

# Computational Methods for Hyperbolic Conservation Laws

E. Bruce Pitman  
University at Buffalo  
*pitman@buffalo.edu*

## 1 Introduction

These notes concern hyperbolic conservation laws. Conservation laws are PDEs with a particular structure. In one space dimension these take the form

$$\partial_t u(x, t) + \partial_x f(u(x, t)) = 0 \tag{1}$$

where  $u : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^m$  is a vector of conserved variables (or state variables). For fluid dynamics,  $u$  is the vector of mass, momentum and energy densities - so that  $\int_a^b u_j(x, t) dx$  is the total quantity of the  $j^{\text{th}}$  state variable in the interval at time  $t$ . Because these variables are conserved,  $\int_{-\infty}^{\infty} u_j(x, t) dx$  should be constant in  $t$ . The function  $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$  is the flux function, which gives the rate of flow of the conserved variables at any point.

The PDEs must be augmented by initial data,  $u(x, 0) = u_0(x)$ . Then the equation (1) and this initial data constitute the Cauchy problem. If (1) holds only on an interval  $[a, b] \in \mathbb{R}$ , boundary data must be specified.

We assume (1) is *hyperbolic*. That is, we assume the  $m \times m$  Jacobian matrix  $A = f' = \frac{df}{du}$  has  $m$  real eigenvalues  $\lambda_j(u)$ ,  $j = 1, \dots, m$  and a complete set of linearly independent eigenvectors. This means that  $A$  is diagonalizable. The system is called strictly hyperbolic if these eigenvalues are distinct.

In two space dimensions, the conservation law takes the form

$$\partial_t u(x, y, t) + \partial_x f(u(x, y, t)) + \partial_y g(u(x, y, t)) = 0 \tag{2}$$

Hyperbolicity requires that any linear combination  $\alpha f' + \beta g'$  be diagonalizable for real  $\alpha$  and  $\beta$ .

## 2 Examples

The Euler equations of gas dynamics can be written

$$\partial_t \begin{bmatrix} \rho \\ \rho u \\ E \end{bmatrix} + \partial_x \begin{bmatrix} \rho u \\ \rho u^2 + p \\ u(E + p) \end{bmatrix} = 0 \tag{3}$$

Here  $\rho$  is the mass density,  $u$  a velocity,  $E$  the energy, and  $p$  the pressure, which is a specified function of the other variables.

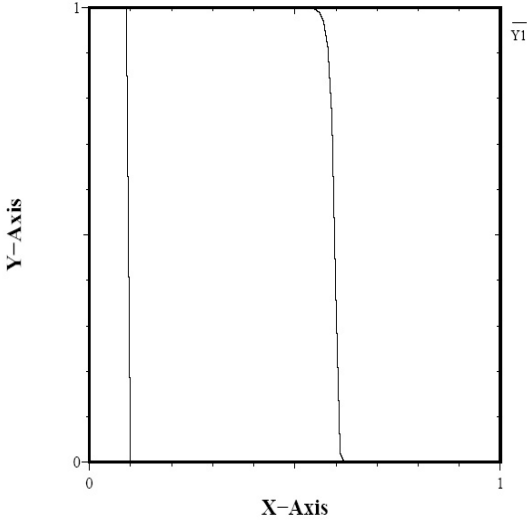


Figure 1: A shockwave solution to Burgers' equation, computed with a high order accurate numerical method.

Isentropic gas dynamics means that the entropy ( $S = c_v \log(p/\rho^\gamma) + \text{const}$ ) of the system is constant, and implies the energy is a specified function of the momentum  $\rho u$  and the density. The system simplifies

$$\partial_t \begin{bmatrix} \rho \\ \rho u \end{bmatrix} + \partial_x \begin{bmatrix} \rho u \\ \rho u^2 + p \end{bmatrix} = 0 \quad (4)$$

where  $p = \rho^\gamma$ .

A special system is isothermal flow, where  $p = a^2 \rho$ .

Another example is the shallow water equations

$$\partial_t \begin{bmatrix} h \\ hu \end{bmatrix} + \partial_x \begin{bmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \end{bmatrix} = 0 \quad (5)$$

where  $g$  is the gravitational constant.

A scalar equation has the form  $\partial_t u + \partial_x f(u) = 0$  where  $a(u) = f'$ . There is one often used nonlinear scalar example, Burgers' equation

$$\partial_t u + \frac{1}{2} \partial_x u^2 = 0 \quad (6)$$

We say the equation is *genuinely nonlinear* if  $a' \neq 0$ .

Of course the linear advection equation is

$$\partial_t u + \partial_x (au) = 0 \quad (7)$$

where  $a$  is a constant.

We sometimes consider diffusive effects. In this case, the flux includes a gradient term. For example, in Burgers' equation,  $f(u) = \frac{1}{2}u^2 - \mu \partial_x u$  where  $\mu$  is the viscosity, so the diffusive version is

$$\partial_t u + \partial_x \left( \frac{1}{2}u^2 - \mu \partial_x u \right) = 0 \quad (8)$$

which we rewrite as

$$\partial_t u + \frac{1}{2} \partial_x u^2 = \mu \partial_x^2 u \quad . \quad (9)$$

Finally, several complicating factors that will arise near the end of our discussions are nicely modeled by a cubic nonlinearity,

$$\partial_t u + \frac{1}{3} \partial_x u^3 = 0 \quad . \quad (10)$$

### 3 The Linear Advection Equation

Let us examine the linear equation

$$\partial_t u + \partial_x (au) = 0 \quad (11)$$

where  $a$  is a positive constant. To solve the Cauchy problem, we specify initial data  $u(x, 0) = u_0(x)$ . Now the equation can be interpreted as saying that the total derivative of  $u$  is zero

$$\frac{du(x(t), t)}{dt} = \partial_t u(x(t), t) + \partial_x u(x(t), t) \frac{dx}{dt} = 0$$

along characteristic curves given by  $\frac{dx}{dt} = a$ . It is then clear that the solution is

$$u(x, t) = u_0(x - at) \quad .$$

So the differential equation simply translates the initial data to the right with speed  $a$ .

For non-constant speed  $a = a(x)$ , we can write the equation as

$$\partial_t u + a(x) \partial_x u = -a'(x)u \quad . \quad (12)$$

That is, along the characteristic  $\frac{dx}{dt} = a(x)$ ,  $u$  is not constant by changes according to

$$\frac{du}{dt} = -a'u \quad . \quad (13)$$

Given a particular point  $(\bar{x}, \bar{t})$ , the *domain of dependance* is the set of all  $x$ s at a time  $t < \bar{t}$  such that characteristics emanating from  $(x, t)$  reach  $(\bar{x}, \bar{t})$ . Notice that if we examine the initial time  $t = 0$ , we could change the initial data outside the domain of dependance and not change the solution  $u(\bar{x}, \bar{t})$ .

A related idea is the *range of influence*. If we consider a point  $(\bar{x}, 0)$ , the range of influence is the set of all  $(x, t)$ ,  $t > 0$  that can be reached by characteristics from  $(\bar{x}, 0)$ .

## 4 Burgers' Equation

The viscous Burgers' equation can be solved by the Hopf-Cole transformation, which converts the equation into a linear diffusion equation. We will examine in some detail the inviscid Burgers' equation

$$\partial_t u + \frac{1}{2} \partial_x u^2 = \partial_t u + u \partial_x u = 0 \quad . \quad (14)$$

As for the linear advection equation, we can interpret this as

$$\frac{du}{dt} = 0 \quad \text{along characteristics} \quad \frac{dx}{dt} = u \quad . \quad (15)$$

Consider smooth initial data  $u(x, 0) = u_0(x)$ . Then the characteristic solution implies (see the solution for the linear advection equation)

$$u(x, t) = u_0(x - u(x, t)t) \quad (16)$$

If we now differentiate

$$\frac{du}{dx} = \frac{d}{dx} u_0(x - u(x, t)t) = u'_0 - t \frac{du}{dx} u'_0 = 0 \quad .$$

Solving we find

$$\frac{du}{dx} = \frac{u'_0}{1 + tu'_0} \quad .$$

That is,  $u'$  blows up - even for arbitrarily smooth  $u_0$  - in finite time if the initial data is ever decreasing. One sees that characteristics intersect at this blow-up time, and the solution becomes multi-valued.

To proceed, we generalize the notion of solution. For smooth test functions  $\phi(x, t) \in C_0^1(x, t)$  with compact support, multiply and integrate to find

$$\int_0^\infty \int_{-\infty}^\infty \{\phi \partial_t u + \phi \partial_x f(u)\} dx dt = 0$$

Integrating by parts, we define a weak solution  $u$  one that satisfies

$$\int_0^\infty \int_{-\infty}^\infty \{u \partial_t \phi + f(u) \partial_x \phi\} dx dt = \int_{-\infty}^\infty u(x, 0) \phi(x, 0) dx$$

## 5 The Riemann Problem

The Riemann problem consists of solving the conservation law 1 for special piecewise constant initial data

$$u(x, 0) = \begin{cases} u_L & x < 0 \\ u_R & x > 0 \end{cases} \quad (17)$$

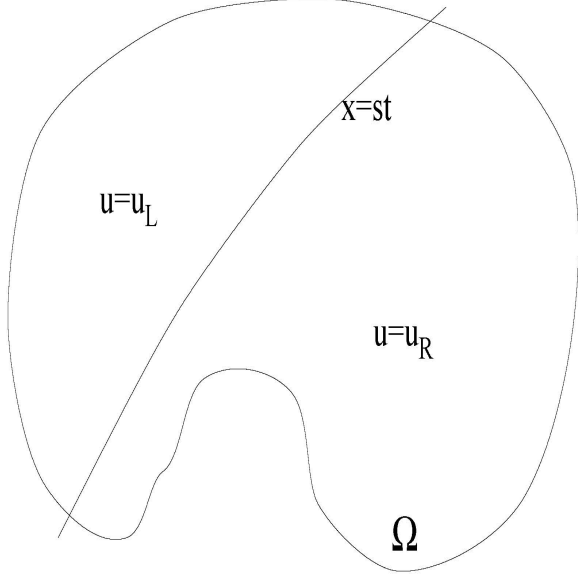


Figure 2: A weak solution  $u$  consisting of piecewise constant states  $u_L$  and  $u_R$  separated by a shock with speed  $s$ .

For Burgers' equation, then, characteristics on the left of the origin travel with speed  $u_L$ , while those on the right travel with speed  $u_R$ .

Now consider a weak solution  $u$  in a compact region  $\Omega$  in the  $(x, t)$ -plane. Consider the solution  $u = u_L$  for  $x < st$ , and  $u = u_R$  for  $x > st$ . By Green Theorem, we find

$$\begin{aligned}
0 &= \int_{\Omega} \{u \partial_t \phi + f(u) \partial_x \phi\} dx dt & (18) \\
&= \int_{x < st} (\partial_t u + \partial_x f) \phi dx dt + \int_{x > st} (\partial_t u + \partial_x f) \phi dx dt \\
&\quad - \int_{x=st^-} \{u_L \phi dx - f(u_L) \phi dt\} - \int_{x=st^+} \{u_R \phi dx - f(u_R) \phi dt\} \\
&= - \int_{T^-}^{T^+} (u_L \phi s - f(u_L) \phi) dt + \int_{T^-}^{T^+} (u_R \phi s - f(u_R) \phi) dt \\
&= - \int_{T^-}^{T^+} \phi [(u_R - u_L) s - (f(u_R) - f(u_L))] dt
\end{aligned}$$

This implies that the shock speed

$$s = \frac{f(u_R) - f(u_L)}{u_R - u_L} = \frac{[f]}{[u]} \quad (19)$$

This is called the Rankine-Hugoniot condition.

