Chapter 6

Hyperbolic Problems

6.1 Vector Systems and Characteristics

Let's resume the study of finite difference techniques for hyperbolic conservation laws by examining the characteristics of one-dimensional vector systems having the form

$$\mathbf{u}_t + \mathbf{f}(\mathbf{u})_x = \mathbf{b}(x, t, \mathbf{u}). \tag{6.1.1a}$$

Here, $\mathbf{u}(x,t)$, $\mathbf{f}(\mathbf{u})$, and $\mathbf{b}(x,t,\mathbf{u})$ are m-dimensional vectors. As in Chapter 1, it is sometimes convenient to write (6.1.1a) in the convective form

$$\mathbf{u}_t + \mathbf{A}\mathbf{u}_x = \mathbf{b}, \tag{6.1.1b}$$

where the Jacobian

$$\mathbf{A}(\mathbf{u}) = \mathbf{f}_{\mathbf{u}}(\mathbf{u}) \tag{6.1.1c}$$

is an $m \times m$ matrix.

Definition 6.1.1. *If* $\mathbf{A}$ *has* m *real and distinct eigenvalues* $\lambda_1 < \lambda_2 < \ldots < \lambda_m$ *and, hence,* m *linearly independent eigenvectors* $\mathbf{p}^{(1)}$, $\mathbf{p}^{(2)}, \ldots, \mathbf{p}^{(m)}$, *then (6.1.1a) is said to be hyperbolic.*

Physical problems where dissipative effects can be neglected often lead to hyperbolic systems. Areas where these arise include acoustics, dynamic elasticity, electromagnetics, and gas dynamics.

Let

$$\mathbf{P} = [\mathbf{p}^{(1)}, \mathbf{p}^{(2)}, \ldots, \mathbf{p}^{(m)}] \tag{6.1.2a}$$

and write the eigenvalue-eigenvector relation in matrix form as

$$\mathbf{AP} = \mathbf{P\Lambda}, \tag{6.1.2b}$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix} \tag{6.1.2c}$$

Multiplication of (6.1.1b) by $\mathbf{P}^{-1}$ and the use of (6.1.2b) gives

$$\mathbf{P}^{-1}\mathbf{u}_t + \mathbf{P}^{-1}\mathbf{A}\mathbf{u}_x = \mathbf{P}^{-1}\mathbf{u}_t + \mathbf{\Lambda}\mathbf{P}^{-1}\mathbf{u}_x = \mathbf{P}^{-1}\mathbf{b}.$$

Let

$$\mathbf{w} = \mathbf{P}^{-1}\mathbf{u} \tag{6.1.3}$$

so that

$$\mathbf{w}_t + \mathbf{\Lambda}\mathbf{w}_x = \mathbf{P}^{-1}\mathbf{u}_t + (\mathbf{P}^{-1})_t\mathbf{u} + \mathbf{\Lambda}[\mathbf{P}^{-1}\mathbf{u}_x + (\mathbf{P}^{-1})_x\mathbf{u}].$$

Using (6.1.3)

$$\mathbf{w}_t + \mathbf{\Lambda}\mathbf{w}_x = \mathbf{Q}\mathbf{w} + \mathbf{g}, \tag{6.1.4a}$$

where

$$\mathbf{Q} = [(\mathbf{P}^{-1})_t + \mathbf{\Lambda}(\mathbf{P}^{-1})_x]\mathbf{P}, \qquad \mathbf{g} = \mathbf{P}^{-1}\mathbf{b}. \tag{6.1.4b}$$

In component form, (6.1.4a) is

$$(w_i)_t + \lambda_i(w_i)_x = \sum_{j=1}^{m} q_{i,j}w_j + g_i, \qquad i = 1, 2, \ldots, m. \tag{6.1.4c}$$

Thus, the transformation (6.1.3) has uncoupled the differentiated terms of the original system (6.1.1b).

