



---

On Upstream Differencing and Godunov-Type Schemes for Hyperbolic Conservation Laws

Author(s): Amiram Harten, Peter D. Lax and Bram Van Leer

Source: *SIAM Review*, Jan., 1983, Vol. 25, No. 1 (Jan., 1983), pp. 35-61

Published by: Society for Industrial and Applied Mathematics

Stable URL: <https://www.jstor.org/stable/2030019>

**REFERENCES**

Linked references are available on JSTOR for this article:

[https://www.jstor.org/stable/2030019?seq=1&cid=pdf-reference#references\\_tab\\_contents](https://www.jstor.org/stable/2030019?seq=1&cid=pdf-reference#references_tab_contents)

You may need to log in to JSTOR to access the linked references.

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

*Society for Industrial and Applied Mathematics* is collaborating with JSTOR to digitize, preserve and extend access to *SIAM Review*

## ON UPSTREAM DIFFERENCING AND GODUNOV-TYPE SCHEMES FOR HYPERBOLIC CONSERVATION LAWS\*

AMIRAM HARTEN†, PETER D. LAX‡, AND BRAM VAN LEER§

**Abstract.** This paper reviews some of the recent developments in upstream difference schemes through a unified representation, in order to enable comparison between the various schemes. Special attention is given to the Godunov-type schemes that result from using an approximate solution of the Riemann problem. For schemes based on flux splitting, the approximate Riemann solution can be interpreted as a solution of the collisionless Boltzmann equation.

**Introduction.** Upstream-differencing schemes attempt to discretize hyperbolic partial differential equations by using differences biased in the direction determined by the sign of the characteristic speed.

In recent years upstream differencing has become very popular, and a multitude of new techniques of implementing directionally biased differencing have been suggested. This popularity is primarily due to the robustness of upstream-differencing schemes, the availability of an underlying physical model, and the possibility of achieving high resolution of stationary discontinuities. Most of these schemes are an extension of the Courant–Isaacson–Rees scheme [1] to nonlinear conservation laws, and therefore a unified description may be given.

The present paper concentrates on reviewing basic concepts and deriving design principles; numerical experimentation is not presented. The paper is built up as follows. Section 1 reviews some properties of the equations essential to their proper numerical approximation. In §2 we discuss a straightforward extension of linear upstream differencing to nonlinear systems. Section 3 introduces the physical picture due to Godunov, useful to interpret certain schemes from §2 and to construct new schemes. Finally, in §4 we comment on flux splitting, another form of upstream differencing, and relate it to a class of schemes motivated by the Boltzmann equation.

**1. Weak solutions and their numerical approximation.** In this paper we consider numerical solutions of the initial-value problem for hyperbolic systems of conservation laws

$$(1.1a) \quad u_t + f(u)_x = 0, \quad u(x, 0) = u_0(x), \quad -\infty < x < \infty.$$

Here  $u(x, t)$  is a column vector of  $m$  unknowns and  $f(u)$ , the flux, is a vector-valued function of  $m$  components. We can write (1.1a) in matrix form

$$(1.1b) \quad u_t + Au_x = 0, \quad A(u) = f_u.$$

(1.1) is called *hyperbolic* if all eigenvalues of the Jacobian matrix  $A$  are real. We assume that the eigenvalues  $a_1(u), \dots, a_m(u)$  are distinct and arranged in an increasing order.

---

\*Received by the editors March 29, 1982, and in revised form May 5, 1982. This research was sponsored by the National Aeronautics and Space Administration under contract NAS1-15810 while the first and third authors were in residence at ICASE, NASA Langley Research Center, Hampton, Virginia.

†Tel Aviv University, Ramat Aviv, Israel. The work of this author was supported in part by the National Aeronautics and Space Administration under contract NCA2-OR525-001 at NASA Ames Research Center, Moffett Field, California, and by the U.S. Department of Energy under contract DE-AC02-76ER03077.

‡Courant Institute of Mathematical Sciences, New York University, New York, New York 10012. The work of this author was supported in part by the U.S. Department of Energy under contract DE-AC02-76ER03077.

§Institute for Computer Applications in Science and Engineering, Leiden State University, Leiden, the Netherlands.

To allow for discontinuous solutions we admit weak solutions that satisfy (1.1) in the sense of distribution theory, i.e.,

$$(1.2a) \quad \int_0^\infty \int_{-\infty}^\infty [w_t u + w_x f(u)] dx dt + \int_{-\infty}^\infty w(x, 0) u_0(x) dx = 0,$$

for all  $C^\infty$  test functions  $w(x, t)$  that vanish for  $|x| + t$  large.

Condition (1.2a) is equivalent to requiring that for all rectangles  $(a, b) \times (t_1, t_2)$  the relation obtained by integrating (1.1a) over the rectangle should hold:

$$(1.2b) \quad \int_a^b u(x, t_2) dx - \int_a^b u(x, t_1) dx + \int_{t_1}^{t_2} f(u(b, t)) dt - \int_{t_1}^{t_2} f(u(a, t)) dt = 0.$$

Clearly, a piecewise-smooth weak solution of (1.1) satisfies (1.1) pointwise in each smooth region; across each curve of discontinuity the Rankine–Hugoniot relation

$$(1.3) \quad f(u_R) - f(u_L) = S(u_R - u_L)$$

holds, where  $S$  is the speed of propagation of the discontinuity, and  $u_L$  and  $u_R$  are the states on the left and the right, respectively.

Since weak solutions of (1.1) are not uniquely determined by their initial data, we select physically relevant solutions, defined as those solutions that are limits as  $\varepsilon \rightarrow 0$  of solutions  $u(\varepsilon)$  of the viscous equations

$$(1.4) \quad u_t + f(u)_x = \varepsilon u_{xx}, \quad \varepsilon > 0.$$

In this paper we consider systems of conservation laws (1.1) that possess an *entropy function*  $U(u)$ , defined as follows:

- (i)  $U$  is a convex function of  $u$ , i.e.,  $U_{uu} > 0$ .
- (ii)  $U$  satisfies

$$(1.5a) \quad U_u f_u = F_u$$

where  $F$  is some other function called *entropy flux*; it follows from (1.5a) that every smooth solution of (1.1) also satisfies

$$(1.5b) \quad U(u)_t + F(u)_x = 0.$$

Limit solutions of (1.4) satisfy, in the weak sense, the following inequality:

$$(1.6a) \quad U(u)_t + F(u)_x \leq 0;$$

i.e., for all nonnegative smooth test functions  $w(x, t)$  of compact support

$$(1.6b) \quad - \int_0^\infty \int_{-\infty}^\infty (w_t U + w_x F) dx dt - \int_{-\infty}^\infty w(x, 0) U(u_0(x)) dx \leq 0.$$

Condition (1.6b) is equivalent to requiring that for all rectangles  $(a, b) \times (t_1, t_2)$  the inequality obtained by integrating (1.6a) over the rectangle should hold:

$$(1.6c) \quad \int_a^b U(u(x, t_2)) dx - \int_a^b U(u(x, t_1)) dx + \int_{t_1}^{t_2} F(u(b, t)) dt - \int_{t_1}^{t_2} F(u(a, t)) dt \leq 0.$$

If  $u$  is piecewise smooth with discontinuities, then (1.5b) holds pointwise in the smooth regions, while across a discontinuity

$$(1.6d) \quad F(u_R) - F(u_L) - S[U(u_R) - U(u_L)] \leq 0.$$

Relations (1.6) are called *entropy conditions* (see [12]).

In the following we shall describe numerical approximations to weak solutions of (1.1) that are obtained by 3-point explicit schemes in conservation form:

$$(1.7a) \quad v_j^{n+1} = v_j^n - \lambda f_{j+1/2}^n + \lambda f_{j-1/2}^n, \quad \lambda = \tau/\Delta,$$

where

$$(1.7b) \quad f_{j+1/2}^n = f(v_j^n, v_{j+1}^n).$$

Here  $v_j^n = v(j\Delta, n\tau)$ , and  $f(u, v)$  is a *numerical flux*. We require the numerical flux to be *consistent* with the physical flux in the following sense:

$$(1.7c) \quad f(u, u) = f(u).$$

We say that the difference scheme (1.7) is *consistent* with the entropy condition (1.6a) if an inequality of the following kind is satisfied:

$$(1.8a) \quad U_j^{n+1} \leq U_j^n - \lambda F_{j+1/2}^n + \lambda F_{j-1/2}^n,$$

where the following abbreviations are used:

$$(1.8b) \quad U_j^{n+1} = U(v_j^{n+1}), \quad U_j^n = U(v_j^n),$$

$$(1.8c) \quad F_{j+1/2}^n = F(v_j^n, v_{j+1}^n);$$

here  $F(u, v)$  is a *numerical entropy flux*, consistent with the entropy flux:

$$(1.8d) \quad F(u, u) = F(u).$$

The following is an easy (but useful) extension of an easy (but useful) theorem of Lax and Wendroff [12]:

**THEOREM 1.1.** *Suppose the difference scheme (1.7) is consistent with the conservation law (1.1a), and with the entropy condition (1.6a). Let  $v_j^n$  be a solution of (1.7), with initial values  $v_j^0 = \phi(j\Delta)$ . Extend the lattice function  $v_j^n$  to continuous values of  $x, t$  by setting, as usual*

$$(1.9) \quad v(x, t) = v_j^n, \quad j = [x/\Delta], \quad n = [t/\tau].$$

*Suppose that for some sequence  $\Delta_k \rightarrow 0$ ,  $\tau/\Delta = \lambda$ , the limit*

$$\lim_{\Delta_k \rightarrow 0} v(x, t) = u(x, t)$$

*exists in the sense of bounded,  $L_1^{\text{loc}}$  convergence. Then the limit  $u$  satisfies the weak form (1.2) of the conservation law, and the weak form (1.6b) of the entropy condition.*

The proof consists, just as in [12], of multiplying (1.7a) by a test function, summing by parts over  $n$  and  $j$ , writing the sum as an integral, and passing to the limit  $\Delta_k \rightarrow 0$ .

Theorem 1.1 remains true, and its proof the same, when the fluxes  $f$  and  $F$  are allowed to be functions of  $2l$  arguments:

$$(1.10) \quad f_{j+1/2} = f(u_{j-l+1}, u_{j-l+2}, \dots, u_{j+l}),$$

and similarly for  $F_{j+1/2}$ .

