem fails to exist. We conjecture that this is true more generally: that is, that if S is a convex set satisfying Theorem 3.2(a) and in which f' is bounded, then S should be invariant for solutions of the Lax-Friedrichs scheme, provided only that the CFL condition (4.1) holds. When such an invariant set does exist, the results of Hoff-Smoller [5] and Nishida-Smoller [9] become applicable to show that the parabolic system (3.10), with arbitrarily large BV Cauchy data, has a global smooth solution.

Last, we shall examine the Osher scheme (see [10, 11]) in the special case that there are no sonic points. Thus, for u in V,

$$(4.5) \qquad \lambda_1(u) < \cdots < \lambda_q(u) < 0 < \lambda_{q+1}(u) < \cdots < \lambda_n(u).$$

In this case Osher's scheme takes the simple form

$$(4.6) \qquad u_k^{m+1} = u_k^m - \frac{\Delta t}{\Delta x} \left[f(u_{k+1}^m) - f(v_q(u_k^m, u_{k+1}^m)) \right.$$
$$\left. + f(v_q(u_{k-1}^m, u_k^m)) - f(u_{k-1}^m) \right],$$

where $v_i(u_l, u_r)$, $i = 1, \ldots, n$, is defined as follows. We connect $u_r = v_0$ to $u_l = v_n$ with a sequence of curves Γ_i and require that Γ_i be an integral curve of r_i. v_{i-1} and v_i are then the endpoints of Γ_i. The reversal of the natural order of the characteristic speeds here seems to be connected to the sharpness with which the scheme resolves shocks and contact discontinuities. On the other hand, we shall show that this reversal can also result in the loss of invariant regions.

Suppose, for example, that the Osher scheme is applied to the p-system (Example 3.5). We let S be the set

$$S = \left\{ (v, u) \colon\ 0 \geq g(v, u) \equiv u - u_0 - \int_{v_0}^{v} \lambda(s)\,ds \right\},$$

which was shown to be an invariant region, and we take $(v_{k-1}^m, u_{k-1}^m) = (v_k^m, u_k^m) = (v_0, u_0)$ and $(v_{k+1}^m, u_{k+1}^m) = (v_2, u_2) \in S$ as in Figure 4. We may arrange these points so that (v_0, u_0) is connected to a point (v_1, u_1) by an integral curve of r_2, and (v_1, u_1) to (v_2, u_2) by an integral curve of r_1, in such a way that

$$(4.7) \qquad\qquad\qquad v_0 < v_2 < v_1.$$

Applying Osher's scheme (4.6), we then obtain

$$v_k^{m+1} \equiv \bar{v} = v_0 + \frac{\Delta t}{\Delta x}(u_2 - u_1),$$

$$u_k^{m+1} \equiv \bar{u} = u_0 - \frac{\Delta t}{\Delta x}[p(v_2) - p(v_1)].$$

Now let $v_1 = v_2 + \varepsilon$, where $\varepsilon > 0$. Then

$$u_2 = u_1 + \int_{v_1}^{v_2} \lambda(s)\, ds$$

so that

$$u_2 - u_1 = -\lambda(v_1)\varepsilon + O(\varepsilon^2).$$

Thus

$$g(\bar{v}, \bar{u}) = \bar{u} - u_0 - \int_{v_0}^{\bar{v}} \lambda(s)\, ds$$

$$= -\frac{\Delta t}{\Delta x}[p(v_2) - p(v_1)] - \int_{v_0}^{v_0 + (\Delta t/\Delta x)(u_2 - u_1)} \lambda(s)\, ds$$

$$= \frac{\Delta t}{\Delta x}p'(v_1)\varepsilon - \frac{\Delta t}{\Delta x}\lambda(v_0)(u_2 - u_1) + O(\varepsilon^2 + |u_2 - u_1|^2)$$

$$= -\frac{\Delta t}{\Delta x}\lambda(v_1)^2\varepsilon + \frac{\Delta t}{\Delta x}\lambda(v_0)\lambda(v_1)\varepsilon + O(\varepsilon^2)$$

$$= \frac{\Delta t}{\Delta x}\lambda(v_1)\varepsilon[\lambda(v_0) - \lambda(v_1)] + O(\varepsilon^2),$$

which is positive for $\varepsilon > 0$ small, by (4.7). (This depends upon the usual assumption that $p'' > 0$, which implies that $\lambda(v)$ is decreasing in v.) Thus $(\bar{v}, \bar{u}) \notin S$, even though (v_0, u_0) and (v_2, u_2) are points of S. Observe also that this failure to respect the invariant sets S cannot be remedied by any restriction on the mesh ratio $\Delta t/\Delta x$.