As in Chapter 1, consider the directional derivative of each component w_i, $i = 1, 2, \ldots, m$, of $\mathbf{w}$,

$$\frac{dw_i}{dt} = (w_i)_t + (w_i)_x \frac{dx}{dt}, \qquad i = 1, 2, \ldots, m,$$

in the directions

$$\frac{dx}{dt} = \lambda_i, \qquad i = 1, 2, \ldots, m, \tag{6.1.5}$$

and use (6.1.4c) to obtain

$$\frac{dw_i}{dt} = \sum_{j=1}^{m} q_{i,j} w_j + g_i, \qquad i = 1, 2, \ldots, m. \tag{6.1.6}$$

As a reminder, the curves (6.1.5) are the *characteristics* of the system (6.1.1a, 6.1.1b). The partial differential system (6.1.1b) may be solved by integrating the $2m$ ordinary differential equations (6.1.5, 6.1.6). This system is uncoupled through its differentiated terms but coupled through $\mathbf{Q}$ and $\mathbf{g}$. This method of solution is, quite naturally, called the method of characteristics. While we could develop effective numerical methods based on the method of characteristics, they are generally not efficient when $m > 2$.

In Chapter 1, we studied homogeneous ($\mathbf{b} \equiv \mathbf{0}$) scalar problems ($m = 1$) and found that the solution u did not change along the characteristic. Examining (6.1.6), we see that w_i will change along the characteristic unless the right side of (6.1.6) vanishes. This generally requires both $\mathbf{b} = \mathbf{0}$ and $\mathbf{P}$ to be constant. Similarly, we'll have to modify the definition of domain of dependence that was introduced in Chapter 2 to accommodate the inhomogeneous term in (6.1.1b).

Definition 6.1.2. *The set of all points that determine the solution at a point $P(x_0, t_0)$ is called the domain of dependence of P.*

Consider the arbitrary point $P(x_0, t_0)$ and the characteristics passing through it as shown in Figure 6.1.1. The solution $u(x_0, t_0)$ depends on the initial data on the interval $[A, B]$ and on the values of $\mathbf{b}$ in the region APB, bounded by $[A, B]$ and the characteristic curves $\dot{x} = \lambda_1$ and $\dot{x} = \lambda_m$. Thus, the region APB is the domain of dependence of P.

Example 6.1.1. Consider the initial value problem for the forced wave equation

$$u_{tt} = a^2 u_{xx} + q(x), \qquad -\infty < x < \infty, \qquad t > 0,$$

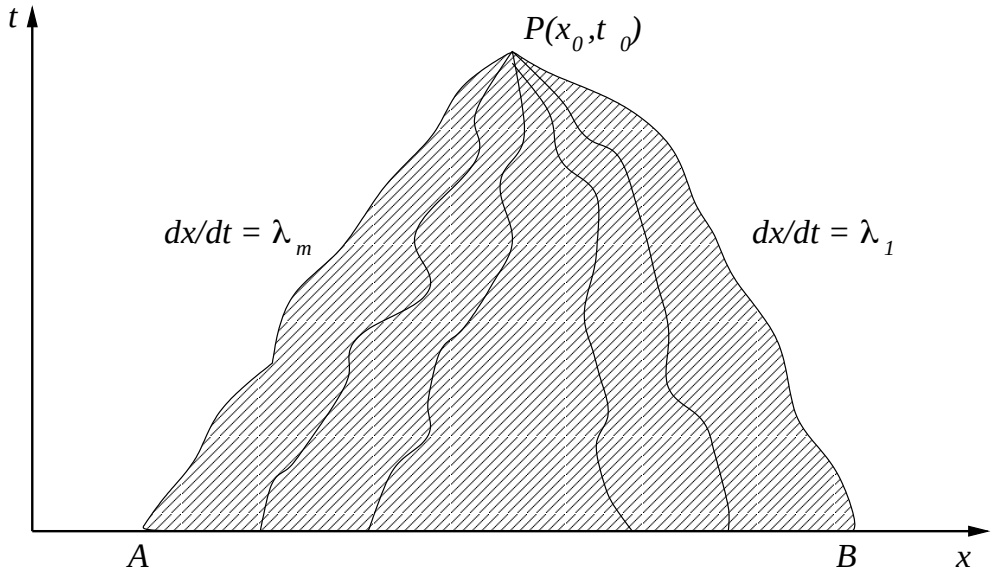

Figure 6.1.1: Domain of dependence of a point $P(x_0, t_0)$. The solution at P depends on the initial data on the line $[A, B]$ and the values of $\mathbf{b}$ within the region APB bounded by the characteristic curves $dx/dt = \lambda_1, \lambda_m$.