For Burgers' equation, we see shocks travel at a speed

$$s = \frac{\frac{1}{2}u_R^2 - \frac{1}{2}u_L^2}{u_R - u_L} = \frac{u_R + u_L}{2}$$

Notice that the form of the equation impacts the shock speed. For example, given the non-conservative form  $\partial_t u + u \partial_x u = 0$ , multiplying by a strong solution  $u$ , we have  $\partial_t \frac{u^2}{2} + \partial_x \frac{u^3}{3} = 0$ . This equation has a shock speed of  $\frac{2}{3} \frac{u_R^2 + u_R u_L + u_L^2}{u_R + u_L}$ . So we need to be sure we understand the origin of the conservation law, that we have the form correct.

## 6 Shock and Rarefaction Solutions

Consider Burgers' equation with Riemann data

$$u(x, 0) = \begin{cases} 1 & x < 0 \\ 0 & x > 0 \end{cases}$$

The discontinuity persists in  $t > 0$ , and the solution consists of the constant states  $u = 1$  for  $x < st$ , and  $u = 0$  for  $x > st$ , where the shock propagates with speed  $s = \frac{1}{2}$ .

Now consider the Riemann data

$$u(x, 0) = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases}$$

One might guess a similar solution  $u = 0$  for  $x < st$ , and  $u = 1$  for  $x > st$ , where again  $s = \frac{1}{2}$ .

There exists another solution, a rarefaction wave. Look for a solution  $u = u(\xi)$  of the similarity variable  $\xi = \frac{x}{t}$ . Substituting,

$$\begin{aligned} -\frac{x}{t^2} u' + \frac{1}{t} u u' &= 0 \\ (u - \xi) u' &= 0 \end{aligned} \tag{20}$$

So we have a solution  $u = \xi$  that is constant along rays from the origin. The characteristics on the left of the origin are vertical, those on the right travel with speed 1. So we can fit the similarity solution  $u = \xi$  into the wedge  $0 < x < t$ . The resulting solution is continuous, though not differentiable along  $x = 0$  nor  $x = t$ .

Thus we see that weak solutions to conservation laws are not unique.

To choose among these solutions, there are several so-called *entropy conditions*, each motivated by physics considerations.

**Lax Geometric Entropy Condition**  $u$  is the correct solution if it satisfies the characteristic condition: The characteristics on either side of a curve of discontinuity, when continued forward in time, must intersect the curve.

The Lax condition implies  $a(u_L) > s > a(u_R)$ , so for Burgers equation we have  $u_L > s > u_R$ . For this second Riemann problem, therefore, the correct solution is the rarefaction wave. Note that a shock is the correct solution to the first Riemann problem.

A second entropy condition is due to Liu:

**Liu Entropy Condition**  $u$  is the entropy solution if, for all  $u$  between  $u_L$  and  $u_R$ ,

$$\frac{f(u) - f(u_R)}{u - u_R} \leq s \leq \frac{f(u) - f(u_L)}{u - u_L} \tag{21}$$

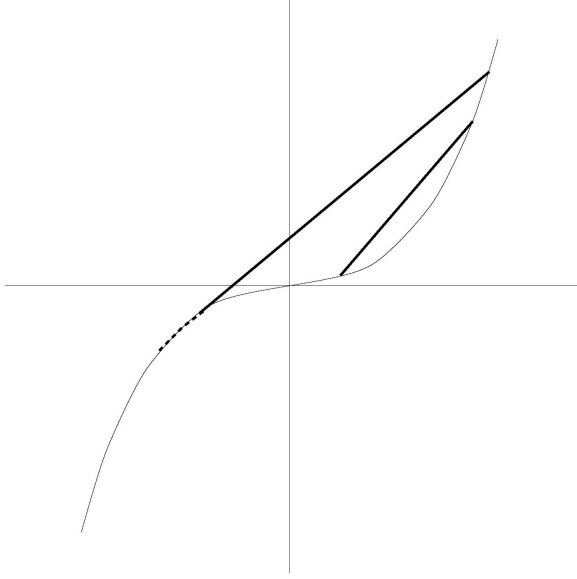


Figure 3: A cubic flux function. If  $u_L > 0$  and  $u_R$  is sufficiently negative, the solution is a compound wave - a shock followed immediately by a rarefaction.

The third condition is from Olenik:

**Olenik Entropy Condition**  $u$  is the entropy solution if there exists  $E > 0$  such that for all  $a > 0$  and for all  $t > 0$ ,

$$\frac{u(x+a, t) - u(x, t)}{a} < \frac{E}{t} \quad (22)$$

Although not explicit in these conditions, it is hoped that the entropy solution is the  $\epsilon = 0$  limit of the diffusion equation  $\partial_t u + \partial_x f(u) = \epsilon \partial_x^2 u$  with the same initial data.

It is something of a metatheorem that for strictly hyperbolic, genuinely nonlinear systems, all entropy conditions give rise to the same admissible discontinuities, and these conditions are somehow related to viscous dissipation and (physical) entropy production. The Kruskov theory for scalar equations  $\partial_t w + \partial_x g(w) = 0$  states there exists a unique entropy solution, which is the limit, as  $\epsilon \rightarrow 0$  of  $\partial_t w + \partial_x g(w) = \epsilon \partial_x^2 w$ . That is, the metatheorem is true for scalars. The conditions of Lax, Liu and Olenik are (essentially) equivalent for genuinely nonlinear scalar equations.

## 7 Cubic Equation

Burgers' equation has a convex flux function,  $f(u) = \frac{1}{2}u^2$ . The solution to the Riemann problem is a shock if  $u_L > u_R$ , and a rarefaction if  $u_L < u_R$ .

A more complex picture emerges if the flux is cubic,  $f(u) = \frac{1}{3}u^3$ . Of course the flux is both convex and concave in different regions, and the nature of the solution depends on the relative location of  $u_L, u_R$ . Although not a complete description of the solution, several cases describe the essential features.

1. If  $u_L > 0, u_R > 0$  and  $u_L < u_R$ , the solution is a rarefaction.

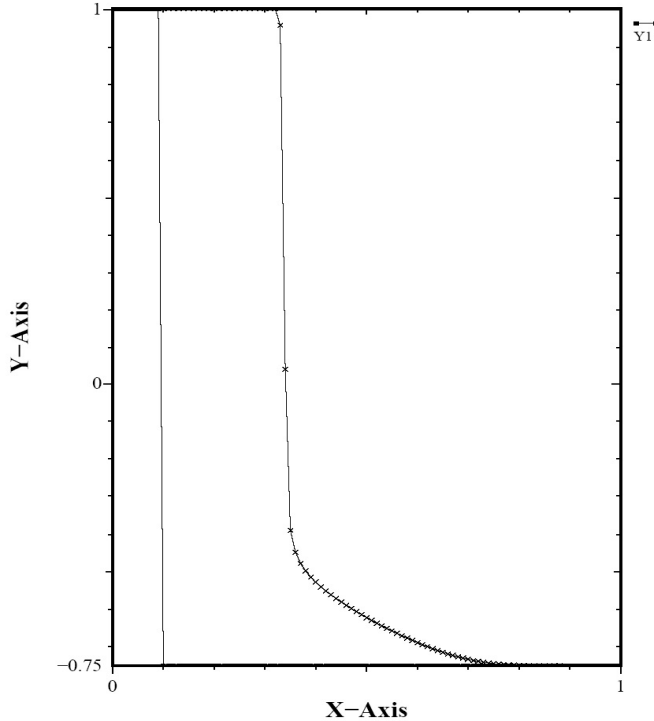


Figure 4: An example of a compound wave, computed using a first order upwind method.

2. If  $u_L > 0$  and  $u_L > u_R > u_*$ , the solution is a shock. Here  $u_*$  is that  $u$  such that the line segment from  $u_L$  is tangent to  $f$ .
3. If  $u_L > 0$  and  $u_L > u_* > u_R$ , the solution is a compound wave. That is, the solution consists of a shock from  $u_L$  to  $u_*$  immediately followed by a rarefaction from  $u_*$  to  $u_R$ .
4. Similar considerations hold if  $u_L < 0$ , where the relationship of  $u_L$  and  $u_R$  are reversed.

## 8 Linear Systems

It is useful to recall the structure of a linear hyperbolic system

$$\partial_t u + A \partial_x u = 0 \tag{23}$$

where  $A$  is an  $m \times m$  matrix with real eigenvalues  $\lambda_1 \leq \lambda_2 \leq \dots < \lambda_m$  and a complete set of (right) eigenvectors  $r_1, \dots, r_m$ . We can diagonalize  $A$  as

$$\Lambda = R^{-1} A R \tag{24}$$

where  $\Lambda$  is the diagonal matrix with the eigenvalues along the diagonal, and  $R$  is the matrix of eigenvectors, so that  $A r_j = \lambda_j r_j$ .

Converting the equation into characteristics variables  $v = R^{-1}u$  find

$$\begin{aligned} R^{-1}\partial_t u + R^{-1}A\partial_x u &= 0 \\ \partial_t v + \Lambda\partial_x v &= 0 \end{aligned} \tag{25}$$

Each variable is decomposed from the others, yielding  $m$  equations  $\partial_t v_j + \lambda_j \partial_x v_j = 0$ . The solution  $v_j(x, t) = v_j(x - \lambda_j t, 0)$ . Then the solution  $u(x, t) = Rv(x, t)$ . Writing this out,

$$\begin{aligned} u(x, t) &= \sum_{j=1}^m v_j(x, t)r_j \\ u(x, t) &= \sum_{j=1}^m v_j(x - \lambda_j t, 0)r_j \end{aligned} \tag{26}$$

To solve the Riemann problem with data  $u_L$  and  $u_R$ , decompose this data  $u_L = \sum_{j=1}^m \alpha_j r_j$  and  $u_R = \sum_{j=1}^m \beta_j r_j$ . Then

$$v_j(x, 0) = \begin{cases} \alpha_j & x < 0 \\ \beta_j & x > 0 \end{cases} \tag{27}$$

and so

$$v_j(x, t) = \begin{cases} \alpha_j & x - \lambda_j t < 0 \\ \beta_j & x - \lambda_j t > 0 \end{cases} \tag{28}$$

Let  $J = J(x, t)$  be the maximum value of  $j$  for which  $x - \lambda_j t > 0$ . Then

$$u(x, t) = \sum_{j=1}^{J(x,t)} \beta_j r_j + \sum_{j=J+1}^m \alpha_j r_j \tag{29}$$

Across the  $j^{\text{th}}$  characteristic, the solution jumps with  $[u] = (\beta_j - \alpha_j)r_j$ . The solution can be written in alternative forms

$$\begin{aligned} u(x, t) &= u_L + \sum_{\lambda_j < x/t} (\beta_j - \alpha_j)r_j \\ &= u_R - \sum_{\lambda_j < x/t} (\beta_j - \alpha_j)r_j \end{aligned} \tag{30}$$

We can view “solving the Riemann problem” as decomposing the initial discontinuity

$$u_R - u_L = (\beta_1 - \alpha_1)r_1 + \dots + (\beta_m - \alpha_m)r_m \tag{31}$$

We will solve nonlinear problems in a similar fashion, especially in numerical approaches to conservation laws.