Assume that  $u_0(x)$  is equal to some reference state  $u_*$  for  $|x|$  large:

$$(1.11a) \quad u_0(x) = u_* \quad \text{for } |x| > M.$$

Then

$$(1.11b) \quad v_j^n = u_* \quad \text{for } \Delta|j| > M + n\Delta.$$

The entropy  $U$  may be altered by adding to it an arbitrary inhomogeneous linear function; this follows from definition (1.5a). Adding such a linear function to  $U$  will not alter its convexity, but achieves the following:

$$(1.12) \quad U(u_*) = 0, \quad U_u(u_*) = 0.$$

Since  $U$  is convex, it follows from (1.12) that  $U(u) > 0$  for  $u \neq u_*$ ; in fact if  $U$  is strictly convex,

$$(1.13) \quad U(u) \geq c|u - u_*|^2.$$

Now sum (1.8a) with respect to  $j$  over all integers  $j$ ; we obtain

$$(1.14a) \quad \sum_j U_j^{n+1} \leq \sum_j U_j^n.$$

In other words: total entropy is a decreasing function of time. In particular

$$(1.14b) \quad \sum_j U_j^n \leq \sum_j U_j^0.$$

This is an a priori inequality for solutions of the difference scheme (1.7), analogous to the energy inequality for linear symmetric hyperbolic differential and difference equations. Since by (1.13)  $U$  is positive for  $u \neq u_*$ , this is an a priori estimate for solutions of the difference scheme (1.7), and indicates that the scheme is stable. However (1.14b) is not strong enough to prove the pointwise boundedness of solutions of (1.7), nor the existence of convergent subsequences.

Theorem 1.1 holds in any number of space variables. Furthermore multidimensional schemes that are composites of one-dimensional fractional steps satisfy the multidimensional analogue of the entropy condition (1.8a) if each individual one-dimensional step satisfies an entropy inequality of the form (1.8a); see Crandall–Majda [2].

A word of caution: when dealing with equations of mathematical physics, in particular the equations of compressible flow, we must make sure that the difference scheme we are using keeps the variables within their physical range, i.e., that density and pressure are always positive quantities.

**2. Upstream-differencing schemes.** We start our review with the description of the first-order-accurate Courant–Isaacson–Rees scheme [1], the simplest upstream differencing scheme for the constant-coefficient scalar equation

$$(2.1) \quad u_t + a u_x = 0, \quad a = \text{constant},$$

$$(2.2a) \quad v_j^{n+1} = v_j^n - \lambda a \cdot \begin{cases} v_{j+1}^n - v_j^n & \text{for } a < 0, \\ v_j^n - v_{j-1}^n & \text{for } a > 0. \end{cases}$$

Introducing the notation

$$a^- \equiv \min(a, 0) = \frac{1}{2}(a - |a|),$$

$$a^+ \equiv \max(a, 0) = \frac{1}{2}(a + |a|),$$

we rewrite (2.2a) as

$$(2.2b) \quad v_j^{n+1} = v_j^n - \lambda[a^+(v_j^n - v_{j-1}^n) + a^-(v_{j+1}^n - v_j^n)],$$

which can be rewritten as

$$(2.2c) \quad v_j^{n+1} = \lambda a^+ v_{j-1}^n + \lambda(1 - |a|) v_j^n - \lambda a^- v_{j+1}^n.$$

Under the Courant–Friedrichs–Lewy condition

$$(2.2d) \quad \lambda|a| \leq 1$$

all coefficients of  $v_i^n$  on the right in (2.2b) are positive. Such a scheme is called *monotone*, and is stable in the maximum norm; that is, for a monotone scheme

$$(2.2e) \quad \max_j |v_j^{n+1}| \leq \max_j |v_j^n|.$$

Equation (2.2b) can be rewritten as

$$(2.2f) \quad v_j^{n+1} = v_j^n - \frac{\lambda}{2} a(v_{j+1}^n - v_{j-1}^n) + \frac{\lambda}{2} |a| (v_{j+1}^n - 2v_j^n + v_{j-1}^n).$$

This shows that solutions of (2.2) can be thought of as approximating solutions of

$$(2.3) \quad w_t + aw_x = \frac{1}{2} \Delta x |a| (1 - \lambda|a|) w_{xx}$$

to second-order accuracy. We observe that the viscosity term in (2.3) vanishes for  $a = 0$ ; this fact later will allow perfectly resolved stationary shocks but may also result in admitting entropy violating discontinuities.

We describe now the extension of (2.2) to systems of equations with constant coefficients:

$$(2.4) \quad u_t + Au_x = 0, \quad A = \text{constant}.$$

Because of the hyperbolicity assumption, the system (2.4) can be diagonalized by a similarity transformation

$$(2.5a) \quad w = T^{-1}u, \quad T^{-1}AT = \Lambda, \quad \Lambda_{ij} = a_i \delta_{ij},$$

$$(2.5b) \quad w_t + \Lambda w_x = 0.$$

The components of  $w$  are called characteristic variables and (2.5b) is a system of decoupled characteristic equations.

We extend the Courant–Isaacson–Rees scheme to systems by applying the scalar scheme (2.2) to each of the decoupled scalar characteristic equations. In matrix form this can be written as

$$(2.6a) \quad w_j^{n+1} = w_j^n - \frac{\lambda}{2} \Lambda (w_{j+1}^n - w_{j-1}^n) + \frac{\lambda}{2} |\Lambda| (w_{j+1}^n - 2w_j^n + w_{j-1}^n),$$

where the diagonal matrix  $|\Lambda|$  is defined by  $|\Lambda|_{ij} = |a_i| \delta_{ij}$ . In the original variables, the scheme (2.6a) takes the form

$$(2.6b) \quad v_j^{n+1} = v_j^n - \frac{\lambda}{2} A (v_{j+1}^n - v_{j-1}^n) + \frac{\lambda}{2} |A| (v_{j+1}^n - 2v_j^n + v_{j-1}^n),$$

where  $|A| = T|\Lambda|T^{-1}$ . Clearly, the stability condition for (2.6a) and (2.6b) is

$$(2.6c) \quad \frac{\Delta t}{\Delta x} \max_k |a_k| \leq 1.$$

In general we define the matrix  $\chi(A)$  by

$$(2.7) \quad \chi(A) = T\chi(\Lambda)T^{-1}, \quad (\chi(\Lambda))_{ij} = \chi(a_i)\delta_{ij}.$$

We remark that under our assumption of a full set of eigenvectors we can compute  $\chi(A)$  by  $\chi(A) = P(A)$ , where  $P(x)$  is the Lagrangian interpolation polynomial such that  $P(a_i) = \chi(a_i)$ ,  $i = 1, \dots, m$ .

In the following we shall describe various techniques to extend the upstream-differencing scheme (2.6b) to nonlinear systems of conservation laws (1.1).

The linear system (2.4) can be regarded as a system of conservation laws (1.1a), where the flux depends linearly on  $u$ :

$$(2.8a) \quad f(u) = Au.$$

The upstream-differencing scheme (2.6b) is in conservation form (1.7), with numerical flux given by

$$(2.8b) \quad f(u, v) = A^+ u + A^- v$$

where  $A^+$  and  $A^-$  are the positive and negative parts of  $A$ , defined by the functional calculus (2.7) as

$$(2.8c) \quad A^+ = \chi^+(A), \quad A^- = \chi^-(A)$$

where we set

$$(2.8d) \quad \chi^+(a) = a^+, \quad \chi^-(a) = a^-.$$

Note that since  $\chi^+(a) + \chi^-(a) = a$ , and  $\chi^+(a) - \chi^-(a) = |a|$ , we can write

$$(2.9) \quad A^+ = \frac{1}{2}(A + |A|), \quad A^- = \frac{1}{2}(A - |A|).$$

DEFINITION. A difference scheme in conservation form (1.7) is said to be an *upstream* scheme if:

(i) For  $u$  and  $v$  nearby states, (2.8b) is a linear approximation to the numerical flux  $f(u, v)$ .

(ii) When all signal speeds associated with the numerical flux  $f(u, v)$  are  $> 0$ ,

$$f(u, v) = f(u).$$

When all signal speeds are  $< 0$ ,

$$f(u, v) = f(v).$$

The relevant signal speeds generally differ from the characteristic speeds of the states  $u$  and  $v$ .

We restate (i) in analytic terms: Suppose  $u$  and  $v$  are near some reference state  $u_*$ ; then we require that

$$(2.10a) \quad \begin{aligned} f(u, v) = & f(u_*) + A^+(u_*)(u - u_*) + A^-(u_*)(v - u_*) \\ & + o(|u - u_*| + |v - u_*|). \end{aligned}$$

A natural choice for  $u_*$  is  $(u + v)/2$ ; setting this into (2.10a), making use of (2.9) to set

$$A^+ - A^- = |A|$$

and noting that

$$f\left(\frac{u+v}{2}\right) = \frac{f(u) + f(v)}{2} + o(|u - v|)$$

we get

$$(2.10b) \quad f(u, v) = \frac{f(u) + f(v)}{2} - \frac{1}{2} \left| A\left(\frac{u+v}{2}\right) \right| (v - u) + o(|u - v|).$$

We can write any numerical flux in the form

$$(2.11a) \quad f(u, v) = \frac{f(u) + f(v)}{2} - \frac{1}{2} d(u, v);$$

for the sake of consistency (1.7c) we need

$$(2.11b) \quad d(u, u) = 0.$$

The upstream condition (2.10b) can be expressed then as

$$(2.11c) \quad d(u, v) = \left| A \left( \frac{u + v}{2} \right) \right| (v - u) + o(|u - v|).$$

Formula (2.6a) shows that linear upstream difference schemes contain a large dose of artificial viscosity, except for those components where  $a_k$  is small, in particular where  $a_k = 0$ . The same appears to be true for all upstream difference schemes for nonlinear conservation laws: when all characteristic speeds are not zero, each component is treated like a scheme with a hefty amount of artificial viscosity, smearing discontinuities. There is however quite a distinction among the schemes when one of the characteristic speeds is zero; this shows up in the way each scheme resolves a stationary shock, centered transonic rarefaction wave, and stationary contact discontinuity. We turn now to examining these matters.