$$u(x,0) = u^0(x), \qquad u_t(x,0) = \dot{u}^0(x), \qquad -\infty < x < \infty.$$

If $u(x,t)$ corresponds to the transverse displacement of a taut string, then the constant $a^2 = T/\rho$, where T is the tension and ρ is the linear density of the string, and $q(x)$ is a lateral load applied to the string in the direction of u. The lateral load could, for example, represent the weight of the string.

We'll transform the wave equation to a first-order system by letting

$$u_1 = u_t, \qquad u_2 = au_x.$$

Physically, $u_1(x,t)$ is the velocity and $u_2(x,t)$ is the stress at point x and time t in the string.

The transformation gives

$$(u_1)_t = a(u_2)_x + q$$

and the compatibility condition

$$(u_2)_t = au_{xt} = au_{tx} = a(u_1)_x.$$

Thus, the first-order system has the form of (6.1.1b) with

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix} 0 & -a \\ -a & 0 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} q \\ 0 \end{bmatrix}.$$

The initial conditions become

$$u_1(x,0) = \dot{u}^0(x), \qquad u_2(x,0) = au_x^0(x), \qquad -\infty < x < \infty.$$

The eigenvalues of $\mathbf{A}$ are $\lambda = \pm a$, the characteristics are

$$\dot{x} = \pm a,$$

and the eigenvectors are

$$\mathbf{P} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Since $\mathbf{P}^{-1} = \mathbf{P}$, we may use (6.1.3) to determine the canonical variables as

$$w_1 = \frac{u_1 + u_2}{\sqrt{2}}, \qquad w_2 = \frac{u_1 - u_2}{\sqrt{2}}.$$

From (6.1.4), the canonical form of the problem is

$$(w_1)_t - a(w_1)_x = \frac{q}{\sqrt{2}}, \qquad (w_2)_t + a(w_2)_x = \frac{q}{\sqrt{2}}.$$

The characteristics integrate to

$$x = x_0 - at, \qquad x = x_0 + at,$$

and along the characteristics, we have

$$\frac{dw_k}{dt} = \frac{q}{\sqrt{2}}, \qquad k = 1, 2.$$

Integrating, we find

$$w_1(x,t) = w_1^0(x_0) + \frac{1}{\sqrt{2}} \int_0^t q(x_0 - a\tau) d\tau$$

or

$$w_1(x,t) = w_1^0(x_0) - \frac{1}{a\sqrt{2}} \int_{x_0}^{x_0 - at} q(\xi) d\xi.$$