This example suggests that we modify the Osher scheme to restore the natural ordering of the characteristic speeds. Specifically, given vectors u_l and u_r, we connect $u_l = v_0$ to $u_r = v_n$ by a sequence of integral curves Γ_i of r_i. If $v_{i-1}(u_l, u_r)$ and $v_i(u_l, u_r)$ are the endpoints of Γ_i, and if the λ_i satisfy (4.5), then the modified scheme which results will be

(4.8) $$u_k^{m+1} = u_k^m - \frac{\Delta t}{\Delta x}[f(v_q(u_k^m, u_{k+1}^m)) - f(v_q(u_{k-1}^m, u_k^m))].$$

In the following theorem, we show that this "modified Osher scheme" does in fact preserve the correct invariant regions. While rather specialized because of assumption (4.4), this result is still broad enough to cover the application to the p-system (Example 3.5).

THEOREM 4.2. *Let S be a convex set satisfying Theorem 3.2(a), and assume that (4.5) holds. Then S is invariant for the modified Osher scheme (4.8) provided that all states are sufficiently close and that one of the following holds:*
   (a) *S is the intersection of V with a half-space;*
   (b) *S is strictly convex, $n = 2$, and the CFL condition (4.1) holds;*
   (c) *S is strictly convex, $n \geq 3$, and $\Delta t/\Delta x$ is sufficiently small.*

PROOF. It is sufficient to consider the case that $u_k^m \in \partial S$. We shall expand the terms $f(v_q)$ in (4.8) about u_k^m. First, let $u_k^m = v_0$ and $u_{k+1}^m = v_n$, and connect

v_{i-1} to v_i by an integral curve of r_i of "length" $\varepsilon_i = \lambda_i(v_i) - \lambda_i(v_{i-1})$. In the following computations we let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_q)$, and functions without arguments are understood to be evaluated at u_k^m:

$$v_q - u_k^m = \sum_{i=1}^{q}(v_i - v_{i-1}) = \sum_{i \leq q}\left(\varepsilon_i r_i + \frac{1}{2}\varepsilon_i^2 r_i' r_i\right)\Bigg|_{v_{i-1}} + O(\varepsilon^3)$$

$$= \sum_{i \leq q}\left(\varepsilon_i r_i + \frac{1}{2}\varepsilon_i^2 r_i' r_i\right) + \sum_{j < i \leq q}\varepsilon_i r_i'(v_j - v_{j-1}) + O(\varepsilon^3)$$

$$= \sum_{i \leq q}\left(\varepsilon_i r_i + \frac{1}{2}\varepsilon_i^2 r_i' r_i\right) + \sum_{j < i \leq q}\varepsilon_i \varepsilon_j r_i' r_j + O(\varepsilon^3).$$

Therefore

$$f(v_q) - f(u_k^m) = f'(v_q - u_k^m) + \tfrac{1}{2}f''(v_q - u_k^m)^2 + O(\varepsilon^3)$$

$$= \sum_{i \leq q}\left(\varepsilon_i \lambda_i r_i + \frac{\varepsilon_i^2}{2}f' r_i' r_i\right) + \sum_{j < i \leq q}\varepsilon_i \varepsilon_j f' r_i' r_j$$

$$+ \tfrac{1}{2}\sum_{i,j \leq q}\varepsilon_i \varepsilon_j f''(r_i, r_j) + O(\varepsilon^3)$$

$$= \sum_{i \leq q}\varepsilon_i \lambda_i r_i + \tfrac{1}{2}\sum_{i \leq q}\varepsilon_i^2[f' r_i' r_i + f''(r_i, r_i)]$$

$$+ \sum_{j < i \leq q}\varepsilon_i \varepsilon_j[f' r_i' r_j + f''(r_i, r_j)] + O(\varepsilon^3).$$

However, if we differentiate the identity $f' r_i = \lambda_i r_i$ in the direction of r_j, we obtain

$$f' r_i' r_j + f''(r_i, r_j) = (\partial \lambda_i / \partial r_j)r_i + \lambda_i r_i' r_j.$$

Therefore

$$(4.9) \quad f(v_q(u_k^m, u_{k+1}^m)) - f(u_k^m) = \sum_{i \leq q}\varepsilon_i \lambda_i r_i + \sum_{i \leq q}\varepsilon_i^2(r_i + \lambda_i r_i' r_i)$$

$$+ \sum_{j < i \leq q}\varepsilon_i \varepsilon_j\left(\frac{\partial \lambda_i}{\partial r_j}r_i + \lambda_i r_i' r_j\right) + O(\varepsilon^3).$$