The most critical difference in performance occurs in resolving a stationary shock (see (1.6d)):

$$(2.12a) \quad u_0(x) = \begin{cases} u, & x < 0, \\ v, & x > 0, \end{cases} \quad f(u) = f(v), \quad F(v) < F(u).$$

The lack of numerical dissipation allows the design of schemes that perfectly resolve stationary shocks, i.e., (2.12a) is a steady solution of the numerical scheme. The condition for that is

$$(2.12b) \quad d(u, v) = 0 \quad \text{if } f(u) = f(v), \quad F(v) < F(u).$$

On the other hand

$$(2.12c) \quad u_0(x) = \begin{cases} u, & x < 0, \\ v, & x > 0, \end{cases} \quad F(v) > F(u)$$

is not an admissible discontinuity and should not be a steady solution of the finite difference scheme, i.e., we require that

$$(2.12d) \quad d(u, v) \neq 0 \quad \text{if } f(u) = f(v), \quad F(v) > F(u).$$

We remark that the danger that a given upstream scheme selects a nonphysical solution will occur only for stationary or near-stationary discontinuities; otherwise there is enough numerical viscosity in (2.3) to enforce the selection of a physically relevant solution. Hence there are two options in designing an upstream-differencing scheme for solving problems with discontinuous solution:

(1) to switch direction of differencing in a way that will effectively introduce nonlinear dissipation at the expense of slightly spreading the shock;

(2) to satisfy (2.12) and thus get perfect resolution of a stationary shock, but to add a mechanism for checking the admissibility of the discontinuity.



We turn now to describing various forms of  $d(u, v)$  in (2.11a). The most straightforward way to generate such functions is by

$$(2.13a) \quad d(u, v) = |A|(u, v)(v - u),$$

where  $|A|(u, v)$  is a matrix function of  $u$  and  $v$  that has nonnegative eigenvalues, and such that

$$(2.13b) \quad |A|(u, u) = |A(u)|;$$

$|A(u)|$  is defined by (2.7).

The simplest forms of (2.13) are

$$(2.14a) \quad |A|(u, v) = \left| A\left(\frac{u+v}{2}\right) \right|,$$

or

$$(2.14b) \quad |A|(u, v) = \frac{1}{2} [|A(u)| + |A(v)|].$$

The latter was used by van Leer in [13], and introduces some nonlinear numerical dissipation that somewhat smears stationary shocks but on the other hand excludes nonphysical discontinuities.

Another form of (2.13a) with (2.14a) has been suggested by Huang [10]:

$$(2.15) \quad d(u, v) = \operatorname{sgn} \left( A\left(\frac{u+v}{2}\right) \right) [f(v) - f(u)];$$

here  $\operatorname{sgn}(x)$  is the sign of  $x$ , and  $\operatorname{sgn}(A)$  is defined by (2.7).

Yet another type of scheme has been designed by Roe [19]. His scheme is of the form (2.13a), where the matrix function  $A(u, v)$  is required to have these properties:

(i)

$$(2.16a) \quad f(v) - f(u) = A(u, v)(v - u).$$

(ii)  $A(u, v)$  has real eigenvalues and a complete set of eigenvectors.

(iii)

$$(2.16b) \quad A(u, u) = A(u).$$

For the Euler equations of compressible flow Roe [19] has constructed a linearization of form (2.16a) having these properties. We show now that such a linearization exists quite generally:

**THEOREM 2.1 (Harten-Lax).** *Suppose (1.1a) has an entropy function. Then (1.1a) has a Roe-type linearization.*

*Proof.* We shall construct an  $A$  satisfying (2.16a) which is of form  $A = BP$ ,  $B$  symmetric,  $P$  positive definite. Clearly, such an  $A$  is similar to the symmetric matrix  $P^{1/2}BP^{1/2}$  and so has property (ii). In our construction we use the entropy function  $U(u)$ ; since  $U$  is convex, the mapping  $u \rightarrow w = U_u$  is one-to-one; we introduce  $w$  as new variable in place of  $u$ . Let  $u_1$  and  $u_2$  be two arbitrary states,  $w_1 = w(u_1)$ ,  $w_2 = w(u_2)$ ,  $f_1 = f(u_1)$ ,  $f_2 = f(u_2)$ . Then

$$(2.17) \quad \begin{aligned} f_2 - f_1 &= \int_0^1 \frac{d}{d\theta} f(\theta w_2 + (1 - \theta)w_1) d\theta \\ &= \int_0^1 f_w d\theta (w_2 - w_1) = B(w_2 - w_1). \end{aligned}$$

We claim that  $f_w$  is a symmetric matrix; this implies that  $B$  is symmetric. To see about  $f_w$ , we use relation (1.5a); differentiating with respect to  $u$  we get

$$U_{uu}f_u + U_u f_{uu} = F_{uu}.$$

The second term on the left is a linear combination of symmetric matrices  $f_{uu}^i$ ; the right side,  $F_{uu}$ , also is symmetric. Therefore so is the first term

$$U_{uu}f_u.$$

It follows then that also

$$(2.18) \quad f_u U_{uu}^{-1}$$

is symmetric.

Differentiating

$$w = U_u$$

with respect to  $u$  shows that  $w_u = U_{uu}$ ; therefore

$$U_{uu}^{-1} = u_w.$$

Substituting this into (2.18) shows that

$$f_{uu}u_w = f_w$$

is symmetric, as asserted.

Next we express

$$(2.19) \quad \begin{aligned} w_2 - w_1 &= \int_0^1 \frac{d}{d\theta} w(\theta u_2 + (1 - \theta)u_1) d\theta \\ &= \int_0^1 w_u d\theta (u_2 - u_1) = P(u_2 - u_1). \end{aligned}$$

Using  $w_u = U_{uu}$  and the convexity of  $U$  we conclude that  $P$  as defined by (2.19) is positive definite.

Combining (2.17) and (2.19) gives

$$f_2 - f_1 = BP(u_2 - u_1) = A(u_2 - u_1).$$

Thus the  $A$  we have constructed can be factored as  $BP$ , as asserted. Condition (2.16b) is clearly satisfied.  $\square$

Note that  $B$  depends symmetrically on  $w_1$  and  $w_2$ , and  $P$  symmetrically on  $u_1$  and  $u_2$ . This shows that  $A$  is a symmetric function of  $u_1$  and  $u_2$ .

A similar result holds for systems of conservation laws in any number of space variables as long as there is an entropy (see Harten [8]).

Having constructed  $A$ , we can define its absolute value by (2.7); then we set

$$(2.20) \quad d(u, v) = |A(u, v)|(v - u).$$

When  $u$  and  $v$  correspond to a stationary discontinuity (2.12), then it follows from  $f(v) - f(u) = 0$  and (2.16a) that  $v - u$  is a null vector of  $A(u, v)$ , and consequently in the null space of  $|A(u, v)|$ ; thus  $d(u, v) = 0$ , whether or not the entropy condition  $F(v) - F(u) < 0$  is satisfied. Therefore the corresponding upstream differencing may admit nonphysical solutions. In an appendix to [9], Harten and Hyman describe a viscosity-like term, i.e., of the form (3.33a), that can be added to (2.15) and (2.16) to

reject inadmissible discontinuities without affecting the perfect resolution of the physical ones.

Yet another way to construct  $d$  in (2.11c) is

(2.21) 
$$d(u, v) = \int_u^v |A(w)| dw,$$

where the integration in (2.21) is carried out on a path in state-space connecting  $u$  and  $v$ . Osher in [16] suggests a path of integration  $\Gamma$  that is piecewise parallel to the right eigenvectors  $R_k$  of  $A$ :

(2.22a) 
$$\Gamma = \bigcup_{k=1}^m \Gamma_k,$$

(2.22b) 
$$\Gamma_k: \begin{cases} \frac{du^k}{dl} = R_k(u^k), & 0 < l < l_k, \quad k = 1, \dots, m, \\ u^{k+1}(l_{k+1}) = u^k(0), \end{cases}$$

(2.22c) 
$$u^m(0) = u, \quad u^1(l_1) = v.$$

Existence of a unique solution to (2.22b)–(2.22c) is guaranteed if  $\|u - v\|$  is sufficiently small [16]. A consequence of this choice of path is that  $d$  in (2.21) decouples into characteristic contributions

(2.23) 
$$d(u, v) = \sum_{k=1}^m \int_0^{l_k} |a_k(u(l))| R_k(u(l)) dl.$$

Osher shows that limit solutions of (2.11) with (2.23) satisfy the entropy condition and that a stationary contact discontinuity is perfectly resolved; the representation of a stationary shock requires at least one intermediate state.

3. Godunov-type schemes.

**A. The general theory.** Godunov, in his construction of the “best” monotone scheme [4], has used the exact solutions of local Riemann problems to obtain an upstream-differencing scheme.

The solution of the Riemann problem

(3.1) 
$$u_t + f(u)_x = 0, \quad u(x, 0) = \begin{cases} u_L, & x < 0, \\ u_R, & x > 0, \end{cases}$$

depends only on the states  $u_L$  and  $u_R$  and the ratio  $x/t$ ; it will be denoted by  $u(x/t; u_L, u_R)$ . Since signals propagate with finite velocity,

(3.2a) 
$$u(x/t; u_L, u_R) = u_L \quad \text{for } x/t \leq a_L,$$

(3.2b) 
$$u(x/t; u_L, u_R) = u_R \quad \text{for } x/t \geq a_R;$$

$a_L$  and  $a_R$  are the smallest and largest signal velocity.