It's usual to eliminate x_0 by using the characteristic equation to obtain

$$w_1(x,t) = w_1^0(x + at) - \frac{1}{a\sqrt{2}} \int_{x+at}^{x} q(\xi) d\xi.$$

Likewise

$$w_2(x,t) = w_2^0(x - at) + \frac{1}{a\sqrt{2}} \int_{x-at}^{x} q(\xi) d\xi.$$

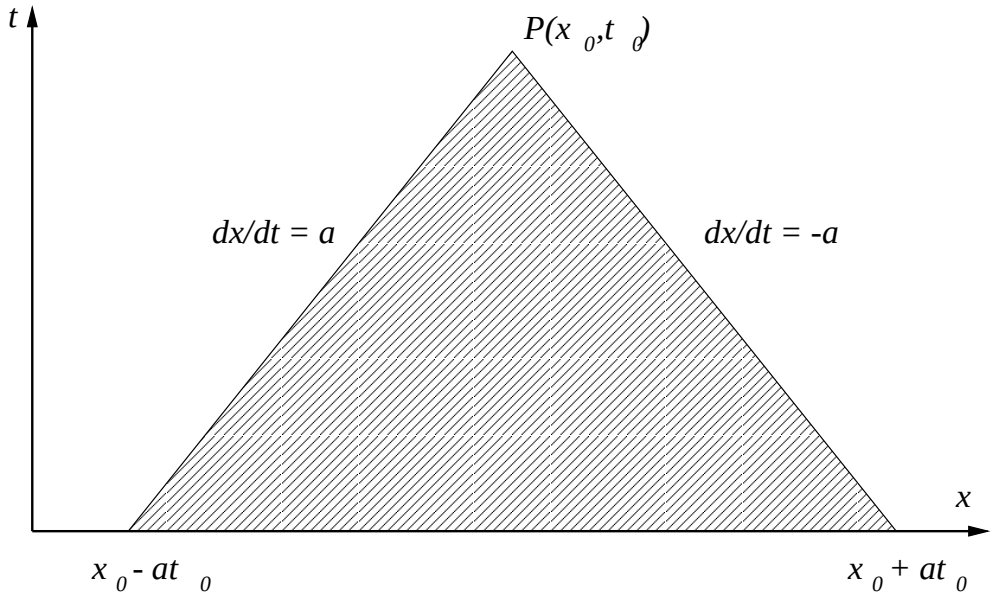

Figure 6.1.2: The domain of dependence of a point $P(x_0, t_0)$ for Example 6.1.1 is the triangle connecting the points P, $(x_0 - at_0, 0)$, and $(x_0 + at_0, 0)$.

The domain of dependence of a point $P(x_0, t_0)$ is shown in Figure 6.1.2. Using the bounding characteristics, it is the triangle connecting the points (x_0, t_0), $(x_0 - at_0, 0)$, and $(x_0 + at_0, 0)$. (Actually, with q being a function of x only, the domain of dependence only involves values of $q(x)$ on the subinterval $(x_0 - at_0, 0)$ to $(x_0 + at_0, 0)$.)

Transforming back to the physical variables

$$u_1(x,t) = \frac{1}{\sqrt{2}}(w_1 + w_2) = \frac{1}{\sqrt{2}}[w_1^0(x+at) + w_2^0(x-at)] + \frac{1}{2a}\int_{x-at}^{x+at} q(\xi)d\xi,$$

$$u_2(x,t) = \frac{1}{\sqrt{2}}(w_1 - w_2) = \frac{1}{\sqrt{2}}[w_1^0(x+at) - w_2^0(x-at)] - \frac{1}{2a}[\int_{x+at}^{x} q(\xi)d\xi + \int_{x-at}^{x} q(\xi)d\xi].$$

Suppose, for simplicity, that $\dot{u}^0(x) = 0$, then

$$u_1(x,0) = 0 = \frac{1}{\sqrt{2}}[w_1^0(x) + w_2^0(x)],$$

$$u_2(x,0) = au_x^0(x) = \frac{1}{\sqrt{2}}[w_1^0(x) - w_2^0(x)].$$

Thus,

$$w_1^0(x) = -w_2^0(x) = \frac{au_x^0(x)}{\sqrt{2}},$$

and

$$u_1(x,t) = \frac{a}{2}[u_x^0(x+at) - u_x^0(x-at)] + \frac{1}{2a}\int_{x-at}^{x+at} q(\xi)d\xi,$$

$$u_2(x,t) = \frac{a}{2}[u_x^0(x+at) + u_x^0(x-at)] - \frac{1}{2a}[\int_{x+at}^{x} q(\xi)d\xi + \int_{x-at}^{x} q(\xi)d\xi].$$

Since $u_2 = au_x$, we can integrate to find the solution in the original variables. In order to simplify the manipulations, let's do this for the homogeneous system ($q(x) = 0$). In this case, we have

$$u_2(x,t) = \frac{a}{2}[u_x^0(x+at) + u_x^0(x-at)];$$

hence,

$$u(x,t) = \frac{1}{2}[u^0(x+at) + u^0(x-at)].$$

The solution for an initial value problem when

$$u^0(x) = \begin{cases} x+1, & \text{if } -1 \le x \le 0 \\ 1-x, & \text{if } 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

is shown in Figure 6.1.3. The initial data splits into two waves having half the initial amplitude and traveling in the positive and negative x directions with speeds a and $-a$, respectively.