## 9 Nonlinear Systems

We consider

$$\partial_t u + \partial_x f(u) = 0 \quad (32)$$

where the Jacobian  $A = \frac{df}{du}$  has real, distinct eigenvalues  $\lambda_j(u)$ . The  $j^{\text{th}}$  variable is called genuinely nonlinear if  $r_j(u) \cdot \nabla \lambda_j(u) \neq 0$ ; by convention, we assume  $\frac{d\lambda_j}{du} > 0$ . [Note: For a scalar equation, this means  $f'' \neq 0$ .] Thus in each eigen-direction, the speed varies monotonically. We will also normalize the eigenvectors so that  $\|r_j(u)\| = 1$ . If a discontinuity has constant states  $\hat{u}$  and  $\tilde{u}$  on either side, the Rankine-Hugoniot condition must hold  $f(\tilde{u}) - f(\hat{u}) = s(\tilde{u} - \hat{u})$ . If the state  $\hat{u}$  on the left is fixed, these provide  $m$  equations for the  $m + 1$  unknowns  $\tilde{u}$  and  $s$ , a one parameter family of solutions indexed by  $\eta$ ,  $u_j(\eta, \hat{u})$ , where  $u_j(0, \hat{u}) = \hat{u}$ . Let  $s_j(\eta)$  be the corresponding shock speed. The Rankine-Hugoniot condition can be written

$$f(\tilde{u}_j(\eta)) - f(\hat{u}) = s_j(\eta)(\tilde{u}_j(\eta) - \hat{u}) \quad (33)$$

Differentiate with respect to  $\eta$  and evaluate at  $\eta = 0$  to find

$$f'(\hat{u})\tilde{u}'_j(0) = s_j(0)\tilde{u}'_j(0) \quad (34)$$

so the jump is (approximately) a multiple of  $r_j(\hat{u})$ , and the speed is (approximately)  $\lambda_j(\hat{u})$ . If  $f$  is smooth, it can be shown by an implicit function argument that these solutions exist locally near  $\hat{u}$ , and that  $r_j$  and  $s_j$  are smooth (see Lax or Smoller). These curves are called the Hugoniot locus. If  $\tilde{u}_j$  lies on the  $j^{\text{th}}$  Hugoniot locus, then  $\hat{u}$  and  $\tilde{u}_j$  are connected by a shock in the  $j^{\text{th}}$  family.

Even for a nonlinear system, it might happen that the characteristic speed is constant in one of the directions - that is,

$$r_j(u) \cdot \nabla \lambda_j(u) \equiv 0 \quad (35)$$

we say the  $j^{\text{th}}$  field is linearly degenerate. A discontinuity in this field is called a contact discontinuity. We generalize the Lax entropy condition and admit discontinuous solutions that satisfy

$$\lambda_j(u_L) \geq s \geq \lambda_j(u_R) \quad (36)$$

We also consider rarefaction solutions  $u(x, t) = w(\xi)$  where  $\xi = \frac{x}{t}$ . Substituting and rearranging we find

$$\begin{aligned} -\frac{x}{t^2}w' + \frac{1}{t}f'(w)w' &= 0 \\ f'(w)w' &= \xi I w' \end{aligned} \quad (37)$$

That is,  $w'$  is proportional to some eigenvector,

$$w'(\xi) = \alpha(\xi)r_j(w(\xi)) \quad (38)$$

and  $\xi = \lambda_j(w(\xi))$ . Hence rarefaction solutions lie along integral curves of  $r_j$ . By our assumption of genuine nonlinearity, the speed  $\lambda_j$  varies monotonically as  $\xi$  increases, and we

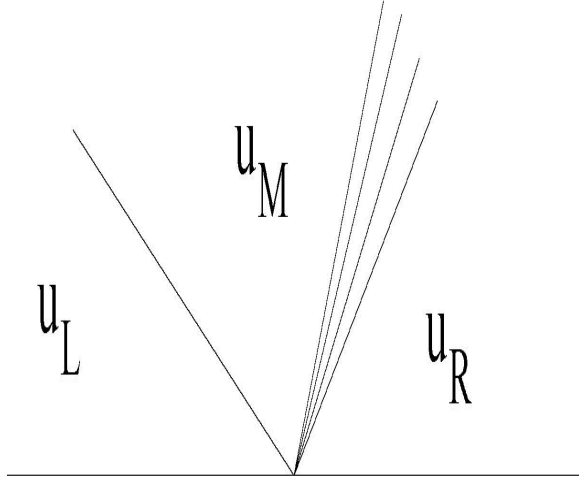


Figure 5: An example of a solution to a  $2 \times 2$  system, consisting of a left-going shock connecting  $u_L$  to an intermediate state  $u_M$ , and a rarefaction from  $u_M$  to  $u_R$ .

can connect left and right states if they lie along the same integral curve and  $\lambda_j(u_L) < \lambda_j(u_R)$ . Finally, differentiating  $\xi = \lambda_j(w(\xi))$ , one finds

$$\begin{aligned} 1 &= \nabla \lambda_j(w(\xi)) \cdot w'(\xi) \\ &= \alpha(\xi) \nabla \lambda_j(w(\xi)) \cdot r_j(w(\xi)) \end{aligned} \quad (39)$$

and we find the ODE for  $w$

$$w'(\xi) = \frac{r_j(w(\xi))}{\nabla \lambda_j(w(\xi)) \cdot r_j(w(\xi))}, \quad \xi_1 \leq \xi \leq \xi_2 \quad (40)$$

with data  $w(\xi_1) = u_L$ .

For the general Riemann problem for nonlinear system, the solution is constructed by piecing together shocks and rarefactions, one in each family. In this fashion, one connects  $u_L$  to  $u_R$  by a set of  $m - 1$  intermediate states. Each state is connected to its neighbor either as a state along a Hugoniot locus or as a state along an integral curve. Because of non-uniqueness, both shock and rarefaction curves exist at each state; entropy conditions are used to discard the non-physical solutions.

## 9.1 Example

To better understand the Riemann problem for systems, consider the system of equations describing a nonlinear wave in one dimension

$$\begin{aligned} \partial_t z - \partial_x v &= 0 \\ \partial_t v - \partial_x \sigma(z) &= 0 \end{aligned} \quad (41)$$

The Jacobian matrix is

$$A = \begin{pmatrix} 0 & -1 \\ -\sigma'(z) & 0 \end{pmatrix} \quad (42)$$

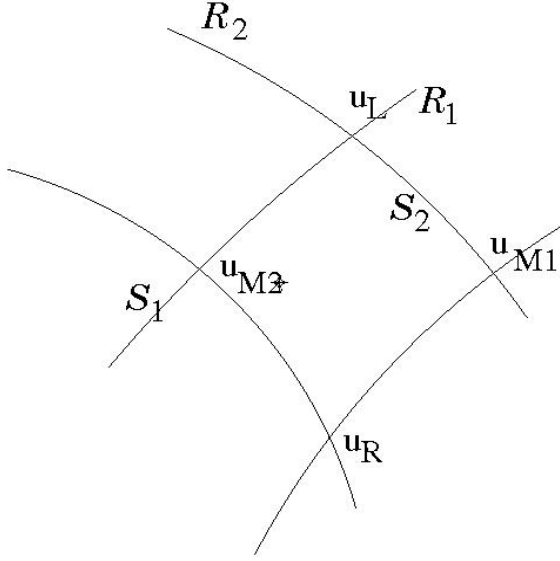


Figure 6: The construction of a solution to the Riemann problem. The rarefaction curve and shock curve have twice-differentiable contact at  $u_L$ . To solve the problem, follow a 1-wave from  $u_L$  to the intermediate state  $u_M$ , and then a two wave from  $u_M$  to  $u_R$ .

with eigenvalues  $\pm\sqrt{\sigma'(z)}$  and right eigenvectors  $r_{\pm} = (1, -\pm\sqrt{\sigma'})^T$ . The Rankine-Hugoniot conditions read

$$\begin{aligned} s(z_2 - z_1) &= -(v_2 - v_1) \\ s(v_2 - v_1) &= -(\sigma(z_2) - \sigma(z_1)) \end{aligned} \quad (43)$$

This means that the shock speed is

$$s_{\pm} = \pm \sqrt{\frac{\sigma(z_2) - \sigma(z_1)}{z_2 - z_1}} \quad (44)$$

It is clear that  $s_{\pm} \rightarrow \lambda_{\pm}$  as  $z_2 \rightarrow z_1$ . Moreover,

$$\lambda_{\pm}' = \frac{\sigma''}{2\sqrt{\sigma'}} = 2s_{\pm}' \quad (45)$$

To be genuinely nonlinear, then, requires  $\sigma'' > 0$  or  $\sigma'' < 0$ . Physics suggests the latter is the better choice.

First we consider shocks. Consider the 1-state as fixed, and parameterize the shock by  $z_2$ . So the solution is given by the shock speed above and

$$v_{2\pm} = v_1 + s_{\pm}(z_1 - z_2) \quad (46)$$

Assuming  $\sigma'' < 0$ , two cases present themselves:  $\sigma'(z_2) < s^2 < \sigma'(z_1)$  and  $\sigma'(z_1) < s^2 < \sigma'(z_2)$ . Using the Lax entropy condition, we find for shocks in the +family

$$\lambda_+(z_2) < s_+ < \lambda_+(z_1) \quad (47)$$

This is true only if  $z_2 > z_1$ . For the  $-$ family, similar arguments show  $z_2 < z_1$ . Then the shock loci are

$$\begin{aligned} s_+ &= \{(z_2, v_2) : v_2 = v_1 - s_+(z_2 - z_1) \quad z_2 > z_1\} \\ s_- &= \{(z_2, v_2) : v_2 = v_1 - s_-(z_2 - z_1) \quad z_2 < z_1\} \end{aligned} \quad (48)$$

For rarefaction waves,

$$Au' = \xi I u' \quad (49)$$

$\xi = \lambda_{\pm}(u(\xi))$  and  $u' = \alpha r_{\pm}$ . We find from our general discussion

$$\begin{pmatrix} z \\ v \end{pmatrix}' = \frac{r_{\pm}}{r_{\pm} \cdot \nabla \lambda_{\pm}} \quad (50)$$

For the  $+$ family,

$$\frac{dv}{dz} = -\sqrt{\sigma'(z)} \Rightarrow v = v_1 - \int_{z_1}^{z_2} \sqrt{\sigma'(z)} dz \quad (51)$$

For these rarefactions,  $z_2 < z_1$ . Similarly, for the  $-$ family,

$$\frac{dv}{dz} = \sqrt{\sigma'(z)} \Rightarrow v = v_1 + \int_{z_1}^{z_2} \sqrt{\sigma'(z)} dz \quad (52)$$

and  $z_2 > z_1$ .

## 9.2 Entropy

The entropy conditions of Olenik cannot be extended to systems of equations. The Liu condition can be generalized. The Lax condition is easily applied - one examines the characteristics impinging on a shock. There is another notion of entropy that is important. Given the conservation law (1), an entropy-entropy flux pair is a pair of scalar functions  $\eta(u)$ ,  $\psi(\eta(u))$  that satisfy the equation

$$\partial_t \eta + \partial_x \psi = 0 \quad (53)$$

Differentiating,

$$\nabla \eta \cdot \partial_t u + \nabla \psi \cdot \partial_x u = 0 \quad (54)$$

To achieve this from the nonconservative form  $\partial_t u + A \partial_x u = 0$ , we find  $\nabla \eta \cdot A = \nabla \cdot \psi$ . This is a system of  $m$  PDEs for  $\eta$  and  $\psi$ , and is in general overdetermined for  $m \geq 2$ . However for symmetric systems - that is, when  $A$  is symmetric, so  $\partial f_j / \partial u_k = \partial f_k / \partial u_j$  - the equations may be solved. Then there is a function  $h$  such that

$$\frac{\partial h}{\partial u_k} = f_k \quad (55)$$

and

$$\eta = \sum u_j^2 \quad \psi = \sum u_j f_j - h \quad (56)$$

satisfies the equations.