Godunov derives his scheme by considering the numerical approximation  $v(x, t_n)$  of the discrete time levels  $t_n, n = 0, 1, \dots$ , to be a piecewise constant function in  $x$ , i.e.,

(3.3a) 
$$v(x, t_n) = v_j^n \quad \text{for } x \text{ in } I_j = ((j - \tfrac{1}{2})\Delta, (j + \tfrac{1}{2})\Delta).$$

To calculate the numerical approximation at the next time level  $t_{n+1} = t_n + \tau$ , we first solve exactly the initial-value problem

(3.3b) 
$$u_t + f(u)_x = 0, \quad u(x, t_n) = v(x, t_n), \quad -\infty < x < \infty,$$

for  $t_n \leq t \leq t_n + \tau$ , and denote its solution by  $u_n(x, t)$ . Each discontinuity in  $v(x, t_n)$  constitutes locally a Riemann problem. If we keep  $\lambda |a_{\max}| < 1/2$  where  $|a_{\max}|$  is the largest signal speed, then because of (3.2) there is no interaction between neighboring Riemann problems, and  $u_n(x, t)$  can be expressed exactly in terms of the solutions of local Riemann problems:

$$(3.3c) \quad u_n(x, t) = u\left(\frac{x - (j + 1/2)\Delta}{t - t_n}; v_j^n, v_{j+1}^n\right) \quad \text{for } j\Delta < x < (j + 1)\Delta, \quad t_n \leq t \leq t_{n+1}.$$

Godunov obtains a piecewise-constant approximation  $v(x, t_{n+1})$  by averaging  $u_n(x, t_{n+1})$ , i.e., he sets

$$(3.3d) \quad v_j^{n+1} = \frac{1}{\Delta} \int_{I_j} u_n(x, t_n + \tau) dx.$$

We can rewrite (3.3d) in terms of the solutions to the local Riemann problem as

$$(3.4a) \quad v_j^{n+1} = \frac{1}{\Delta} \int_0^{\Delta/2} u(x/\tau; v_{j-1}^n, v_j^n) dx + \frac{1}{\Delta} \int_{-\Delta/2}^0 u(x/\tau; v_j^n, v_{j+1}^n) dx.$$

Since  $u_n$  is an exact solution of the conservation laws (3.1), we can evaluate the integral defining  $v_j^{n+1}$  in (3.3d) by applying (1.2b) over  $I_j \times (t_n, t_{n+1})$ ; we get

$$(3.4b) \quad v_j^{n+1} = v_j^n - \lambda [f(\hat{v}_{j+1/2}) - f(\hat{v}_{j-1/2})],$$

where

$$(3.4c) \quad \hat{v}_{j+1/2} = u(0; v_j^n, v_{j+1}^n).$$

This shows that (3.4b) is in conservation form, with

$$(3.4d) \quad f(v, w) = f(u(0; v, w)).$$

The exact solution  $u_n(x, t)$  of the Riemann problem satisfies the entropy condition (1.6c):

$$\int_{I_j} U(u_n(x, t_{n+1})) dx \leq \Delta U(v_j^n) - \tau F(\hat{v}_{j+1/2}) + \tau F(\hat{v}_{j-1/2}).$$

Since  $U$  is a convex function, Jensen's inequality holds:

$$U\left(\frac{1}{\Delta} \int_{I_j} u(x, t) dx\right) \leq \frac{1}{\Delta} \int_{I_j} U(u(x, t)) dx.$$

Combining the last two inequalities, we deduce that Godunov's scheme satisfies the entropy inequality (1.8a).

The description (3.3c) makes sense only if the local Riemann problems don't interact, i.e., if

$$\lambda |a_{\max}| < 1/2.$$

On the other hand, (3.4b) remains consistent with (3.3d) as long as the waves issuing from  $(j \pm 1/2)\Delta$  do not reach  $(j \mp 1/2)\Delta$  during the time interval  $t_n \leq t \leq t_{n+1}$ . This will be the case as long as

$$\lambda |a_{\max}| \leq 1.$$

It follows from the Rankine–Hugoniot relation (1.3) that  $f(u(s; v, w)) - su(s; v, w)$

is a continuous function; however it is only piecewise differentiable. It follows that the Godunov flux function  $f(v, w)$  defined by (3.4d) is only piecewise differentiable.

Godunov's scheme satisfies criterion (ii) for upstream schemes because of (3.2). To

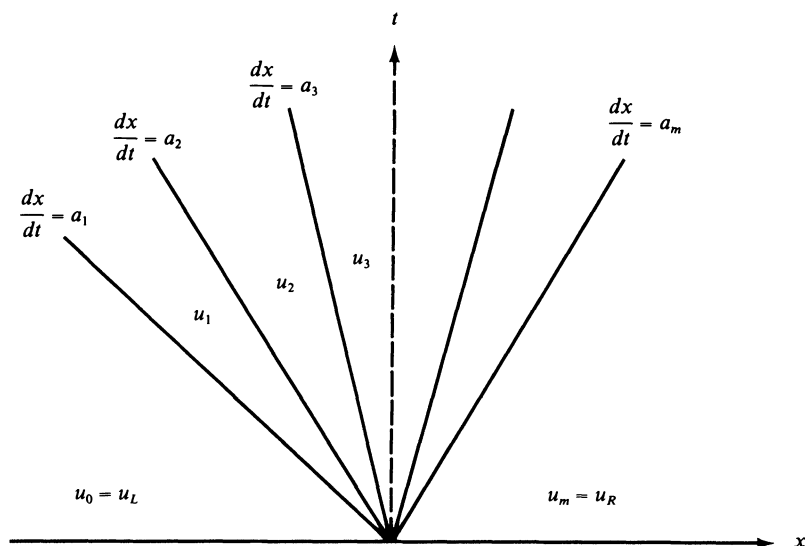


FIG. 1.

verify that it also satisfies criterion (i) we shall show that Godunov's scheme, when applied to linear equations, reduces to (2.6b).

Consider

$$(3.5a) \quad u_t + Au_x = 0,$$

$A$  a constant matrix. Here the solution of the Riemann problem is composed of constant states separated by a fan of  $m$  characteristic lines (see Fig. 1).

$$(3.5b) \quad u(x/t; u_L, u_R) = u_k \quad \text{for } a_k < x/t < a_{k+1}, \quad k = 0, \dots, m,$$

where we have defined

$$u_0 = u_L, \quad u_m = u_R, \quad a_0 = -\infty, \quad a_{m+1} = +\infty.$$

The intermediate states  $u_R$  can be calculated from the representation of  $u_R - u_L$  in terms of the right eigenvectors  $R_k$  of  $A$  in the following way:

$$(3.5c) \quad u_R - u_L = \sum_{i=1}^m J_i R_i,$$

$$(3.5d) \quad u_k = u_L + \sum_{i=1}^k J_i R_i.$$

We can write this by (2.7) as

$$(3.5e) \quad u_k = u_L + \sigma_k(A)(u_R - u_L)$$

where  $\sigma_k$  is the function

$$(3.5f) \quad \sigma_k(a) = \begin{cases} 1 & \text{for } a < a_k, \\ 0 & \text{for } a > a_k. \end{cases}$$

Let  $N$  be an integer such that

$$a_N < 0 < a_{N+1}.$$

Then by (3.5e)

$$(3.6) \quad \begin{aligned} u(0; u_L, u_R) &= u_N = u_L + \sigma_N(A)(u_R - u_L) \\ &= (I - \sigma_N(A))u_L + \sigma_N(A)u_R. \end{aligned}$$

Since by (2.8a), in the linear case  $f(u) = Au$ , substituting (3.6) into (3.4d) gives

$$f(u, v) = A(I - \sigma_N(A))u + A\sigma_N(A)v.$$

From the functional calculus (2.7) and (2.8c) we deduce that

$$A(I - \sigma_N(A)) = A^+, \quad A\sigma_N(A) = A^-,$$

so that

$$f(u, v) = A^+u + A^-v,$$

in full agreement with (2.8b). Thus in the linear case Godunov's scheme reduces to the upstream scheme (2.6b).

The solution to the Riemann problem (3.1) has a rather complicated structure: as in the constant coefficient case (3.5), the solution to (3.1) depends on  $x/t$  and consists of constant states  $u_k$ ,  $k = 0, \dots, m$ ;  $u_0 = u_L$ ,  $u_m = u_R$ , separated by a fan of waves. Unlike in the constant coefficient case, the  $k$ -wave separating  $u_{k-1}$  and  $u_k$  is not necessarily a single line having a characteristic speed  $a_k$ . If the  $k$ th characteristic field is genuinely nonlinear then the  $k$ -wave is either a rarefaction wave ( $a_k(u_{k-1}) < a_k(u_k)$ ) or a shock propagating with speed  $S$ , ( $a_k(u_{k-1}) > S > a_k(u_k)$ ). If the  $k$ th characteristic field is linearly degenerate then the  $k$ -wave is a contact discontinuity propagating with speed  $a_k(u_{k-1}) = a_k(u_k)$  (see [12]).

It is evident from (3.3d) that, due to averaging, the Godunov scheme does not make use of all the information contained in the exact solution of the Riemann problem. We therefore consider replacing the exact solution to the Riemann problem  $u(x/t; u_L, u_R)$  in (3.4a) by an approximation  $w(x/t; u_L, u_R)$ ; the latter can have a much simpler structure as long as it does not violate the essential properties of conservation and entropy inequality. The following theorem due to Harten and Lax [7, Thm. 2.1] shows that this type of approximation is consistent:

**THEOREM 3.1 (Harten-Lax).** *Let  $w(x/t; u_L, u_R)$  be an approximation to the solution of the Riemann problem that satisfies the following conditions:*

(i) *consistency with the integral form of the conservation law in the sense that*

$$(3.7a) \quad \int_{-\Delta/2}^{\Delta/2} w(x/t; u_L, u_R) dx = \frac{\Delta}{2} (u_L + u_R) - \tau f_R + \tau f_L$$

for  $\Delta/2 > \tau \max |a_k|$ , where

$$f_R = f(u_R), \quad f_L = f(u_L);$$

(ii) *consistency with the integral form of the entropy condition in the sense that*

$$(3.7b) \quad \int_{-\Delta/2}^{\Delta/2} U(w(x/t; u_L, u_R)) dx \leq \frac{\Delta}{2} (U_L + U_R) - \tau F_R + \tau F_L$$

for  $\Delta/2 > \tau \max |a_k|$ , where

$$F_R = F(u_R), \quad F_L = F(u_L).$$

Using the approximation  $w$  to the Riemann problem, we can define a Godunov-type scheme as follows:

$$(3.8) \quad v_j^{n+1} = \frac{1}{\Delta} \int_0^{\Delta/2} w(x/t; v_{j-1}^n, v_j^n) dx + \frac{1}{\Delta} \int_{-\Delta/2}^0 w(x/t; v_j^n, v_{j+1}^n) dx.$$

Assertion. If conditions (3.7a) and (3.7b) are satisfied, the scheme (3.8) is in conservation form consistent with (3.1), and satisfies the entropy inequality (1.8a).

For proof, see [7, after Thm. 2.1]. It is shown there that the Godunov-type averaging can be replaced by Glimm-type sampling. [7] contains a good account of this technique and an analysis of its accuracy.

Theorem 3.1 shows that Godunov-type schemes that satisfy conditions (i) and (ii) above satisfy the hypotheses of Theorem 1.1; this shows that if such a scheme converges, the limit satisfies the conservation law and the entropy condition in the weak sense.

We note that Godunov's scheme is of Godunov type.