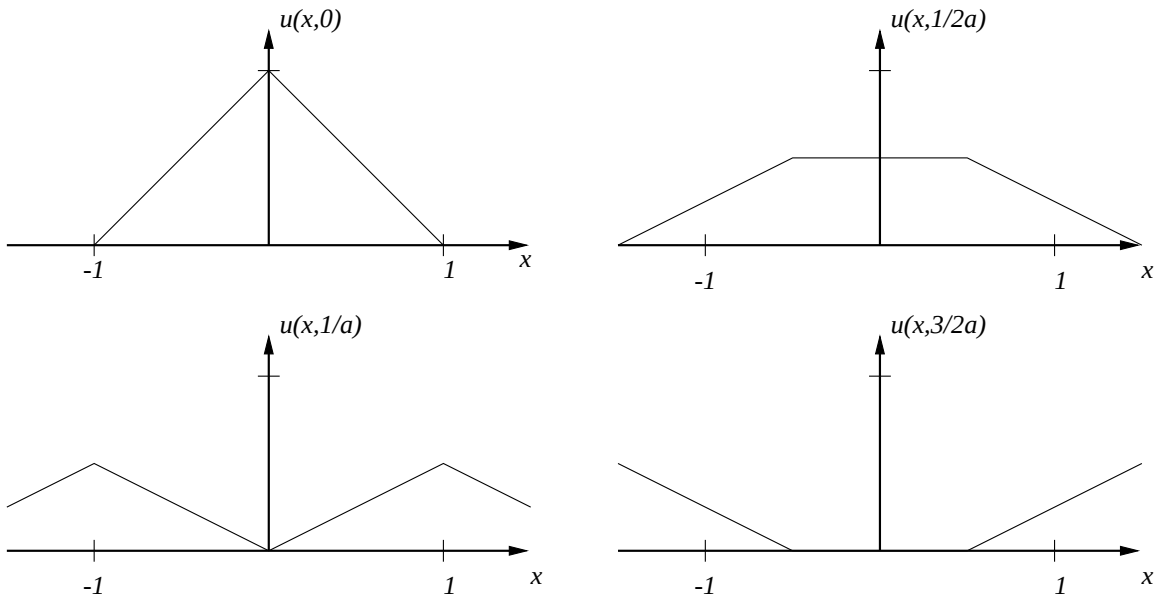

Figure 6.1.3: Solution of Example 6.1.1 at $t = 0$ (upper left), $1/2a$ (upper right), $1/a$ (lower left), and $3/2a$ (lower right).

6.2 The Courant, Friedrichs, Lewy Theorem

All finite difference schemes that we have developed and studied for scalar problems extend to vector systems of the form (6.1.1a) or (6.1.1b). Thus, for example, the Lax-Friedrichs scheme for (6.1.1b) is

$$\frac{\mathbf{U}_j^{n+1} - (\mathbf{U}_{j+1}^n + \mathbf{U}_{j-1}^n)/2}{\Delta t} + \mathbf{A}_j^n \frac{\mathbf{U}_{j+1}^n - \mathbf{U}_{j-1}^n}{2\Delta x} = \mathbf{b}_j^n,$$

where $\mathbf{A}_j^n \equiv \mathbf{A}(j\Delta x, n\Delta t \mathbf{U}_j^n)$, *etc.* The Courant, Friedrichs, Lewy Theorem also applies to vector systems of the form (6.1.1b).