Next, let $u_{k-1}^m = v_0$ be connected to $u_k^m = v_n$ by a sequence of integral curves Γ_i of r_i. Let v_{i-1} and v_i be the endpoints of Γ_i and $\varepsilon_i = \lambda_i(v_i) - \lambda_i(v_{i-1})$. Then letting $\varepsilon = (\varepsilon_{q+1}, \ldots, \varepsilon_n)$, we obtain, in a similar manner,

$$(4.10) \qquad f(u_k^m) - f(v_q(u_{k-1}^m, u_k^m))$$

$$= \sum_{q+1 \leq i}\varepsilon_i \lambda_i r_i - \tfrac{1}{2}\sum_{q+1 \leq i}\varepsilon_i^2(r_i + \lambda_i r_i' r_i)$$

$$- \sum_{q+1 \leq i < j}\varepsilon_i \varepsilon_j\left(\frac{\partial \lambda_i}{\partial r_j}r_j + \lambda_i r_i' r_j\right) + O(\varepsilon^3).$$

Now let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$, where $(\varepsilon_1, \ldots, \varepsilon_q)$ appears in (4.9) and $(\varepsilon_{q+1}, \ldots, \varepsilon_n)$ in (4.10). Adding (4.9) and (4.10), we obtain

$$(4.11) \quad f(v_q(u_k^m, u_{k+1}^m)) - f(v_q(u_{k-1}^m, u_k^m))$$
$$= \sum_i \varepsilon_i \lambda_i r_i - \frac{1}{2} \sum_i (\operatorname{sgn} \lambda_i) \varepsilon_i^2 (r_i + \lambda_i r_i' r_i)$$
$$+ \left(\sum_{j < i \leq q} - \sum_{q+1 \leq i < j} \right) \left[\varepsilon_i \varepsilon_j \left(\frac{\partial \lambda_i}{\partial r_j} r_i + \lambda_i r_i' r_j \right) \right] + O(\varepsilon^3).$$

Finally, let S be given locally as $\{u: g(u) \leq 0\}$, so that $\nabla g(u_k^m)$ is parallel to $l_p(u_k^m)$. Then from (4.8) and (4.11),

$$g(u_k^{m+1}) = g(u_k^m) + \nabla g^t (u_k^{m+1} - u_k^m)$$
$$+ \frac{1}{2} (u_k^{m+1} - u_k^m)^t g'' (u_k^{m+1} - u_k^m) + O(\varepsilon^3)$$
$$= -\alpha \varepsilon_p \lambda_p \nabla g^t r_p [1 + O(\varepsilon)] + \frac{\alpha}{2} \sum_{i \neq p} \varepsilon_i^2 |\lambda_i| \nabla g^t r_i' r_i$$
$$- \alpha \left(\sum_{\substack{j < i \leq q \\ i, j \neq p}} - \sum_{\substack{q+1 \leq i < j \\ i, j \neq p}} \right) (\varepsilon_i \varepsilon_j \lambda_i \nabla g^t r_i' r_j)$$
$$+ \frac{\alpha^2}{2} \sum_{i, j \neq p} \varepsilon_i \varepsilon_j \lambda_i \lambda_j r_j^t g'' r_i + O(\bar{\varepsilon}^3),$$

where α is the mesh ratio, $\alpha = \Delta t / \Delta x$, and $\bar{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_{p-1}, \varepsilon_{p+1}, \ldots, \varepsilon_n)$. However, for $i \neq p$, $\nabla g^t r_i = 0$ on ∂S, so that for $j \neq p$,

$$r_i^t g'' r_j + \nabla g^t r_i' r_j = 0.$$

Thus

$$(4.12) \qquad g(u_k^{m+1}) = -\alpha \varepsilon_p \lambda_p \nabla g^t r_p [1 + O(\varepsilon)] - \frac{\alpha}{2} \sum_{i \neq p} \varepsilon_i^2 |\lambda_i| r_i^t g'' r_i$$
$$+ \alpha \left(\sum_{\substack{j < i \leq q \\ i, j \neq p}} - \sum_{\substack{q+1 \leq i < j \\ i, j \neq p}} \right) (\varepsilon_i \varepsilon_j \lambda_i r_j^t g'' r_i)$$
$$+ \frac{\alpha^2}{2} \sum_{i, j \neq p} \varepsilon_i \varepsilon_j \lambda_i \lambda_j r_j^t g'' r_i + O(\bar{\varepsilon}^3).$$

We claim that the first term on the right side is nonpositive. Suppose first that $\lambda_p < 0$. Then $p \leq q$ and the curve Γ_p occurs in the path joining u_k^m to u_{k+1}^m. Since $u_k^m = v_0$ is in ∂S, so is v_{p-1}; and since $v_n = u_{k+1}^m$ is in S, so is v_p. Thus

$$0 \geq g(v_p) = g(v_p) - g(v_{p-1}) = \varepsilon_p \nabla g^t r_p |_{v_{p-1}} + O(\varepsilon_p^2).$$

Thus if $\varepsilon_p \neq 0$, ε_p and $\nabla g^t r_p |_{v_{p-1}}$ have opposite signs. Since all states are close and $\lambda_p < 0$, we may then conclude that

$$-\alpha \varepsilon_p \lambda_p \nabla g \cdot r_p |_{u_l} [1 + O(\varepsilon)] \leq 0.$$

The proof is similar for the case that $\lambda_p > 0$. Now, when (a) of the theorem holds, g is linear and the above estimate therefore shows that $g(u_k^{m+1}) \leq 0$, which implies that $u_k^{m+1} \in S$.