We have shown at the beginning of this section that Godunov's scheme (3.3) can also be written as a scheme (3.4) in conservation form. The appropriate numerical flux was obtained from the integral conservation laws (1.2b). We show now that all schemes of Godunov type can be expressed in conservation form; we can obtain the appropriate numerical flux by applying the integral conservation law (1.2b) to the approximate solution of the Riemann problem over the rectangle  $(-\Delta/2, 0) \times (0, \tau)$ :

$$(3.9a) \quad \int_{-\Delta/2}^0 w(x/\tau; u_L, u_R) dx - \frac{\Delta}{2} u_L + \tau [f_{LR} - f_L] = 0,$$

where

$$f_{LR} = f(u_L, u_R).$$

This gives

$$(3.9b) \quad f_{LR} = f_L - \tau^{-1} \int_{-\Delta/2}^0 w(x/\tau; u_L, u_R) dx + \frac{\Delta}{2\tau} u_L.$$

If we apply the integral conservation law (1.2b) over the rectangle  $(0, \Delta/2) \times (0, \tau)$ , we obtain

$$(3.9c) \quad f'_{LR} = f_R + \tau^{-1} \int_0^{\Delta/2} w(x/\tau; u_L, u_R) dx - \frac{\Delta}{2\tau} u_R.$$

The equality of (3.9b) and (3.9c) is just the content of the consistency relation (3.7a).

Using formula (3.9b) for  $f_{j+1/2}$  in (1.7b) and (3.9c) for  $f_{j-1/2}$  in (1.7b) and setting the resulting expressions into (1.7a) gives (3.8), i.e., it puts the Godunov-type scheme in conservation form (1.7a):

$$(3.10) \quad v_j^{n+1} = v_j^n - \frac{\tau}{\Delta} [f(v_j, v_{j+1}) - f(v_{j-1}, v_j)].$$

If all signal speeds are positive, then  $w(s; u_L, u_R) = u_L$  for  $s < 0$ ; according to (3.9b) in this case  $f_{LR} = f_L$ . Similarly, if all signal speeds are negative, then  $w(s; u_L, u_R) = u_R$  for  $s > 0$ ; it follows from (3.9c) that in this case  $f_{LR} = f_R$ . This is property (ii) of upstream schemes, which is thus satisfied by all Godunov-type schemes.

In a scheme of Godunov type we can incorporate into the numerical flux all physical insights that we can put into the approximate solution of the Riemann problem. Also, as Harten and Lax pointed out in [7], a Godunov-type scheme (3.8) can be used just as easily

on a grid that varies in time, by adjusting the intervals of integration on the right in (3.8). This makes these schemes the natural choice for adaptive grids; further development of such algorithms and numerical experiments are described in Harten and Hyman [9].

**B. Construction of particular Riemann solvers.** We turn now to describing two different approximate Riemann solvers, and the Godunov-type schemes corresponding to them. The first, due to Roe, is based on a linearization notion of type (2.16). Roe approximates solutions of the Riemann problem for (3.1) by exact solutions of the Riemann problem for the following linear hyperbolic equation with constant coefficients:

$$(3.11a) \quad w_t + A_{LR} w_x = 0, \quad w(x, 0) = \begin{cases} u_L, & x < 0, \\ u_R, & x > 0. \end{cases}$$

Here  $A_{LR}$  is a matrix that satisfies (2.16a)–(2.16b) and has properties (i)–(iii) listed there. Combining (2.16a) with (3.5c) yields

$$(3.11b) \quad f_R - f_L = A_{LR}(u_R - u_L) = \sum a_i J_i R_i,$$

where  $a_i$  are the eigenvalues of  $A_{LR}$ ,  $R_i$  the corresponding right eigenvectors, and  $J_i$  the coefficients in the resolution (3.5c):

$$(3.11c) \quad u_R - u_L = \sum J_i R_i.$$

The approximate Riemann solver is given by (3.5b), with  $u_k$  defined by (3.5d).

The numerical flux associated with an approximate Riemann solver is given by (3.9b); substituting (3.5d) into (3.9b) we get

$$(3.12a) \quad f_{LR} = f_L + \sum a_i^- J_i R_i$$

where

$$a^- = \min(a, 0) = \frac{1}{2}(a - |a|).$$

Substituting this into (3.12) and using (3.11b) gives

$$(3.12b) \quad \begin{aligned} f_{LR} &= \frac{1}{2}(f_L + f_R) - \frac{1}{2} \sum |a_i| J_i R_i \\ &= \frac{1}{2}(f_L + f_R) - \frac{1}{2} |A_{LR}| (u_R - u_L); \end{aligned}$$

in the last step we have used the definition of  $|A|$  as given by (2.7). Indeed, (3.12b) is Roe's scheme defined in (2.11a), (2.20).

As already pointed out in §2, Roe's scheme admits nonphysical, i.e. entropy-violating, stationary discontinuities. In an appendix to [9], Harten and Hyman show how to modify Roe's scheme to eliminate such entropy-violating discontinuities while retaining those that satisfy the entropy law.

We note that the numerical flux (3.12b) of Roe's scheme resembles Osher's scheme (2.23). There the jumps  $J_k$  in the characteristic state variables are represented by the path length  $l_k$ . Osher's scheme, however, is not of Godunov type in the sense of (3.9) since the integration path  $\Gamma$  in state-space does not correspond to a univalued approximate Riemann solution  $w(x/t, u_L, u_R)$  as in (3.8).

Roe's Riemann solver contains a great amount of detail:  $m - 1$  intermediate states.



We describe next a hierarchy of Riemann solvers where much of this detail is lumped together. The simplest of these schemes contains only one intermediate state.

i) Denote by  $a_L$  and  $a_R$  lower and upper bounds, respectively, for the smallest and largest signal velocity, calculated according to some algorithm. Define the approximate Riemann solver by

$$(3.13) \quad u(x/t; u_L, u_R) = \begin{cases} u_L & \text{for } x/t < a_L, \\ u_{LR} & \text{for } a_L < x/t < a_R, \\ u_R & \text{for } a_R < x/t, \end{cases}$$

where the state  $u_{LR}$  is determined from the conservation law (3.7a):

$$(\tau a_L + \Delta/2)u_L + \tau(a_R - a_L)u_{LR} + (\Delta/2 - \tau a_R)u_R = \frac{\Delta}{2}(u_L + u_R) - \tau[f_R - f_L].$$

This gives

$$(3.14) \quad u_{LR} = \frac{a_R u_R - a_L u_L}{a_R - a_L} - \frac{f_R - f_L}{a_R - a_L}.$$

We turn now to the determination of the associated numerical flux. We substitute (3.14) into (3.13), and then into (3.9b):

$$(3.15a) \quad f_{LR} = \begin{cases} f_L & \text{when } 0 < a_L, \\ \frac{-a_L}{a_R - a_L} f_R + \frac{a_R}{a_R - a_L} f_L + \frac{a_L a_R}{a_R - a_L} (u_R - u_L) & \text{when } a_L < 0 < u_R, \\ f_R & \text{when } a_R < 0. \end{cases}$$

This can be combined into a single formula

$$(3.15b) \quad f_{LR} = \frac{a_R^- - a_L^-}{a_R - a_L} f_R + \frac{a_R^+ - a_L^+}{a_R - a_L} f_L - \frac{1}{2} \frac{a_R |a_L| - a_L |a_R|}{a_R - a_L} (u_R - u_L).$$

Since  $u_{LR}$  was chosen to satisfy the conservation law (3.7a), we conclude that  $u_{LR}$  is the mean value of the exact solution over the interval  $(\tau a_L, \tau a_R)$ . It follows therefore from Jensen's inequality that (3.15) satisfies the entropy inequality (3.7b).

Suppose that  $u_L$  and  $u_R$  can be connected with a shock of the first or the  $m$ th family. In these cases the exact solution is

$$(3.16) \quad u(x, t) = \begin{cases} u_L & \text{for } x/t < S, \\ u_R & \text{for } S < x/t, \end{cases}$$

where  $S$  is the speed of propagation of the shock. Suppose the algorithm for calculating  $a_L$  and  $a_R$  is such that it furnishes  $a_L = S$  or  $a_R = S$ , depending on whether the shock belongs to the first or the  $m$ th family. Then it follows from the equality of  $u_{LR}$  with the mean value of the exact solution that (3.15) is the exact solution.

ii) We describe next a class of approximate Riemann solvers, where  $u_L$  and  $u_R$  are linked through *two* intermediate states. These states are so chosen that:

a) The conservation laws are satisfied.

b) If the exact solution of the Riemann problem links  $u_L$  and  $u_R$  through a single shock (or contact discontinuity) of any of the  $m$  families of waves, then so does the approximate Riemann solver.

c) The entropy law is satisfied.

Such an approximate Riemann solver was constructed in [7]; the one presented here differs from it in some important details. We are grateful to Paul Woodward for a suggestion which has been incorporated in the scheme.

Let the velocities  $a_L$  and  $a_R$  be defined as in approximation i) described above. We define a velocity  $V$  as follows: Let  $U$  be an entropy function defined by (1.5); denote its gradient by  $w = U_u$ , and introduce the abbreviation

$$(3.17a) \quad w(u_R) - w(u_L) = l_{LR}.$$

We set

$$(3.17b) \quad V = l_{LR} \cdot (f_R - f_L) / l_{LR} \cdot (u_R - u_L)$$

where the dot denotes the Euclidean scalar product.

Next we show that  $V$  is well defined, and derive its salient properties:

LEMMA 3.2: i) *The denominator in (3.17b) is positive for  $u_L \neq u_R$ .*

ii)  *$V$  is uniformly bounded.*

iii) *If  $u_L$  and  $u_R$  satisfy the Rankine–Hugoniot condition (1.3),*

$$(3.18) \quad f_R - f_L = S(u_R - u_L),$$

then  $V = S$ .

*Proof.* i) Combining (3.17a) and (2.19),

$$(3.19) \quad l_{LR} \cdot (u_R - u_L) = P(u_R - u_L) \cdot (u_R - u_L);$$

since  $P$  is positive definite, the above quantity is positive for  $u_R \neq u_L$ .

ii) Use (2.17), (3.17a) and (2.19) to express the numerator of (3.17b), and (3.19) for the denominator; we get

$$V = \frac{P(u_R - u_L) \cdot BP(u_R - u_L)}{(u_R - u_L) \cdot P(u_R - u_L)}.$$

This is a ratio of two quadratic forms and therefore lies between the smallest and largest eigenvalue of  $P^{1/2}BP^{1/2}$ . These are equal to the eigenvalues  $a_j$  of  $A = BP$  constructed for Theorem 2.1. In fact,  $V$  can be represented as a weighted average of the eigenvalues of  $A$ .

iii) Setting (3.18) into (3.17b) gives

$$V = S.$$

This completes the proof of Lemma 3.2.