Theorem 6.2.1. *(Courant, Friedrichs, Lewy). A necessary condition for the convergence of a finite difference scheme is that its domain of dependence contain the domain of dependence of the partial differential system (6.1.1b).*

Proof. We'll prove the theorem for constant-coefficient problems. Proofs in more general situations are similar. The matrix $\mathbf{Q}$ of (6.1.4b) is zero for constant-coefficient problems and the canonical form of (6.1.1b) becomes

$$\mathbf{w}_t + \mathbf{\Lambda} \mathbf{w}_x = \mathbf{g}. \tag{6.2.1}$$

The eigenvalues $\lambda_1 < \lambda_2 < \ldots < \lambda_m$ are constant and the characteristics are straight lines. A typical set of characteristics passing through an arbitrary point $P(x_0, t_0)$ is shown in Figure 6.2.1. Again, for simplicity, suppose that a four-point one-level explicit finite difference scheme

$$\mathbf{W}_j^{n+1} = \mathbf{C}_{-1}\mathbf{W}_{j-1}^n + \mathbf{C}_0\mathbf{W}_j^n + \mathbf{C}_1\mathbf{W}_{j+1}^n + \Delta t \mathbf{g}_j^n. \tag{6.2.2}$$

is used to obtain the numerical solution of (6.2.1). The stencil of this difference scheme and its domain of dependence are shown in Figure 6.2.2. The domain of dependence of the finite difference scheme is the region ABP, which is bounded by the x axis and the lines of slope $\pm\Delta t/\Delta x$. The domain of dependence of the partial differential equation is shown as the region CDP in Figures 6.2.1 and 6.2.2. We suppose that the domain of dependence of the partial differential equation does not contain that of the difference

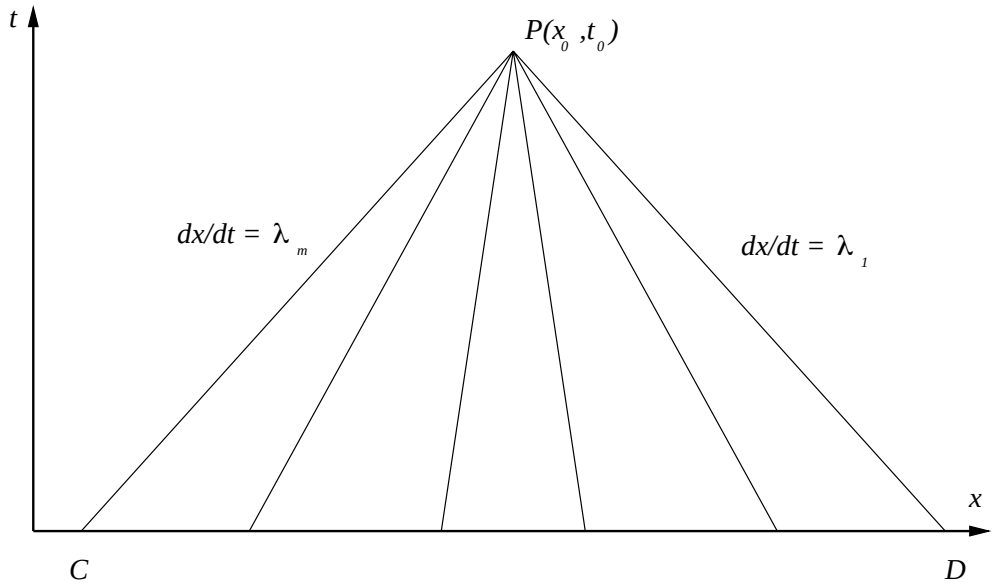

Figure 6.2.1: Characteristics and domain of dependence for a constant coefficient problem of the form (6.1.1b).