Consider now adding to (1) a small dose of viscosity

$$\partial_t u + A \partial_x u = \epsilon \partial_x^2 u \quad (57)$$

Multiply through by  $\nabla \eta$ , and assume the existence of an entropy-entropy flux pair to find

$$\partial_t \eta + \partial_x \psi = \epsilon \nabla \eta \cdot \partial_x^2 u \quad (58)$$

Now

$$\partial_x^2 \eta = \nabla \eta \partial_x^2 u + \frac{\partial^2 \eta}{\partial u_j \partial u_k} \partial_x u_j \partial_x u_k \quad (59)$$

Assume  $\eta$  is *convex*, so the matrix of second derivatives is positive definite. Then

$$\partial_x^2 \eta \geq \nabla \eta \cdot \partial_x^2 u \quad (60)$$

This implies

$$\partial_t \eta + \partial_x \psi \leq \epsilon \partial_x^2 \eta \quad (61)$$

Now let  $\epsilon \rightarrow 0$  to find the entropy inequality  $\partial_t \eta + \partial_x \psi \leq 0$ . One can also show that, across a discontinuity,

$$s[u] - [f(u)] \leq 0 \quad (62)$$

## 10 Classical Numerical Methods

We restrict attention for the moment to a scalar linear equation

$$\partial_t u + a \partial_x u = 0 \quad (63)$$

for  $x \in [a, b]$  and  $t > 0$ . We discretize space into  $N$  equal subintervals of width  $\Delta x = \frac{b-a}{N}$ . We discretize space into steps of size  $\Delta t$ . Denote  $u(i\Delta x, n\Delta t) = u_i^n$ . One might well consider advancing the computed solution by a forward time - centered space differencing

$$u_i^{n+1} = u_i^n - \frac{a\Delta t}{2\Delta x} (u_{i+1}^n - u_{i-1}^n) \quad (64)$$

Now if any error is introduced in going from the continuous equation to the discrete equation, that error should not be amplified as the calculation proceeds. That is, we say the numerical method is stable if

$$\|u^{n+1}\| \leq \|u^n\| \quad (65)$$

where the norm  $\|u^n\| = \Delta x \sum_i |u_i^n|$ . We ask Is there a value of  $\Delta t$  that ensures this scheme is stable? By constructing a difference grid, and noting that we must impose a boundary

condition  $u(a, t) = u_a$  on the left, one can show that this FTCS scheme is unconditionally unstable. [Note: Fourier methods prove the same. See Strikwerda.]

The Lax-Friedrichs scheme is written

$$u_i^{n+1} = \frac{1}{2}(u_{i+1}^n + u_{i-1}^n) - \frac{a\Delta t}{2\Delta x}(u_{i+1}^n - u_{i-1}^n) \quad (66)$$

Claim: The L-F scheme is stable provided the Courant-Friedrichs-Lewy condition is satisfied

$$\left| \frac{a\Delta t}{\Delta x} \right| \leq 1 \quad (67)$$

To see this, note

$$\begin{aligned} \|u^{n+1}\| &= \Delta x \sum_i |u_i^{n+1}| \quad (68) \\ &\leq \frac{\Delta x}{2} \left[ \sum_i \left| \left(1 - \frac{a\Delta t}{\Delta x}\right) u_{i+1}^n \right| + \sum_i \left| \left(1 + \frac{a\Delta t}{\Delta x}\right) u_{i-1}^n \right| \right] \\ &\leq \frac{\Delta x}{2} \left[ \left(1 - \frac{a\Delta t}{\Delta x}\right) \sum_i |u_{i+1}^n| + \left(1 + \frac{a\Delta t}{\Delta x}\right) \sum_i |u_{i-1}^n| \right] \\ &= \frac{1}{2} \left[ \left(1 - \frac{a\Delta t}{\Delta x}\right) \|u^n\| + \left(1 + \frac{a\Delta t}{\Delta x}\right) \|u^n\| \right] \\ &= \|u^n\| \end{aligned}$$

One might also consider taking advantage of the direction of propagation of waves, and write the upwind method

$$u_i^{n+1} = u_i^n - \frac{a\Delta t}{\Delta x}(u_i^n - u_{i-1}^n) \quad (69)$$

A similar analysis shows this method to be stable if the CFL condition holds.

For linear systems, it is helpful to think of upwinding in each characteristic direction - using the difference  $v_i - v_{i-1}$  if the corresponding characteristic speed is positive, and  $v_{i+1} - v_i$  if the speed is negative. Then one transforms back to conserved ( $u$ ) variables. One has

$$\begin{aligned} V_i^{n+1} &= V_i^n - \frac{\Delta t}{\Delta x} \Lambda^+ (V_i^n - V_{i-1}^n) - \frac{\Delta t}{\Delta x} \Lambda^- (V_{i+1}^n - V_i^n) \quad (70) \\ U_i^{n+1} &= U_i^n - \frac{\Delta t}{\Delta x} A^+ (U_i^n - U_{i-1}^n) - \frac{\Delta t}{\Delta x} A^- (U_{i+1}^n - U_i^n) \end{aligned}$$

The notation here is that  $\lambda_j^+ = \max(\lambda_j, 0)$ ,  $\lambda_j^- = \min(\lambda_j, 0)$ , and the matrices  $\Lambda^+ = \text{diag}(\lambda_j^+)$ ,  $\Lambda^- = \text{diag}(\lambda_j^-)$ ,  $\Lambda = \Lambda^+ + \Lambda^-$ . The corresponding  $A$  matrices are obtained by conjugation by  $R$ , the matrix of right eigenvectors:  $A^\pm = R\Lambda^\pm R^{-1}$ .

All these methods are first order accurate. That is, the local truncation error is of order  $O(\Delta t)$  provided  $\Delta t = c \frac{\Delta x}{\max(a)}$  for some constant  $c$ . To achieve second order accuracy, Lax and Wendroff expanded by a Taylor series

$$u(x, t + \Delta t) = u(x, t) + \partial_t u(x, t) \Delta t + \partial_t^2 u(x, t) \frac{\Delta t^2}{2} + \dots$$

Now note  $\partial_t u = -a\partial_x u$ , and  $\partial_t^2 u = \partial_t(-a\partial_x u) = a^2\partial_x^2 u$ . Substituting and replacing differentiation by centered difference, the Lax-Wendroff method is

$$u_i^{n+1} = u_i^n - \frac{a\Delta t}{2\Delta x}(u_{i+1}^n - u_{i-1}^n) + \frac{a^2\Delta t^2}{2\Delta x^2}(u_{i+1}^n - 2u_i^n + u_{i-1}^n) \quad (71)$$

The LW method is stable under the CFL constraint.

## 11 Computing with Classical Methods

The plots show the computed solution to the Riemann Problem for Burgers' equation

$$\partial_t u + \partial_x \frac{u^2}{2} = 0 \quad u(x, 0) = \begin{cases} 1 & x < 0 \\ 0 & x > 0 \end{cases} \quad (72)$$

Two points are apparent from these figures.

- The first order methods, LF and upwind, smear the discontinuity. The LF method is the worse of the two.
- The LW method, although formally of higher order accuracy, introduces oscillations near the discontinuity.

The smearing in the first order methods is due to artificial viscosity. That is, the method, first order accurate for the advection equation, actually approximates a diffusion equation to higher order. This is shown below. The oscillations in the LW method can be thought of as due to a quadratic approximation near a rapid change; alternatively one sees that LW differentiates across the discontinuity, which is certainly not correct physically.

## 12 Modified Equation

A numerical method of specified truncation error for a conservation law can be viewed as approximating a different equation to higher order. Consider the LF method (66) for a linear equation. Expand each term in a Taylor approximation to find

$$\begin{aligned} 0 &= \frac{1}{\Delta t}[(u + \partial_t u \Delta t + \partial_t^2 u \frac{\Delta t^2}{2} + \dots) - (u + \partial_x^2 u \frac{\Delta x^2}{2} + \dots)] \\ &\quad + \frac{1}{2\Delta x} a [2\partial_x u \Delta x + \partial_x^3 u \frac{\Delta x^3}{3} + \dots] \\ &= \partial_t u + a\partial_x u + \frac{1}{2}(\Delta t \partial_t^2 u - \frac{\Delta x^2}{\Delta t} \partial_x^2 u) + O(\Delta x^2) \end{aligned} \quad (73)$$

Of course,  $u$  is a solution of the PDE. Using again that  $\partial_t^2 u = a^2\partial_x^2 u$  and rearranging, we find the LF method solves to second order the modified equation

$$\partial_t u + a\partial_x u = \frac{\Delta x^2}{2\Delta t} (1 - a^2 \frac{\Delta t^2}{\Delta x^2}) \partial_x^2 u \quad (74)$$

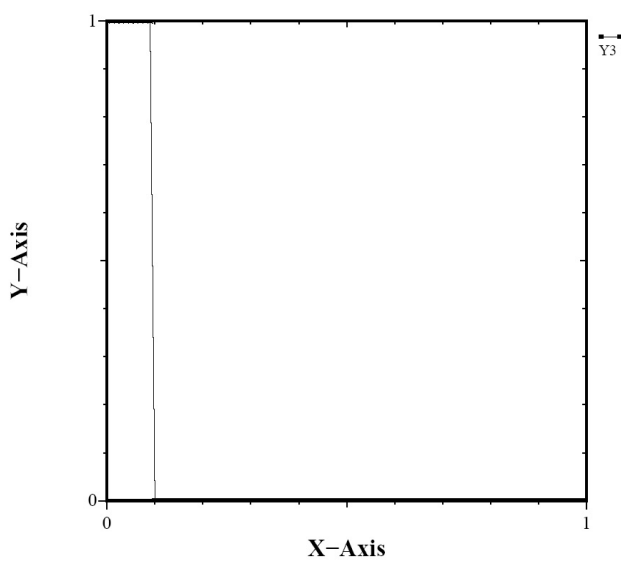
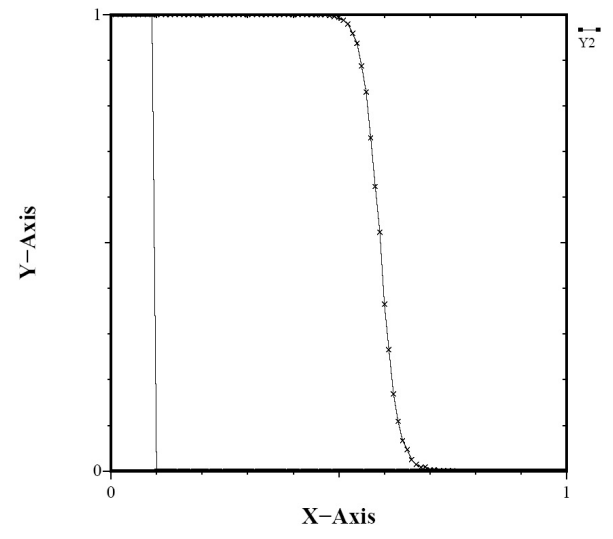
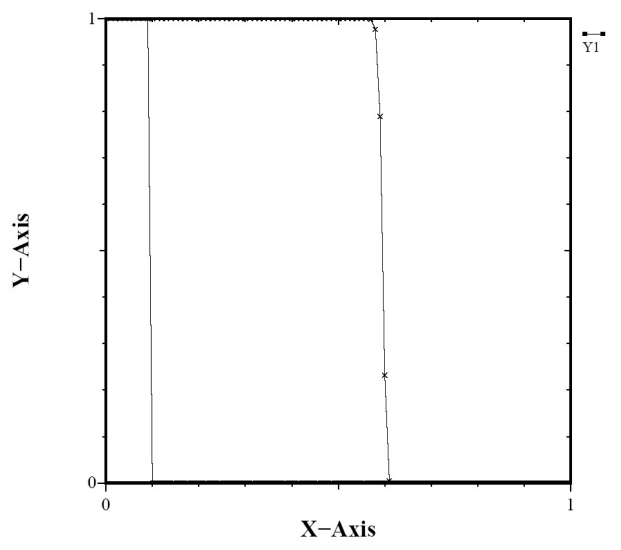


Figure 7: A solution to Burgers' equation, computed with the upwind method, the Lax-Friedrichs method, and a non-conservative difference method. Initially  $u_L = 1$  for  $x < 0.1$ , and  $u_R = 0$  for  $0.1 < x < 1.0$ . It is apparent that the non-conservative scheme does not compute a correct solution - the method yields a steady-state. Also it is clear that the LF method is much more viscous than upwinding.