When $n = 2$, the quadratic terms on the right side of (4.12) simplify to

$$-(\alpha/2)\varepsilon_i^2 |\lambda_i| r_i^t g'' r_i + (\alpha^2/2)\varepsilon_i^2 \lambda_i^2 r_i^t g'' r_i = -(\alpha/2)\varepsilon_i^2 |\lambda_i| r_i^t g'' r_i (1 - \alpha|\lambda_i|),$$

which is negative definite in $\bar{\varepsilon}$ by (4.1) and the strict convexity of S. This proves (b).

To prove (c) we write the quadratic terms on the right side of (4.11) as $-\alpha A(\bar{\varepsilon}, \bar{\varepsilon}) + \alpha^2 B(\bar{\varepsilon}, \bar{\varepsilon})$, where A and B are the obvious bilinear forms. Clearly,

$$0 < \underline{\mu}|\bar{\varepsilon}|^2 \leq B(\bar{\varepsilon}, \bar{\varepsilon}) \leq \overline{\mu}|\bar{\varepsilon}|^2.$$

Now, if we were to repeat the entire computation leading up to (4.12) instead for the Godunov scheme (4.4), the result would have been exactly the same, since the shock and rarefaction curves agree up to third order. And since S is invariant for the Godunov scheme, it follows that the form $-\alpha A + \alpha^2 B$ is nonpositive when α is smaller than the CFL number α_0 in (4.1). Therefore

$$A(\bar{\varepsilon}, \bar{\varepsilon}) \geq \alpha_0 B(\bar{\varepsilon}, \bar{\varepsilon}) \geq \alpha_0 \underline{\mu}|\bar{\varepsilon}|^2,$$

so that

$$-\alpha A(\bar{\varepsilon}, \bar{\varepsilon}) + \alpha^2 B(\bar{\varepsilon}, \bar{\varepsilon}) \leq -\alpha(\alpha_0 \underline{\mu} - \alpha\overline{\mu})|\bar{\varepsilon}|^2 \leq -C|\bar{\varepsilon}|^2,$$

provided that $\alpha < (\underline{\mu}/\overline{\mu})\alpha_0$. In this case the expansion (4.12) shows that $g(u_k^{m+1}) \leq 0$ as required. $\square$

REFERENCES

1. R. Aris and N. Amundson, *Mathematical methods in chemical engineering*, Vol. 2, Prentice-Hall, Englewood Cliffs, N.J., 1973.
2. K. Chueh, C. Conley and J. Smoller, *Positively invariant regions for systems of nonlinear diffusion equations*, Indiana Univ. Math. J. **26** (1977), 373–392.
3. J. Glimm, *Solutions in the large for nonlinear hyperbolic systems of equations*, Comm. Pure Appl. Math. **18** (1965), 95–105.
4. David Hoff, *A finite difference scheme for a system of two conservation laws with artificial viscosity*, Math. Comp. **33** (1979), 1171–1193.
5. David Hoff and Joel Smoller, *Error bounds for finite difference approximations for a class of nonlinear parabolic systems*, Math. Comp. **45** (1985).
6. J. Johnson and J. Smoller, *Global solutions for an extended class of hyperbolic systems of conservation laws*, Arch. Rational Mech. Anal. **32** (1969), 169–189.
7. P. Lax, *Hyperbolic systems of conservation laws*. II, Comm. Pure Appl. Math. **10** (1957), 537–566.
8. _____, *Shock waves and entropy*, Contributions to Nonlinear Functional Analysis (E. Zarantonello, ed.), Academic Press, New York, 1971.
9. T. Nishida and J. Smoller, *A class of convergent finite difference schemes for certain nonlinear parabolic systems* Comm. Pure Appl. Math. **23** (1983), 785–808.
10. S. Osher, *Numerical solution of singular perturbation problems and hyperbolic systems of conservation laws*, Analytical and Numerical Approaches to Asymptotic Problems in Analysis (S. Axelsson, L. S. Frank, A. van der Sluis, eds.), North-Holland, Amsterdam, 1981.
11. S. Osher and F. Solomon, *Upwind difference schemes for hyperbolic systems of conservation laws*, Math. Comp. **158** (1982), 339–374.
12. F. Valentine, *Convex sets*, McGraw-Hill, New York, 1964.

DEPARTMENT OF MATHEMATICS, INDIANA UNIVERSITY, BLOOMINGTON, INDIANA 47405