We now outline two methods for constructing approximate Riemann solvers  $w(x/t; u_L, u_R)$  with two intermediate states  $u_L^*$  and  $u_R^*$  separated by the line  $dx/dt = V$ ; i.e.,  $w$  is of the form

$$(3.20) \quad w(x/t; u_L, u_R) = \begin{cases} u_L, & x/t < a_L, \\ u_L^*, & a_L < x/t < V, \\ u_R^*, & V < x/t < a_R, \\ u_R, & a_R < x/t. \end{cases}$$

(See Fig. 2.)

The flux across the line  $x = st$  for equation (1.1) is defined as

$$(3.21) \quad f_s(u) = f(u) - su.$$

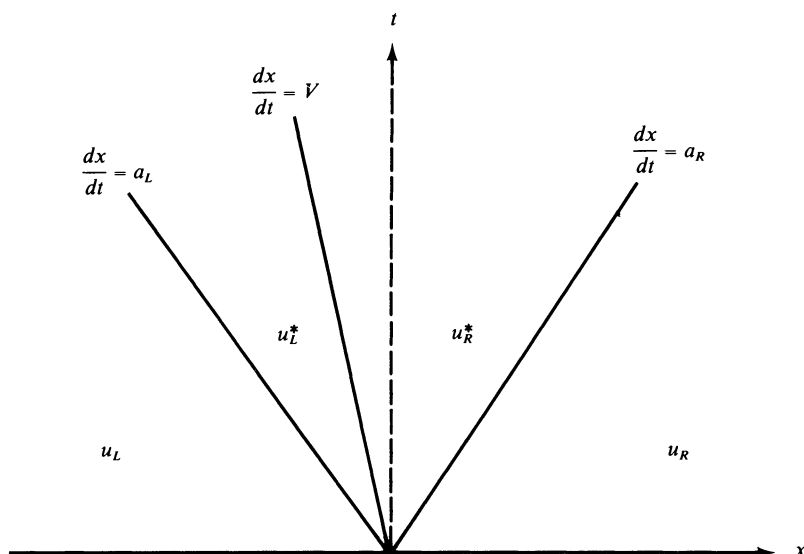


FIG. 2.

We introduce a numerical flux across the line  $x = Vt$ , denoted as

$$f_V(u_L, u_R).$$

We require consistency with the exact flux:

$$(3.22) \quad f_V(u, u) = f_V(u) = f(u) - Vu.$$

Having introduced flux (3.21) across lines, we can write approximate conservation laws for the triangular regions bounded by  $t = \tau$ ,  $x = Vt$ , and  $x = a_L t$  or  $x = a_R t$  respectively:

$$(3.23a) \quad (V - a_L)u_L^* + f_V(u_L, u_R) - f_{a_L}(u_L) = 0$$

and

$$(3.23b) \quad (a_R - V)u_R^* + f_{a_R}(u_R) - f_V(u_L, u_R) = 0;$$

$u_L^*$  and  $u_R^*$  can be determined from (3.23). Clearly, since (3.23a,b) are conservation laws, the resulting scheme (3.20) satisfies the consistency relation (3.7a). Thus requirement a) on conservation is fulfilled.

We turn now to requirement b), the exact resolution of single shocks and contact discontinuities. A shock or contact discontinuity is characterized by the Rankine–Hugoniot condition (3.18) and the entropy condition (1.6d). Using the notation (3.21), these can be written as follows:

$$(3.24) \quad f_S(u) = f_S(u_R), \quad F_S(u_L) \geq F_S(u_R)$$

where

$$(3.25) \quad F_S(u) = F(u) - SV.$$

We denote by  $\hat{f}_V$  a vector function that has the following property:

If  $f_V(u_L) = f_V(u_R)$ , then

$$(3.26) \quad \hat{f}_V(u_L, u_R) = f_V(u_L) = f_V(u_R).$$

Below we shall specify two distinct ways of constructing such an  $\hat{f}_V$ . We now set

$$(3.27) \quad f_V(u_L, u_R) = \hat{f}_V(u_L, u_R) - \beta(u_L, u_R)(u_R - u_L),$$

where  $\beta$  has the following property:

$$(3.28) \quad \beta(u_L, u_R) = 0 \quad \text{when (3.24) holds.}$$

We take (3.27) as our numerical flux. It follows from (3.26) that this flux satisfies the consistency condition (3.22), and from (3.26), (3.28) and part iii) of Lemma 3.2 that single shocks and contact discontinuities are resolved exactly.

Here are our choices for  $\hat{f}_V$  and  $\beta$ : we define  $\delta_R$  and  $\delta_L$  by

$$(3.29a) \quad \delta_R = \frac{a_R - V}{a_R - a_L}, \quad \delta_L = \frac{V - a_L}{a_R - a_L}.$$

Note that

$$(3.29b) \quad 0 \leq \delta_R, \quad 0 \leq \delta_L, \quad \delta_R + \delta_L = 1.$$

Then we define

$$(3.30a) \quad \hat{f}_V^a(u_L, u_R) = \delta_R f_V(u_L) + \delta_L f_V(u_R)$$

and

$$(3.30b) \quad \hat{f}_V^b(u_L, u_R) = f_V(u_{LR}) - f(\delta_L u_L + \delta_R u_R) + \delta_L f_L + \delta_R f_R,$$

where  $u_{LR}$  is defined by (3.14).

It can be verified immediately that  $f_V^a$  satisfies (3.26). To verify it for  $f_V^b$  we note that if  $f_V(u_L) = f_V(u_R)$ , then  $u_{LR} = \delta_R u_R + \delta_L u_L$ ; setting this into the right side of (3.30b) we see that it equals the right side of (3.30a).

We define  $\beta$  as follows:

$$(3.31a) \quad \beta = C_1 \beta_1 + C_2 \beta_2$$

where

$$(3.31b) \quad \beta_1(u_L, u_R) = [F_V(u_R) - F_V(u_L) - \frac{1}{2}(u_L + u_R) \cdot (f_V(u_R) - f_V(u_L))]^+ \|u_R - u_L\|^{-2}$$

where  $p^+$  denotes  $\max(0, p)$ , and

$$(3.31c) \quad \beta_2(u_L, u_R) = (a_R - a_L)^{-1} \|f_V(u_R) - f_V(u_L)\|^2 \|u_R - u_L\|^{-2}.$$

The analysis in [7, §4] shows  $\beta_1$  is a bounded function.

By construction,  $\beta_1 = 0$  when the shock condition (3.24) is satisfied;  $\beta_2 = 0$  when the Rankine–Hugoniot condition (3.18) alone is satisfied. Thus our choice of  $\beta$  satisfies (3.28), and so requirement b) is fulfilled.

An analysis similar to that carried out in [7, §4] shows that the positive constants  $C_1$  and  $C_2$  in (3.31a) can be so chosen that the entropy condition (3.7b) is satisfied. Thus is requirement c) fulfilled.

To derive the numerical flux associated with the above Godunov-type scheme we substitute (3.20) into (3.9b), using (3.23) to express  $u_L^*$  and  $u_R^*$ :

$$(3.32a) \quad f_{LR} = \frac{1}{2} \{ f_L + f_R + \gamma_L [f_V(u_L) - f_V(u_L, u_R)] \\ + \gamma_R [f_V(u_L, u_R) - f_V(u_R)] - |V|(u_R - u_L) \}$$

where

$$(3.32b) \quad \gamma_L = \frac{|V| - |a_L|}{V - a_L}, \quad \gamma_R = \frac{|a_R| - |V|}{a_R - V}.$$

One can easily verify that when  $a_L \geq 0, f_{LR} = f_R$ , and that when  $V = 0, f_{LR} = f_V(u_L, u_R)$ .

If on the right side of (3.32a) we substitute (3.27) for  $f_V(u_L, u_R)$ , the following term containing  $\beta$  appears:

$$(3.33a) \quad -\frac{1}{2}(\gamma_R - \gamma_L)\beta(u_R - u_L).$$

From (3.32b)

$$(3.33b) \quad \frac{1}{2}(\gamma_R - \gamma_L) = \begin{cases} 0 & \text{if } a_L > 0 \text{ or } a_R < 0, \\ \frac{|a_L|}{|a_L| + |V|} & \text{if } a_L < 0 < V < a_R, \\ \frac{|a_R|}{|a_R| + |V|} & \text{if } a_L < V < 0 < a_R, \end{cases}$$

a nonnegative quantity. This shows that  $\beta$  enters the difference scheme as an artificial viscosity. Note that, unlike classical artificial viscosity, our  $\beta$  is zero across a shock and is positive across an incipient rarefaction wave.

Unlike the previous schemes described in this review, the schemes (3.32) are nonlinear even when applied to linear equations. Thus it is not upstream in the sense of our definition in §2. The decrease of entropy guarantees the  $L_2$  stability of the scheme.

For systems with many components, (3.32) requires less computational effort than either Godunov's or Roe's scheme; more experimentation is needed to see whether its accuracy is comparable.

Schemes of type (3.32) are especially suitable for computation on a moving mesh: we move each meshpoint with velocity  $V$ . Such mesh algorithms have been studied in [9].

We remark that any scheme in conservation form (1.7) with a numerical flux  $f(u, v)$  that yields perfect resolution of discontinuities but also admits entropy-violating ones may be corrected by modifying its numerical flux to be

$$f(u, v) - C_1\beta(v - u).$$

**4. Flux-splitting.** In this section we discuss generalizations of the upstream-differencing scheme (2.6b) to nonlinear systems of conservation laws that are based on the flux-splitting

$$(4.1a) \quad f(w) = f^+(w) + f^-(w).$$

We consider schemes in conservation form (1.7a) with the numerical flux

$$(4.1b) \quad f(u, v) = f^+(u) + f^-(v);$$

clearly (4.1a) implies the consistency relation (1.7c). Let us define

$$(4.2a) \quad f^a(w) = f^+(w) - f^-(w),$$

and rewrite (4.1b):

$$(4.2b) \quad f(u, v) = \frac{1}{2} [f(u) + f(v) - (f^a(v) - f^a(u))].$$

Recalling the notation (2.11a) we write

$$(4.2c) \quad d(u, v) = f^a(v) - f^a(u).$$

It is easy to see that (4.1)–(4.2) reduces to (2.6b) in the constant-coefficient case if and only if  $f^a(w)$  becomes  $|A|w$ .