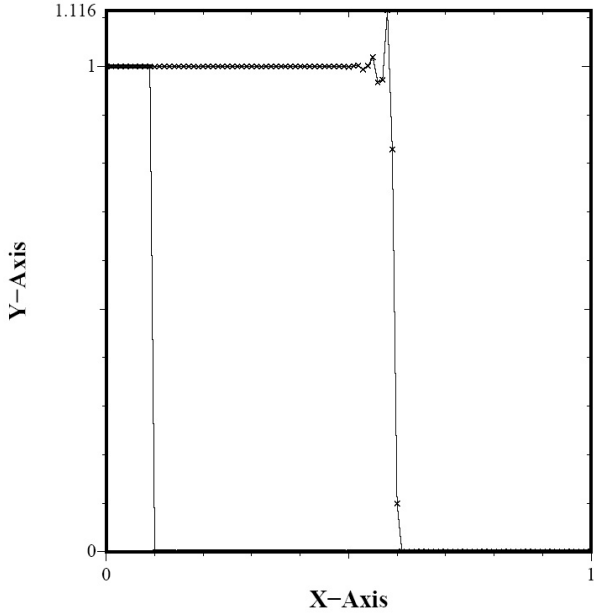


Figure 8: A solution to Burgers' equation computed with the Lax-Wendroff method. The overshoot in LW lags the shock.

Notice this viscosity coefficient vanishes (i) in the limit  $\Delta x \rightarrow 0$  and (ii) if  $\frac{a\Delta t}{\Delta x} = 1$  (the maximum of the CFL constraint).

The upwind method has a modified equation

$$\partial_t u + a\partial_x u = \frac{a\Delta x}{2} \left(1 - a\frac{\Delta t}{\Delta x}\right) \partial_x^2 u \quad (75)$$

In typical computations, one may select  $\Delta t = \frac{3\Delta x}{4a}$ , and the viscosity for upwinding is smaller than for LF.

In contrast, a modified equation analysis for LW shows a dispersive character

$$\partial_t u + a\partial_x u = \frac{a\Delta x^2}{6} \left(a^2 \frac{\Delta t^2}{\Delta x^2} - 1\right) \partial_x^3 u \quad (76)$$

Different frequency waves travel at different speeds. This is the source of the oscillation in the solution plots.

### 13 Conservation and Computation

For Burgers' equation with  $u_L = 1$ ,  $u_R = 0$  as Riemann data, the plots illustrate the need to compute using conservative numerical methods. A non-conservative formulation is

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x} u_i^n (u_i^n - u_{i-1}^n) \quad (77)$$

Notice the shock wouldn't move. But we know that is not the correct solution.

Instead we consider schemes that mimic the conservation law

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x} [F(u_{i-p}^n, \dots, u_{i+q}^n) - F(u_{i-p-1}^n, \dots, u_{i+q-1}^n)] \quad (78)$$

for some numerical flux function  $F$ . The simplest case is  $p = 0$ ,  $q = 1$ , which yields

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x} [F(u_i^n, u_{i+1}^n) - F(u_{i-1}^n, \dots, u_i^n)] \quad (79)$$

This form is natural if we consider the integral form

$$\int_{n\Delta t}^{(n+1)\Delta t} \int_{(i-\frac{1}{2})\Delta x}^{(i+\frac{1}{2})\Delta x} \partial_t u \, dx dt + \int_{n\Delta t}^{(n+1)\Delta t} \int_{(i-\frac{1}{2})\Delta x}^{(i+\frac{1}{2})\Delta x} \partial_x f(u) \, dx dt = 0 \quad (80)$$

Define the cell average of  $u$  as  $\bar{u}_i^n = \int_{(i-\frac{1}{2})\Delta x}^{(i+\frac{1}{2})\Delta x} u(x, n\Delta t) \, dx$  Then we have

$$\bar{u}_i^{n+1} = \bar{u}_i^n - \frac{1}{\Delta x} \left[ \int_{n\Delta t}^{(n+1)\Delta t} f(u(x_{i+\frac{1}{2}}, t)) \, dt - \int_{n\Delta t}^{(n+1)\Delta t} f(u(x_{i-\frac{1}{2}}, t)) \, dt \right] \quad (81)$$

The numerical flux at  $(i + \frac{1}{2})$  is approximated as  $\frac{1}{\Delta t} \int_t^{t+\Delta t} f(u(x_{i+\frac{1}{2}}, t)) \, dt$ . Thus a conservative numerical method is a discrete formulation of the integral form of the conservation laws.

A conservative formulation of upwinding for Burgers equation is

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x} \left[ \frac{1}{2} (u_i^n)^2 - \frac{1}{2} (u_{i-1}^n)^2 \right] \quad (82)$$

We require a flux consistency condition, namely that the numerical flux function  $F$  reduce to the physical flux  $f$  if both arguments are the same - that is,  $F(u^*, u^*) = f(u^*)$ . We require Lipschitz continuity of  $F$  in each argument.

Analytically we have a conservation principle

$$\int_a^b u(x, (n+1)\Delta t) \, dx = \int_a^b u(x, n\Delta t) \, dx - \int_{n\Delta t}^{(n+1)\Delta t} (f(u(b, t)) - f(u(a, t))) \, dt \quad (83)$$

Our discrete formulation has a discrete analogue:

$$\Delta x \sum_{i=J}^K u_i^{n+1} = \Delta x \sum_{i=J}^K u_i^n - \Delta t \sum_{i=J}^K [F(u^n; (i + \frac{1}{2})) - F(u^n; (i - \frac{1}{2}))] \quad (84)$$

This last sum telescopes, and all fluxes drop out except those cells at the edges.

**Lax-Wendroff Theorem** Consider a sequence of computational grids indexed by  $l = 1, 2, \dots$  with mesh parameters  $\Delta x_l, \Delta t_l$  which both  $\rightarrow 0$  as  $l \rightarrow \infty$ . Let  $u_l(x, t)$  denote the computed solution on grid  $l$ , computed with a consistent, conservative method. Assume

$u_l$  converges to some function  $u^*$  as  $l \rightarrow \infty$ . That is,  $\int_0^T \int_a^b |u_l(x, t) - u^*(x, t)| dx dt \rightarrow 0$ . Then  $u^*(x, t)$  is a weak solution of the conservation law.

As a technical point, we need to also assume that the grid functions have bounded total variation for all  $l$ . The total variation is defined as  $TV(v) = \sup \sum_{j=1}^L |v(\xi_j) - v(\xi_{j-1})|$ . This is the analogue of the analytic definition. Also note the theorem does not guarantee the weak solution will satisfy the entropy condition.

**Homework** Find the numerical fluxes for the LF, upwind and LW methods.

## 14 The Godunov Method

To solve (1) with Cauchy data  $u(x, 0) = u_0(x)$ , discretize space into equal length subintervals of size  $\Delta x$ . Assign a discrete approximation to the initial data, where we interpret the discrete values as cell averages:

$$u_i^n = \frac{1}{\Delta x} \int_{(i-\frac{1}{2})\Delta x}^{(i+\frac{1}{2})\Delta x} u(x, n\Delta t) dx$$

At each cell edge, say  $(i + \frac{1}{2})\Delta x$ , there is a Riemann problem to be solved, with data  $u_i^n$  on the left and  $u_{i+1}^n$  on the right. If  $\Delta t$  is small enough, neighboring Riemann problems will not interact. Thus we have the exact solution  $\tilde{u}(x, t)$  for all times  $n\Delta t \leq t \leq (n+1)\Delta t$ . The new, piecewise constant grid solution is the average of  $\tilde{u}(x, (n+1)\Delta t)$  in each cell.

In practice, this algorithm is simplified by observing that it is a version of the integral conservation law.

$$\begin{aligned} & \int_{(i-\frac{1}{2})\Delta x}^{(i+\frac{1}{2})\Delta x} \tilde{u}(x, (n+1)\Delta t) dx = \int_{(i-\frac{1}{2})\Delta x}^{(i+\frac{1}{2})\Delta x} \tilde{u}(x, n\Delta t) dx \\ & - \int_{n\Delta t}^{(n+1)\Delta t} (f(\tilde{u}((i+\frac{1}{2})\Delta x, t)) - f(\tilde{u}((i-\frac{1}{2})\Delta x, t))) dt \end{aligned} \quad (85)$$

Divide through by  $\Delta x$ , this reduces to

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x} [F(u_i^n, u_{i+1}^n) - F(u_{i-1}^n, u_i^n)] \quad (86)$$

where the numerical flux was defined before. This is Godunov's method, a conservative numerical algorithm. Because the analytic solution to the Riemann problem is constant along rays from their vertex, the solution to the cell edge Riemann problems are constant at the stationary state, which is the average demanded in finding  $F$ .

Of course neighboring waves will interact if the timestep is large enough. It can be shown that the Godunov method is stable if

$$\left| \frac{\Delta t}{\Delta x} \lambda_j(u_i^n) \right| \leq 1 \quad (87)$$

for all  $\lambda_j$  at each  $u_i^n$  (the CFL condition). This condition does allow the waves from the Riemann problems at  $(i + \frac{1}{2})$  and  $(i - \frac{1}{2})$  to interact, but the entire interaction occurs within cell  $i$ . By conservation, the cell average at  $(n+1)\Delta t$  remains as prescribed.

The difficult part of Godunov's method is solving each of the cell edge Riemann problems. To do so exactly is a very intensive calculation. Instead we approximate the solution. And it is here that a large number of variations appear. Most of these can be categorized into one of two groups - either solve the nonlinear problem approximately, or solve a related linear problem exactly. We will describe several alternatives. Before doing so we present an interesting approach due originally to Glimm.

## 14.1 Glimm's Method

In the Godunov scheme, the last step is to average the exact Riemann solution over each cell. Glimm proved an existence theorem for solutions to conservation laws by proposing an alternative. Instead of averaging, randomly sample the exact solution at  $(n + 1)\Delta t$  in each cell. He showed that if the sampling is uniform, this random solution converges, with probability 1, to the weak solution of the conservation law. Because each local Riemann problem picks out the entropy solution, this random solution is the entropy satisfying weak solution.

Although we have not demonstrated it yet, the Godunov method introduces artificial viscosity into the computation. Glimm's method has no artificial viscosity.

## 14.2 $2^{nd}$ Order Accuracy

As presented, the Godunov method is akin to a forward Euler step in its time advance, and piecewise linear data for the Riemann problem is first order in space. To achieve second order accuracy, we present an approach which resembles the mid-point rule in time, and a piecewise linear approximation in space. A formal derivation follows, making use of the nonconservative form of the equations for the mid-time prediction followed by a conservative final time update. The Jacobian  $f' = A$ .

By a Taylor expansion,

$$\begin{aligned}
 u_{(i+\frac{1}{2})}^{L(n+\frac{1}{2})} &= u_i^n + (\partial_t u)_i^n \frac{\Delta t}{2} + (\partial_x u)_i^n \frac{\Delta x}{2} \\
 &= u_i^n + (-A \partial_x u)_i^n \frac{\Delta t}{2} + (\partial_x u)_i^n \frac{\Delta x}{2} \\
 &= u_i^n + \frac{1}{2} \Delta u_i^n \left(1 - A \frac{\Delta t}{\Delta x}\right)
 \end{aligned} \tag{88}$$

A similar expression holds at  $(i - \frac{1}{2})$ , providing  $u_{(i-\frac{1}{2})}^{R(n+\frac{1}{2})}$ . In this fashion, at each cell edge there is a predicted value for  $u_{(i+\frac{1}{2})}^{L(n+\frac{1}{2})}$  on the left due to the expansion from cell  $i$ , and one on the right  $u_{(i+\frac{1}{2})}^{R(n+\frac{1}{2})}$ , due to the expansion from cell  $i + 1$ . These provide the cell edge Riemann data, the solution to which is averaged to provide  $u_i^{n+1}$ .

Two modifications to this procedure need to be included. First, the expansion as presented makes sense only if we have converted to characteristic variables. The multiplication by  $A$  is just multiplication by the eigenvalues. So to proceed, one must convert

from conservative to characteristic variables, perform the mid-time predictor, then convert back to conserved variables. The second modification is to observe that  $u_{(i+\frac{1}{2})}^{L(n+\frac{1}{2})}$  should include only those waves from  $i\Delta x$  moving to the right. Likewise,  $u_{(i+\frac{1}{2})}^{L(n+\frac{1}{2})}$  should include only waves in cell  $i+1$  that move to the left. To restrict which waves contribute to the predictor required appropriate multiplication by matrices that project onto the left- and right-going wave families.

It remains to prescribe how to approximate the derivative  $\partial_x u \approx \frac{\Delta u}{\Delta x}$ . There are several approximations that suggest themselves - forward, backward and centered difference approximations. We may be cautious about any approximation, recalling the oscillations in the LW method. Indeed, there is a theorem due to Godunov that says the only linear, monotone preserving methods for conservation laws must be at most first order accurate. This seems to pose an insurmountable obstacle, but the way out is the word ‘linear’. van Leer observed that a nonlinear calculation of  $\frac{\Delta u}{\Delta x}$  is the key. We use the “minmod” prescription. Define

$$\Delta_+ u = u_{i+1} - u_i \quad \Delta_- u = u_i - u_{i-1} \quad \Delta_c u = \frac{1}{2}(u_{i+1} - u_{i-1}) \quad (89)$$

Then

$$\Delta u = \text{minmod}(\Delta_+ u, \Delta_- u, \Delta_c u) = \frac{\text{sgn}(\Delta_+ u) + \text{sgn}(\Delta_- u)}{2} \min(2|\Delta_+ u|, 2|\Delta_- u|, |\Delta_c u|) \quad (90)$$

Notice that, near extrema,  $\Delta u = 0$ .

Figure (1) shows the sharp resolution of a second order method, compared with the first order upwind and LF methods and the oscillations of the LW scheme.

## 15 Approximate Riemann Solvers

Here several approaches are given to solving the local Riemann problems approximately.

**1. MUSCL approach** This algorithm uses the fact that both the shock and rarefaction loci through a state  $u^*$  are tangent to the right eigenvectors  $r_j(u^*)$ . Given left and right states at  $(i + \frac{1}{2})$ , decompose the jump

$$u_R - u_L = \sum_j \alpha_j r_j \quad (91)$$

Then the stationary state is given either as

$$\begin{aligned} u^s &= u_L + \sum_{j, \lambda_j < 0} \alpha_j r_j \\ &= u_R - \sum_{j, \lambda_j > 0} \alpha_j r_j \end{aligned} \quad (92)$$

In practice, it is common to average the states found by the two summations. One then uses as the numerical flux  $F_{(i+\frac{1}{2})} = f(u^s)$ .

What is not apparent is at what state should one evaluate the eigenvectors  $r_j$ ? A common approach, though without justifiable merit, is to use  $(u_L + u_R)/2$ .

This MUSCL approach is reasonable robust. There are two cases where it can fail.

- If any wave, shock or rarefaction, is too strong, the eigenvectors  $r_j$  are not good approximations to the true wave curves (Hugoniot locus or integral curve) and the approximation is in error.
- The approximation lumps the entire strength of a rarefaction wave jump as a single jump across the eigenvector. This does not introduce too much error unless the rarefaction is transonic that is, unless the left edge of the rarefaction travels with negative speed and the right edge with positive speed. In such a case, the stationary state lies within the rarefaction fan. An accurate estimate of this state requires a very accurate approach to solving the entire Riemann problem.

**2. The Roe Solver** Roe proposed solving a linearized version of the Riemann problem exactly, through the use of a special matrix  $\hat{A}(u_L, u_R)$ . Roe required:

1.  $\hat{A}(u_L, u_R)(u_R - u_L) = f(u_R) - f(u_L)$
2.  $\hat{A}(u_L, u_R)$  is diagonalizable with real eigenvalues
3.  $\hat{A}(u_L, u_R) \rightarrow f'(u)$  as  $u_L, u_R \rightarrow u$

Roe matrices have been developed for many systems, but there is no guarantee that it exists (proved to exist for systems with convex entropy) nor that it is easy to find.

Given  $\hat{A}$ , the numerical flux is evaluated as

$$\begin{aligned}
 F(u_L, u_R) &= f(u_L) + \hat{A} \sum_{j, \lambda_j < 0} \alpha_j r_j \\
 &= f(u_R) - \hat{A} \sum_{j, \lambda_j > 0} \alpha_j r_j \\
 &= \frac{f(u_L) + f(u_R)}{2} - \frac{1}{2} |\hat{A}| (u_R - u_L)
 \end{aligned} \tag{93}$$

This last formula is reminiscent of the flux for upwinding in a linear system.

The Roe approach also suffers error when transonic rarefactions occur. There is a standard fix for this problem - essentially split the rarefaction into a left-going and a right-going wave.

**3. The Davis Scheme** The Roe solver contains  $m - 1$  intermediate states. As noted by Harten, Lax and vanLeer, only the stationary state is required. They proposed an approach to use fewer than all  $m - 1$  intermediate states. Davis proposed a specific algorithm that uses only one. Actually, Davis proposed two versions of his algorithm; we present only one of those here, a symmetric version.

Define  $a_{max} = \max_j (|\lambda_j(u_L)|, |\lambda_j(u_R)|)$ . The numerical flux is then

$$F(u_L, u_R) = \frac{f(u_L) + f(u_R)}{2} - \frac{a_{max}}{2}(u_R - u_L) \quad (94)$$

Obviously this is akin to replacing the Roe matrix with an  $a_{max}$  multiple of the identity - thus adding more artificial viscosity than is present in the Roe solver, but at a fraction of the cost. It is useful at this point to recall the numerical flux for the LF scheme. It can be written similarly to (94), but replacing  $a_{max}$  with  $\frac{\Delta x}{\Delta t}$ , which is even more diffusive.

Davis offered a different prescription for computing the mid-time state. Instead of characteristic tracing, he computes a mid-time state  $u_i^{(n+\frac{1}{2})}$  as

$$u_i^{(n+\frac{1}{2})} = u_i^n - \frac{1}{2} \frac{\Delta t}{\Delta x} A_i^n \Delta u \quad (95)$$

The limited slope formula is used to construct  $\Delta u$ . Then the edge states are found by adding the slope computed at time  $n\Delta t$ :

$$\begin{aligned} u_{(i+\frac{1}{2})}^{L(n+\frac{1}{2})} &= u_i^{(n+\frac{1}{2})} + \frac{1}{2} (\Delta u)_i^n \\ u_{(i+\frac{1}{2})}^{R(n+\frac{1}{2})} &= u_{i+1}^{(n+\frac{1}{2})} - \frac{1}{2} (\Delta u)_{i+1}^n \end{aligned} \quad (96)$$

These are the states used in the approximate Riemann solve.

There are several developments generally referred to as ‘‘central schemes’’ which are similar to the Davis approach. These central schemes are essentially  $2^{nd}$  order extensions of the LF scheme, which do not require any Riemann solve (note Davis’ does not require one either). They are generally quite diffusive, and do not usually respect steady state solutions. However they are relatively cheap and easy to apply, so have a certain appeal.

**4. Wave Propagation Algorithm** LeVeque proposed a wave propagation method that uses a flux decomposition similar to Roe’s, but with a flux limiting. Without limiting, the method reduces to a version of the Law-Wendroff method.

LeVeque splits the wave jump

$$\tilde{A}\Delta u = \tilde{A}(u_R - u_L) = \sum_1^m \lambda_j W_j \quad (97)$$

One may think of the fluctuation matrix  $\tilde{A}$  as being decomposed similar to the decomposition for linear systems, so

$$\tilde{A}\Delta u = \tilde{A}^- \Delta u + \tilde{A}^+ \Delta u \quad (98)$$

where  $\tilde{A}^- \Delta u + \tilde{A}^+ \Delta u = f(u_R) - f(u_L)$ . This completely specified the first order scheme. The second order method requires a limiting. Without limiting, it is written

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x} (\tilde{A}^+ \Delta u_i + \tilde{A}^- \Delta u_i) - \frac{\Delta t}{\Delta x} (\tilde{F}_{i+1} - \tilde{F}_i) \quad (99)$$

where

$$\tilde{F}_i = \frac{1}{2} \sum_{j=1}^m |\lambda_j| \left(1 - \frac{\Delta t}{\Delta x} |\lambda_j|\right) W_j \quad (100)$$

Here again  $W_j$  is the  $j^{\text{th}}$  wave. Limiting is done on the waves. If  $W_j = \alpha_j r_j$ , we limit

$$W_j = \tilde{\alpha}_j r_j \quad (101)$$

where  $\tilde{\alpha} = \Phi(\theta)\alpha$ , and at grid point  $i$ ,  $\theta_i = \frac{\alpha_I}{\alpha_i}$  with  $I = i - 1$   $\lambda_i > 0$  or  $I = i + 1$   $\lambda_i \leq 0$ . Finally,  $\Phi(\theta)$  is a minmod function, perhaps as

$$\Phi(\theta) = \max(0, \min(1, \theta)) \quad (102)$$

At <http://www.amath.washington.edu/~rjl/clawpack.html>, LeVeque has made available the CLAWPACK software for hyperbolic systems.

In one form or another, each of these generalizations of the Godunov method can be shown to be total variation decreasing. That is,

$$TV(u^{n+1}) \leq TV(u^n) \quad (103)$$

The reader may consult the web book *Numerical Methods for 1D Compressible Flows*, available at [http://www.crs4.it/HTML/int\\_book/NumericalMethods/int\\_book.html](http://www.crs4.it/HTML/int_book/NumericalMethods/int_book.html).

## 15.1 Other Comments

For multiple dimensions, the MUSCL generalization is due to Colella, and is quite complicated. A simpler approach is to use dimension splitting - get a partial update  $u^*$  as a solution of  $\partial_t u + \partial_x f(u) = 0$ , and a final update  $u^{\text{new}}$  as a solution of  $\partial_t u + \partial_y g(u) = 0$ . Strang splitting is often used to maintain accuracy. Approaches like the Davis scheme can be directly generalized to multiple dimensions without significant difficulty.