Steger and Warming [22] introduced the notion of flux splitting for the equations of gas dynamics. They took advantage of the fact that in gas dynamics  $f$  is a homogeneous function of  $w$  of degree one. Then Euler's identity holds:

$$(4.3a) \quad f(w) = A(w)w, \quad A = f_w.$$

Steger and Warming define

$$(4.3b) \quad f^+(w) = A^+(w)w, \quad f^-(w) = A^-(w)w,$$

where  $A^+$  and  $A^-$  are defined by (2.7), (2.9). Clearly (4.1a) is satisfied; (4.2a) becomes

$$(4.3c) \quad f^a(w) = |A(w)|w,$$

where  $|A(w)|$  is defined by (2.7).

Substituting (4.3c) into (4.2c) gives, after rearrangement,

$$(4.3c') \quad d(u, v) = \frac{1}{2} (|A(u)| + |A(v)|)(v - u) + \frac{1}{2} (|A(v)| - |A(u)|)(u + v).$$

The right-hand side is of the upstream form (2.11c), except for those nearby values of  $u$  and  $v$  for which  $\text{sgn } a_k(u) \neq \text{sgn } a_k(v)$  for some  $k$ . The consequence of this nonsmoothness is a kink in computed solutions near such transitions, e.g. near sonic points. This can be rectified to some extent if one replaces  $|A|$  by  $\chi(A)$ , where  $\chi(s)$  is a smooth approximation to  $|s|$ ; see Steger [23] or Harten [5]. Van Leer [15] has derived a smooth flux splitting for the Euler equations without reference to (4.3a).

We describe now a way of splitting when  $f(u)$  is not a homogeneous function of  $u$ , by introducing a reference state  $u_0$  and applying the mean-value linearization (2.16a) to  $u$  and  $u_0$ :

$$(4.4a) \quad f(u) - f(u_0) = A(u_0, u)(u - u_0).$$

We consider  $u - u_0$  to be a new state variable, and  $f(u) - f(u_0)$  to be a new flux function; denote them again by  $u$  and  $f(u)$ , respectively. Thus any solution of the conservation law (1.1) satisfies

$$(4.4b) \quad u_t + [A(u_0, u)u]_x = 0.$$

The new flux

$$(4.5a) \quad f(u) = A(u_0, u)u$$

can be split as

$$(4.5b) \quad f(u) = A^+(u_0, u)u + A^-(u_0, u)u = f^+(u) + f^-(u),$$

exactly as in the homogeneous case (4.3a).

We turn now to a class of upstream schemes that are a natural generalization of Steger and Warming's flux splitting. These schemes are obtained by approximating each conservation law in (1.1a) by a collisionless Boltzmann equation.

Let  $\phi(x, t, q)$  be a vector function whose  $i$ th component denotes the density of a particle of  $i$ th kind at position  $x$  and time  $t$ , travelling with velocity  $q$ . We assume that the particles stream freely, i.e. that  $\phi$  satisfies the collisionless Boltzmann equation:

$$(4.6a) \quad \phi_t + q\phi_x = 0.$$

Note that the densities of the different kinds of particles are completely decoupled.

Using equation (4.6a) we can determine the value of  $\phi$  for  $t > 0$  in terms of  $\phi_0(x, q) = \phi(x, 0, q)$ :

$$(4.6b) \quad \phi(x, t, q) = \phi_0(x - qt, q).$$

We denote by  $z$  and  $g$  the total density and flux associated with the density  $\phi$ :

$$(4.7a) \quad z(x, t) = \int \phi(x, t, q) dq,$$

$$(4.7b) \quad g(x, t) = \int q \phi(x, t, q) dq.$$

These satisfy the conservation law obtained by integrating (4.6a) with respect to  $q$ :

$$(4.8) \quad z_t + g_x = 0.$$

The initial values of  $z$  and  $g$  can be obtained by setting  $t = 0$  in (4.7):

$$(4.9) \quad z_0(x) = \int \phi_0(x, q) dq, \quad g_0(x) = \int q \phi_0(x, q) dq.$$

If the initial values of  $z$  and  $g$  are equal to those of  $u$  and  $f(u)$ :

$$(4.10) \quad z_0(x) = u_0(x), \quad g_0(x) = f(u_0(x));$$

then for  $t$  small enough the solution  $z$  of (4.8) would be a reasonable approximation to the solution  $u$  of (1.1a). We show now how to choose  $\phi_0$  so that (4.10) holds: we introduce a vector distribution  $\mu(q, u)$ , depending on a vector parameter  $u$ , satisfying

$$(4.11a) \quad \int \mu(q, u) dq = u,$$

$$(4.11b) \quad \int q \mu(q, u) dq = f(u).$$

Then we simply set

$$(4.12a) \quad \phi_0(x, q) = \mu(q, u_0(x)).$$

Substituting (4.12a) into (4.9) and using (4.11) shows that (4.10) is satisfied.

Equation (4.8) can be solved explicitly; as in the Godunov-type schemes, averages of these explicit solutions will be used to approximate solutions of (1.1a).

Let  $v(x)$  be an approximation at  $t_n$  to the solution  $u$  of (1.1a). We define the initial value  $\phi_0$  by (4.12a):

$$(4.12b) \quad \phi_0(x, q) = \mu(q, v(x)).$$

Using this in formula (4.6b) for the exact solution of equation (4.6a) gives

$$(4.12c) \quad \phi(x, t, q) = \mu(q, v(x - qt)).$$

Substituting this into (4.7a) results in

$$(4.13) \quad z(x, t) = \int \mu(q, v(x - qt)) dq.$$

We assume that  $v = v^n$  is piecewise constant. We define an approximation  $v^{n+1}$  to  $u$  at  $t_{n+1} = t_n + \tau$  that is piecewise constant on each interval  $I_j = (x_{j-1/2}, x_{j+1/2})$  by defining the value  $v_j^{n+1}$  of  $v^{n+1}$  on  $I_j$  to be the average of  $z(x, \tau)$  over  $I_j$ . Using (4.13) we get

$$(4.14) \quad v_j^{n+1} = |I_j|^{-1} \int_{I_j} \int \mu(q, v^n(x - q\tau)) dq dx.$$

We show next how to express (4.14) as a scheme in conservation form. We integrate (4.6a) over the rectangle  $I_j \times (0, \tau)$ ; we get

$$\int_{I_j} \phi \, dx \Big|_0^\tau + \int_0^\tau q \phi \, dt \Big|_{x_{j-1/2}}^{x_{j+1/2}} = 0.$$

We integrate this with respect to  $q$ , obtaining the conservation form

$$(4.15a) \quad |I_j|(v_j^{n+1} - v_j^n) + \tau(f_{j+1/2}^n - f_{j-1/2}^n) = 0,$$

where  $v_j^{n+1}$ ,  $v_j^n$  are the values of  $v^{n+1}$ ,  $v^n$  on  $I_j$ , and  $f_{j+1/2}^n$  is defined by

$$\tau f_{j+1/2}^n = \int_0^\tau \int q \phi(x_{j+1/2}, t, q) \, dt \, dq,$$

with  $f_{j-1/2}^n$  defined similarly. Using (4.12c) to express  $\phi$  on the right-hand side we get

$$(4.15b) \quad \tau f_{j+1/2}^n = \int_0^\tau \int q \mu(q, v^n(x_{j+1/2} - qt)) \, dt \, dq.$$

This is the numerical flux associated with the scheme (4.14).

We show now that this numerical flux is consistent with the flux  $f$  in (1.1a). Let's take the case, sure to be satisfied in any scheme of practical significance, that  $\mu$  has bounded  $q$ -support. Then it follows from (4.15b), and the fact that  $v^n$  is piecewise constant, that there is an integer  $N$ , whose value depends on  $\tau$ , such that

$$f_{j+1/2}^n = f(v_{j-N+1}^n, \dots, v_{j+N}^n).$$

Suppose  $v_{j+k}^n = v$ ,  $k = -N+1, \dots, N$ ; then  $v^n(x_{j+1/2} - qt)$  on the right-hand side of (4.15b) equals  $v$  on the  $q$ -support of  $\mu$ ; using (4.11b) we see that the right-hand side of (4.15b) equals  $\tau f(v)$ . This proves consistency.

We turn now to the task of determining the distribution  $\mu$ . Clearly, since only the first two moments of  $\mu$  are specified by conditions (4.11), there is a great deal of leeway. For guidance we turn to the linear case,

$$f(u) = Au, \quad A \text{ constant.}$$

The solution of the linear equation

$$(4.16a) \quad u_t + Au_x = 0,$$

with initial values  $u(x, 0) = u_0(x)$ , has the form

$$(4.16b) \quad u(x, t) = \sum P_i u_0(x - a_i t).$$

Here  $a_j$  are the eigenvalues of  $A$ , and  $P_i$  is projection onto the line spanned by the right eigenvector  $R_i$ . On the other hand, substituting (4.6b) into (4.7a) gives the following expression for  $z$ :

$$(4.17) \quad z(x, t) = \int \phi_0(x - qt, q) \, dq.$$

Clearly, comparing (4.16b) and (4.17) we see that

$$(4.18a) \quad z(x, t) \equiv u(x, t)$$

if

$$(4.18b) \quad \phi_0(x, q) = \sum P_i u_0(x) \delta(q - a_i).$$



It is not hard to show that (4.18a) holds for no other choice of  $\phi_0$  whose support is bounded in  $x$  and  $q$ . Setting (4.18b) for  $\phi_0$  into (4.12a) we conclude that  $\mu$  must be of the form

$$(4.19) \quad \hat{\mu}(q, u) = \sum \delta(q - a_i) P_i u.$$

We turn now to the nonlinear case. As we have shown earlier in this section, the flux can be put in form (4.5a):

$$f(u) = A(u)u$$

where the matrix  $A(u)$  has real eigenvalues  $a_i(u)$  and a complete set of eigenvectors. According to the spectral theory of matrices,

$$(4.20) \quad \sum P_i = I, \quad \sum a_i P_i = A.$$

It follows from this that the distribution  $\hat{\mu}$  defined by (4.19) satisfies relations (4.11) even when  $a_i$  and  $P_i$  are functions of  $u$ .

Since relations (4.11) are linear, the most general  $\mu$  that satisfies (4.11) is of the form

$$(4.21) \quad \mu(q, u) = \hat{\mu}(q, u) + \eta(q, u),$$

where  $\eta$  is any distribution whose first two  $q$ -moments are zero for all values of  $u$ .

We call schemes of form (4.14) Boltzmann-type; we list some of their properties.