It has been shown that this compactness in the TV norm is too strong in multiple dimensions. LeVeque and Goodman showed that an TVD scheme in two dimensions is at most first order accurate.

## 16 ENO Methods

As mentioned, the MUSCL approach and its relatives are TVD. Harten and colleagues studied schemes that allow the variation to increase at any step by an amount  $c\Delta x^q$ , where  $q$  is related to the order of the scheme. Their approach is informed by the work on TVD methods, but the ENO (Essentially Non-Oscillatory) approach does not have the intimate connections to Riemann problems and physics. The ENO methodology begins with cell average values  $u_i^n$ . From these, it produces a high order reconstruction of a smooth interpolating function. This interpolant is time advanced, and the solution projected back onto discrete cell averages. When looked at in this way, the method is a prescription. One chooses the order of the

spatial interpolation, and the order of the time integration to write this prescription. The spatial reconstruction depends on the upwind direction, and builds from a divided difference stencil. The time integration has been examined by Osher and Shu; for second order time stepping, as we will show here, they recommend a trapezoidal rule integration.

We illustrate the method for a scalar conservation law. The numerical flux is built in steps.

1. If  $a_{(i+\frac{1}{2})} = (f(u_R) - f(u_L))/(u_R - u_L) > 0$ ,  $F_{(i+\frac{1}{2})}^* = f(u_i)$ ; else  $F_{(i+\frac{1}{2})}^* = f(u_{i+1})$ .
2. The next piece depends on the sign of  $a_{(i+\frac{1}{2})}$ . If it is positive,  $F_{(i+\frac{1}{2})}^* = F_{(i+\frac{1}{2})}^* + \frac{1}{2} \minmod(f_i - f_{i-1}, f_{i+1} - f_i)$ . If it is negative,  $F_{(i+\frac{1}{2})}^* = F_{(i+\frac{1}{2})}^* - \frac{1}{2} \minmod(f_{i+1} - f_i, f_{i+2} - f_{i+1})$ .
3. The choice at third order depends on the stencil at second order. Again, one grid point is added to the second order stencil, and two second order differences are computed; the one with the smaller variation is used (the minmod). For example, for positive  $a_{(i+\frac{1}{2})}$ , we might have  $F_{(i+\frac{1}{2})}^* = f(u_i) + \frac{1}{2}(f_{i+1} - f_i) - \frac{1}{3} \minmod(f_{i+2} - 2f_{i+1} + f_i, f_{i+1} - 2f_i + f_{i-1})$ . However a different stencil at second order might have produced  $F_{(i+\frac{1}{2})}^* = f(u_i) + \frac{1}{2}(f_i - f_{i-1}) + \frac{2}{3} \minmod(f_{i+1} - 2f_i + f_{i-1}, f_i - 2f_{i-1} + f_{i-2})$ . Similar considerations occur if  $a_{(i+\frac{1}{2})}$  is negative.

Given the formulae for the flux pieces, write  $\partial_t u = -\frac{1}{\Delta x}(F_{(i+\frac{1}{2})}^* - F_{(i-\frac{1}{2})}^*) \equiv \ell(u^n)$ . The time stepping can then be written

$$\begin{aligned} u_i^{(n+\frac{1}{2})} &= u_i^n + \Delta t \ell(u^n) \\ u_i^{n+1} &= \frac{u_i^n + u_i^{(n+\frac{1}{2})}}{2} + \frac{\Delta t}{2} \ell(u^{(n+\frac{1}{2})}) \end{aligned} \tag{104}$$

For systems, one could work in characteristic variables to compute the numerical fluxes  $F$ . Osher and Shu have developed a recipe that allows use of the conserved variables instead.

## 17 Loss of Strict Hyperbolicity

In the sections above, we have used strict hyperbolicity to stitch together local coordinates in state space given by shock and rarefaction loci, into global coordinates. When strict hyperbolicity fails, this global coordinate structure is destroyed, and approximate methods can fail. We illustrate this, and the consequent failures of numerical methods for conservation laws, by an example.

A conservation law such as (1) has rotational symmetry if  $f \circ O = O \circ f$  for all  $O \in O(n)$ . If a system is symmetric,  $f \sim \varphi(|u|^2)u$ . This idea of rotational symmetry is fundamental

in multi-dimensional physics - it encompasses the notion of isotropy. For example, isotropic elasticity and ideal MHD have the form

$$\begin{aligned} \partial_t w - \partial_x v &= 0 \\ \partial_t v - \partial_x g(w) &= 0 \\ \partial_t(e(w) + \frac{|v|^2}{2}) - \partial_x(v \cdot g(w)) &= 0 \end{aligned} \tag{105}$$

where  $g$  is equivariant.

As a model, we consider  $U = (u, v)^T$

$$\partial_t U + \partial_x(|U|^2 U) = 0 \tag{106}$$

Related is the parabolic regularization

$$\partial_t U + \partial_x(|U|^2 U) = \epsilon \partial_x^2 U \tag{107}$$

There are two modes for hyperbolic waves:

- the azimuthal mode  $\lambda_a = |U|^2$  with  $r_a = (-v, u)$ ; this wave is linearly degenerate;
- the radial mode  $\lambda_r = 3|U|^2$  with  $r_r = (u, v)$ ; this wave is genuinely nonlinear away from origin.

Note the system is hyperbolic, but not strictly hyperbolic. Freistühler showed there is an isomorphism between the wave pattern of the model and (part of) the waves of a generic, rotationally invariant system.

We study the Riemann problem with data  $U_L$  and  $U_R$ . In particular, it is helpful to normalize and take  $U_L = (1, 0)^T$ . For general  $U_R$ , the solution consists of an azimuthal wave - a rotation - in the  $(u, v)$ -plane to a state  $U_m$ , where  $|U_m| = |U_L|$ ; this wave is a contact discontinuity. This rotation is followed by a radial wave to  $U_r$ ; this is a shock if  $|U_r| < |U_m|$ , and a rarefaction if  $|U_r| > |U_m|$ .

Consider for a moment the special data  $U_R = (-1, 0)^t$ . The centered wave construction produces a single contact discontinuity, with  $v = 0$ . However there is another solution. If  $v \equiv 0$ , (106) reduces to the scalar  $\partial_t u + \partial_x u^3 = 0$ , an equation we have seen before. We know the solution consists of a composite wave. Thus we are confronted with a new kind of nonuniqueness.

In the figures, we present computed solutions using the Random Choice method, and a second order Godunov method. We also computed using Random Choice but adding viscosity explicitly (the parabolic regularization). The data  $U_L$  is always  $(1, 0)$ . The data  $U_R$  is presented as  $U_R = (\rho \cos(\delta), \rho \sin(\delta))$ . Figure 1 illustrates how viscosity, whether artificial or explicit, corrupts the centered wave construction. In Figure 2, viscous dissipation is shown to have a non-uniform impact, increasing as the Riemann data approaches  $\delta = 1$ .

*Lemma 1* Given initial data  $U_0$ , denote the centered wave construction solution operator by  $S$ . That is, the solution is given by  $U(x, t) = S(U_0)$ . Then  $S$  is continuous in  $L^1_{loc}$ , and stable against small perturbations.

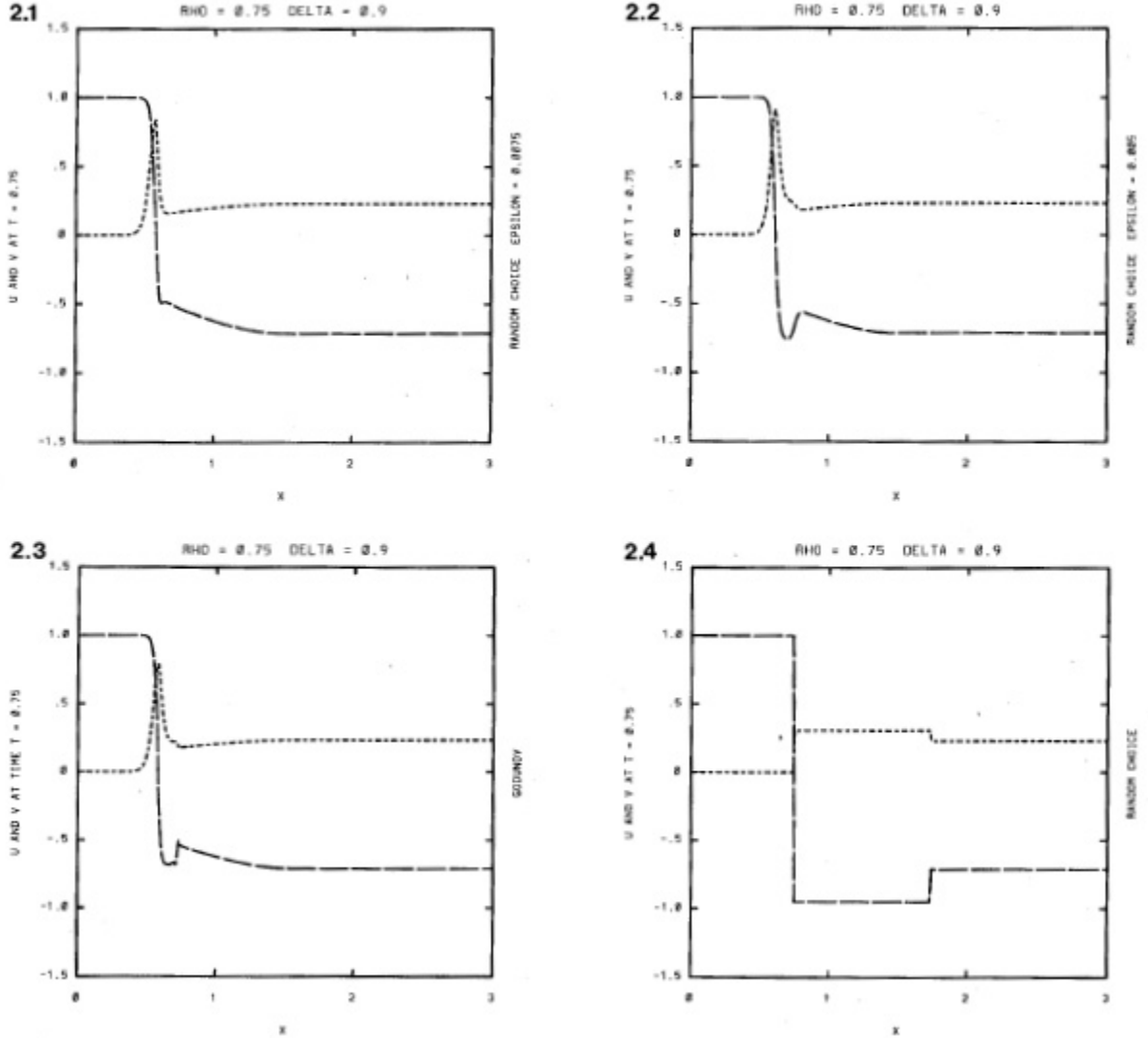


FIG. 2. A comparison of the Random Choice scheme, with various values of the viscosity parameter  $\epsilon$ , and the Godunov scheme. For these plots,  $\rho = 0.75$  and  $\delta = 0.9\pi$ . In Figure 2.1,  $\epsilon = 0.0075$ ; in Figure 2.2,  $\epsilon = 0.005$ . Figure 2.3 shows results using the Godunov scheme. Figure 2.4 shows results from the inviscid Random Choice scheme, i.e.,  $\epsilon = 0.0$ .

Figure 9: Results from computations for  $U_R : \rho = 0.75, \delta = 0.9$ . The Godunov scheme and Random Choice schemes are shown, as are computations with the Random Choice method together with a viscous regularization. Clearly schemes with the viscosity levels presented miss the two waves of the centered wave construction.