(a) Suppose the support of the distribution  $\mu$  is contained in  $|q| \leq Q$ . Suppose for simplicity that the mesh over which we discretize is uniform, i.e., that each interval  $I_j$  has the same length  $\Delta$ . Then it follows easily that (4.14) is a three-point scheme if  $\tau$  is chosen so that

$$(4.22) \quad \tau Q \leq \Delta.$$

For Boltzmann-type schemes, the flux (4.15b) splits naturally into two parts. We define

$$(4.23) \quad \begin{aligned} \mu_+(q, u) &= \begin{cases} \mu(q, u) & \text{for } q \geq 0, \\ 0 & \text{for } q < 0, \end{cases} \\ \mu_-(q, u) &= \begin{cases} 0 & \text{for } q > 0, \\ \mu(q, u) & \text{for } q < 0. \end{cases} \end{aligned}$$

Clearly  $\mu = \mu_+ + \mu_-$ ; substituting this into (4.15b) we obtain a splitting

$$f_{j+1/2} = f_{j+1/2}^+ + f_{j+1/2}^-.$$

If (4.22) holds,  $f_{j+1/2}^+$  depends only on  $v_j$ , and  $f_{j+1/2}^-$  only on  $v_{j+1}$ , as in (4.1b).

(b) Suppose  $f$  is of the form (4.5a) and  $\mu$  is chosen to be of form (4.19). The support of  $\hat{\mu}$  extends from  $a_{\min}$  to  $a_{\max}$ , so  $Q = \max_i |a_i|$ , and the restriction (4.22) is the Courant–Friedrichs–Lewy condition. The decomposition of  $\hat{\mu}$  is

$$\hat{\mu}_+ = \sum_{a_i \geq 0} \delta(q - a_i) P_i u, \quad \hat{\mu}_- = \sum_{a_i < 0} \delta(q - a_i) P_i u.$$

Substituting this into (4.15b) gives

$$f_{j+1/2}^+ = A^+(v_j) v_j \quad f_{j+1/2}^- = A^-(v_{j+1}) v_{j+1}$$

where  $A^+$  and  $A^-$  are the positive and negative parts of the matrix  $A$  defined by (2.7) and (2.9). This is the Steger–Warming scheme (4.3).

(c) Consider the equations of compressible flow in Euler coordinates. In this case the three components of  $\phi$  describe the transport of mass, momentum and energy. Since momentum is mass  $\times$  velocity, and  $q$  is velocity, it is reasonable to stipulate that the second component of  $\phi$  be  $q$  times its first component. In view of (4.12a), this would be the case if and only if the same relation holds for the components of  $\mu$ :

$$\mu^{(2)} = q\mu^{(1)}.$$

We claim that this is true for  $\hat{\mu}$  as given by (4.19). For

$$P_i u = w_i R_i$$

and so, by (4.19),

$$(4.24a) \quad \hat{\mu} = \sum \delta(q - a_i) w_i R_i,$$

$$(4.24b) \quad q\hat{\mu} = \sum \delta(q - a_i) a_i w_i R_i.$$

For the equations of compressible flow, the first component  $f^{(1)}$  of flux and the second conserved quantity  $u^{(2)}$  are equal. This implies that the first row of  $A = f_u$  is  $(0, 1, 0)$ . It follows then from the eigenvalue equation

$$AR_i = a_i R_i$$

that the second component of  $R_i$  is  $a_i$  times its first component. This shows that the second component of (4.24a) equals the first component of (4.24b), as asserted.

The use of  $\hat{\mu}$  as given in (4.19) in a Boltzmann-like difference scheme for the Euler equations goes back to Sanders and Prendergast [21]. Other choices of  $\mu$  are found in [11], [17], and [18].

(d) We analyze now the stability of Boltzmann-type schemes for linear equations with constant coefficients. In this case we take  $\mu$  to depend linearly on  $u$ ,

$$(4.25) \quad \mu(q, u) = M(q)u,$$

$M$  a matrix-valued function. The consistency conditions (4.11) become

$$(4.26) \quad \int M(q) dq = I, \quad \int q M(q) dq = A.$$

The scheme (4.13) is

$$z(x, t) = \int M(q) v(x - qt) dq.$$

Taking the Fourier transform we get

$$\tilde{z}(\xi, t) = \tilde{M}(t\xi) \tilde{v}(\xi).$$

A condition for stability is that  $\tilde{M}(\xi)$  should be power bounded, uniformly for all  $\xi$ . Note that there is no stability restriction of the time step  $t$ ; the reason for this is that Boltzmann-type schemes automatically adjust their domain of dependence.

Note that we have analyzed here the stability of scheme (4.13). The full scheme (4.14) is a combination of (4.13) and projection onto the space of piecewise-constant

functions. The latter decreases every weighted  $L_2$  norm. Therefore if (4.13) decreases some weighted  $L_2$  norm, the combined scheme (4.14) is  $L_2$ -stable.

For the case (4.19), we have

$$M(q) = \sum \delta(q - a_i) P_i \quad \text{and} \quad \tilde{M}(\xi) = \sum e^{ia_j \xi} P_j.$$

If  $A$  is symmetric, the  $P_j$  are orthogonal projections, and  $\|\tilde{M}(\xi)\| \equiv 1$ ; so in this case the scheme is stable. For  $A$  nonsymmetric, stability can be proved by replacing the Euclidean norm by some matrix-weighted norm.

(e) We have not carried out any stability analysis in the nonlinear case, nor studied the interesting question of how to assure the entropy condition.

(f) We conclude by observing that flux-splitting schemes cannot exactly resolve stationary discontinuities. For suppose that the stationary Rankine–Hugoniot condition  $f(u) = f(v)$  is satisfied. It does not follow from this that also

$$(4.27) \quad f^a(u) = f^a(v).$$

But for a split-flux scheme, (4.2b) shows that (4.27) is necessary for the exact resolution of stationary discontinuities.

The split-flux scheme of van Leer [15] for the Euler equations can represent a stationary shock with, at best, one intermediate state, just as the Osher [16] scheme. A contact discontinuity, however, will keep on spreading with the use of any split flux scheme. This follows immediately from the diffusive nature of the underlying Boltzmann model.

## REFERENCES

- [1] R. COURANT, E. ISAACSON AND M. REES, *On the solution of nonlinear hyperbolic differential equations*, Comm. Pure Appl. Math., 5 (1952), pp. 243–255.
- [2] M. G. CRANDALL AND A. MAJDA, *Monotone difference approximations for scalar conservation laws*, Math. Comp., 34 (1980), pp. 1–21.
- [3] J. GLIMM, *Solutions in the large for nonlinear hyperbolic systems of equations*, Comm. Pure Appl. Math., 18 (1965), pp. 697–715.
- [4] S. K. GODUNOV, *A difference scheme for numerical computation of discontinuous solutions of equations of fluid dynamics*, Math. Sbornik, 47 (1959), pp. 271–306. (In Russian.)
- [5] A. HARTEN, *High resolution schemes for hyperbolic conservation laws*, Report DOE/ER/03077-.67, New York Univ., March 1982.
- [6] A. HARTEN, J. M. HYMAN AND P. D. LAX, *On finite-difference approximations and entropy conditions for shocks*, Comm. Pure Appl. Math., 29 (1976), pp. 297–322.
- [7] A. HARTEN AND P. D. LAX, *A random choice finite-difference scheme for hyperbolic conservation laws*, SIAM J. Numer. Anal., 18 (1981), pp. 289–315.
- [8] A. HARTEN, *On the symmetric form of systems of conservation laws with entropy*, ICASE Report 81-34, ICASE, Hampton, VA, 1981.
- [9] A. HARTEN AND J. M. HYMAN, *A self-adjusting grid for the computation of weak solutions of hyperbolic conservation laws*, Report LA9105, Center for Nonlinear Studies, Theoretical Division, Los Alamos National Lab., Los Alamos, NM, 1981.
- [10] L. C. HUANG, *Pseudo-unsteady difference schemes for discontinuous solutions of steady-state one-dimensional fluid dynamics problems*, J. Comp. Phys., 42 (1981), pp. 195–211.
- [11] S. KANIEL AND J. FALCOVITZ, *Transport approach for compressible flow*, Proc. 4th International IRIA Symposium on Computing Methods in Science and Engineering, Versailles, Dec. 1979, R. Glowinski and P. L. Lions, eds., North Holland, Amsterdam, 1981.
- [12] P. D. LAX, *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*, CBMS Regional Conference Series in Applied Mathematics 11, Society for Industrial and Applied Mathematics, Philadelphia, 1972.

- [13] B. VAN LEER, *Towards the ultimate conservative difference scheme. III*, J. Comp. Phys., 23 (1977) pp. 263–275.
- [14] ———, *Upwind differencing for hyperbolic systems of conservation laws*, in Numerical Methods for Engineering, Vol. 1, Dunod, Paris, 1980, pp. 137–149.
- [15] ———, *Flux-vector splitting for the Euler equations*, presented at the 8th Internat. Conference on Numerical Methods for Engineering, Aachen, June, 1982.
- [16] S. OSHER, *Numerical solution of singular perturbation problems and hyperbolic systems of conservation laws*, in Mathematics Studies, 47, North-Holland, Amsterdam, 1981, pp. 179–205.
- [17] D. J. PULLIN, *Direct simulation methods for compressible inviscid ideal gas flow*, J. Comp. Phys., 34 (1980), pp. 231–244.
- [18] R. D. RIETZ, *One-dimensional compressible gas dynamics calculations using the Boltzmann equation*, J. Comp. Phys., 42 (1981), pp. 108–123.
- [19] P. L. ROE, *The use of the Riemann problem in finite-difference schemes*, in Proc. 7th International Conference on Numerical Methods in Fluid Dynamics, Stanford/NASA Ames, June 1980, Lecture Notes in Physics, 141, Springer-Verlag, New York, 1981, pp. 354–359.
- [20] ———, *Approximate Riemann solvers, parameter vectors, and difference schemes*, J. Comp. Phys., 43 (1981), pp. 357–372.
- [21] R. H. SANDERS AND K. H. PRENDERGAST, *On the origin of the 3-kiloparsec arm*, Ap. J., 188 (1974), pp. 489–500.
- [22] J. L. STEGER AND R. F. WARMING, *Flux-vector splitting of the inviscid gas dynamic equations with applications to finite-difference methods*, J. Comp. Phys., 40 (1981), pp. 263–293.
- [23] J. L. STEGER, *A preliminary study of relaxation methods for the inviscid conservative gas dynamics equations using flux-vector splitting*, Report 80-4, Flow Simulations, Inc., Sunnyvale, CA, August 1980.