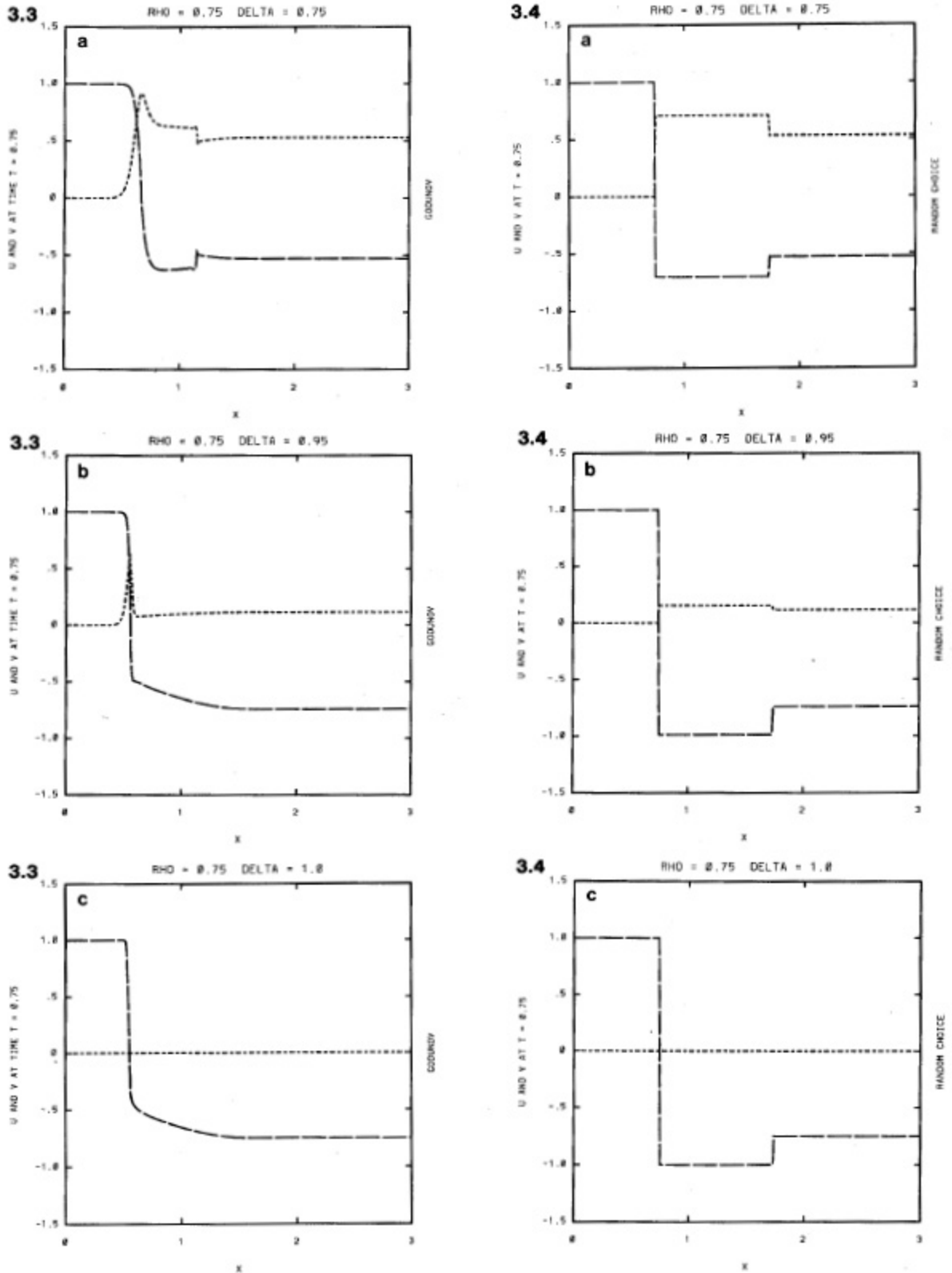


FIG. 3—Continued

Figure 10: Results from computations for  $U_R$  :  $\rho = 0.75$  and several angles  $\delta$  approaching  $\delta = 1$ . The Godunov scheme and Random Choice schemes are shown. As the angle increase, the Godunov method, which contains some viscosity, fails to capture the centered waves.

*Lemma 2* Given colinear initial data  $U_0$  (i.e.  $v_L = v_R = 0$ ), denote the viscous limit construction solution operator by  $s$ . That is, the solution  $U(x, t) = s(U_0)$ . Then  $s$  is continuous in  $L^1_{loc}$ , stable against small perturbations.

*Lemma 3* If the vanishing viscosity method converges to a piecewise continuous wave pattern, then this limit equals  $S(U_0)$  provided  $0 \notin (U_L, U_R)$ . If, however,  $0 \in (U_L, U_R)$ , then the solution of the viscous problem converges to  $s(U_0)$ .

Thus (1) and (2) are NOT uniformly ‘close’.

As a particular application of this difficulty situation, the so-called intermediate shocks in MHD have a structure like the model. These shocks are stable to perturbations as viscous waves, but split into 2 waves as the viscosity goes to zero.

What to do? Well, one idea is to augment the system to enforce aspects of the centered wave symmetry. To this end, write  $U = r\theta$ ,  $\theta \in S^1$ . From (106), one can derive the equations

$$\begin{aligned}\partial_t r + \partial_x r^3 &= 0 \\ \partial_t \theta + r^2 \partial_x \theta &= 0\end{aligned}\tag{108}$$

This is a nonlinear scalar for  $r$ , and, given  $r$ , a linear transport equation for  $\theta$ . Now returning to our model system and motivated by this  $r, \theta$  construction, we propose a new and improved system

$$\begin{aligned}\partial_t r + \partial_x r^3 &= 0 \\ \partial_t U + \partial_x (r^2 U) &= 0\end{aligned}\tag{109}$$

The eigenvalues of this augmented system are  $3r^2$ , with eigenvector  $(r, 2u, 2v)^t$ ,  $r^2$  with  $(0, u, v)^t$ , and  $r^2$  with  $(0, -v, u)^t$ .

Let  $r_0 = |U_0|$ . Then there is a solution  $U$  to the augmented equations with  $U = S(U_0)$  where  $S$  is centered wave solution operator. Thus the new system ‘tracks a constraint of the system’.

Now turn to a parabolic regularization

$$\begin{aligned}\partial_t r + \partial_x r^3 &= \epsilon \partial_x^2 r \\ \partial_t U + \partial_x (r^2 U) &= \epsilon \partial_x^2 U\end{aligned}\tag{110}$$

*Lemma 4* For all  $\epsilon > 0$  and all  $U_0 \in L^\infty$  and  $r_0 \in L^\infty$  satisfying  $|U_0| \leq r_0$ , (110) has bounded smooth solutions  $(r_\epsilon, U_\epsilon)$  with data  $(r_0, U_0)$ . The pair  $(r_\epsilon, U_\epsilon)$  converge, as  $\epsilon \rightarrow 0$ , in  $L^\infty$ -weak-\* to a unique limit  $(r, U) \in L^\infty$ , which solves (106). If additionally  $r_0 = |U_0|$ , then this limit yields  $U = S(U_0)$ .

In spite of this finding, we are left with the questions: What are the appropriate invariants for MHD (and for elasticity)? and Can we define augmented systems and solve these numerically, with well-understood finite difference or finite element methods?

## 18 Current Frontiers

There are three topics at the forefront of current research in hyperbolic conservation laws. One is motivated by physics - How does one solve systems of equations which vary over

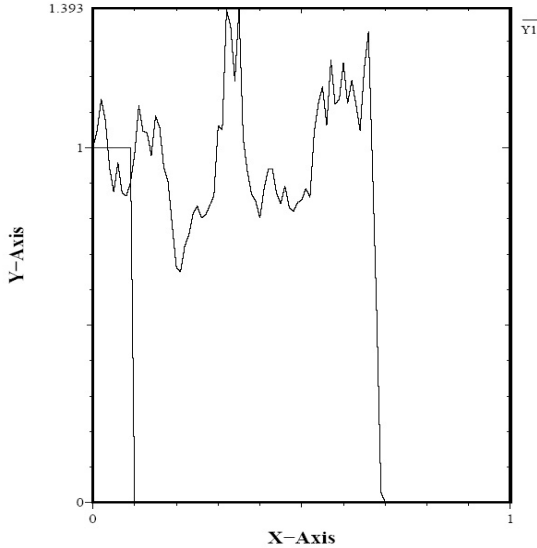


Figure 11: A calculation of a random Burgers' equation, computed with a first order upwind method.

several length and/or time scales? For example, in reactive flow problems, the chemistry of the system often varies over 2 or 3 timescales. In addition, flow typically occurs on scales of tens of centimeters to meters, while, to get the reactions accurately requires resolving on the length of microns. For fluid dynamics, there is a reasonably well understood continuum of equations from molecular dynamics or Monte Carlo to the Boltzman equations to Euler and Navier-Stokes equation. There are proposals which solve the MD equations in a limited region, say near the wall of a vessel where the chemical reactions may take place, and which solve the compressible Navier-Stokes equations away from that wall. Between the two regions is a “handshake” region where both systems are imposed, to generate correct boundary conditions for the two flows.

The second topic asked how stochastic variables affect flow. For example, if the shallow water equations were to include topography, the gravitational acceleration would vary from point to point. If the digital elevation model contained errors, this  $g$  would contain a stochastic component. How is the flow impacted by such variability? The figure shows the variation in a calculation of Burgers' equation  $\partial_t u + \partial_x (\frac{1+\eta}{2} u^2) = 0$ , computed with a first order upwind method. Here  $\eta$  is a uniformly distributed random number in the interval  $[-0.25, 0.25]$ . One sees the significant impact the stochastic fluctuations can have on the shock solution - the mean might be correct, but variations can matter. (The average weather forecast might be partly cloudy with a chance of a shower, but I wouldn't want to be outside when the inevitable heavy rain comes!)

The third topic is more computational. We have presented here a fixed grid, Eulerian framework for the numerical computations of conservation laws. Certainly there is a Lagrangian approach; indeed, the original MUSCL scheme was presented first in a Lagrangian framework. But more generally one may be interested in methods which are Lagrangian in the sense that the grid or elements more or less move with the flow. There are emerging so-called “quasi-particle” methods. Here the basic computational element is a computational

particle, but these have no relation to physical particles. Inside each quasi-particle standard conservation properties hold. Quasi-particles interact with each other through kernel functions. They are advected with the flow.

## References

- [1] Colella, P. and Woodward, P. (1984) *J. Comp. Phys* **54** 174.
- [2] Freistühler, H. and Pitman, E.B. (1992) *J. Comp. Phys* **100** 306.
- [3] Glimm, J. (1965) *Comm. Pure Appl. Math* **18** 95.
- [4] Lax, P.D. (1973) *Hyperbolic Systems of Conservation laws and the Mathematical Theory of Shock Waves* CBMS-NSF Series in Applied Mathematics, SIAM
- [5] LeVeque, R (1990) *Numerical Methods for Conservation Laws* Birkhäuser.
- [6] Roe, P.. (1981) *J. Comp. Phys* **43** 357.
- [7] Shu, C.-W. and Osher, S. (1988) *J. Comp. Phys* **77** 439.
- [8] Smoller, J. (1983) *Shock Waves and Reaction Diffusion Equations* Springer.
- [9] Strikwerda, J.C. (1989) *Finite Difference Schemes and partial Differential Equations* Wadsworth & Brooks-Cole.
- [10] van Leer, B. (1979) *J. Comp. Phys* **37** 101.