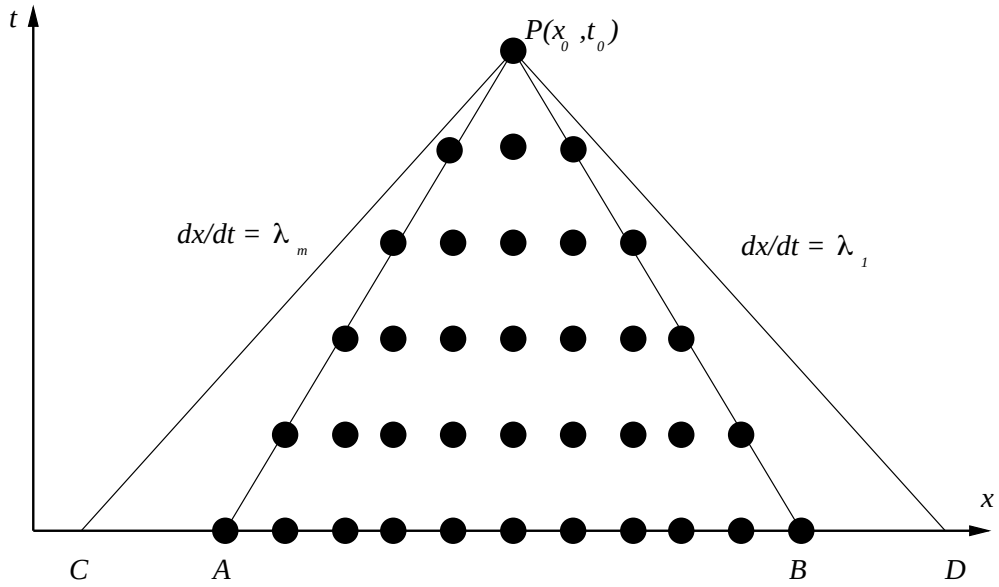

Figure 6.2.2: Sample domains of dependence of constant coefficient partial differential equation of the form (6.1.1b) and a difference scheme of the form of (6.2.2).

scheme, again, as shown in Figure 6.2.2. Consider a problem where the initial data vanishes on AB and the source $\mathbf{g}$ vanishes on ABP. The finite difference solution at P is, therefore, trivial. However, if the initial data is nonzero on CA or BD or the source $\mathbf{g}$ is nonzero on CAP or BDP, the solution of the partial differential system need not

be zero at P. Moreover, the solution of the difference scheme cannot converge to that of the partial differential equation as long as the mesh is refined with $\Delta t/\Delta x$ fixed, since the solution of the difference equation will always be zero at P. We, therefore, have a generalization of the Courant, Friedrichs, Lewy Theorem considered in Chapter 2. □

Remark 1. The Courant, Friedrichs, Lewy Theorem restricts all explicit finite difference schemes to satisfy

$$\max_{1\le i\le m} |\lambda_i| \frac{\Delta t}{\Delta x} \le 1. \tag{6.2.3}$$

Remark 2. The Courant, Friedrichs, Lewy Theorem is usually stated as a stability restriction rather than a convergence restriction. Our proof would have to be supplemented by the Lax equivalence theorem in order to demonstrate that it also imposes a stability restriction.

The von Neumann method can be used to study the stability of homogeneous ($\mathbf{b} = \mathbf{0}$) constant coefficient systems with periodic initial data. Thus, consider

$$\mathbf{u}_t + \mathbf{A}\mathbf{u}_x = \mathbf{0}, \qquad -\infty < x < \infty, \qquad t > 0, \tag{6.2.4a}$$

$$\mathbf{u}(x,0) = \boldsymbol{\phi}(x), \qquad -\infty < x < \infty, \tag{6.2.4b}$$

where $\mathbf{A}$ is constant and $\boldsymbol{\phi}(x)$ is periodic in x. As an example, consider a one-level finite difference scheme having the form

$$\sum_{|s|\le S} \mathbf{D}_s \mathbf{U}_{j+s}^{n+1} = \sum_{|s|\le S} \mathbf{C}_s \mathbf{U}_{j+s}^{n}, \tag{6.2.5}$$

where $\mathbf{C}_s$ and $\mathbf{D}_s$ are $m \times m$ matrices and S is an integer.

As in the scalar case, assume a solution in the form of a discrete Fourier series

$$\mathbf{U}_j^n = \sum_{k=0}^{J-1} \mathbf{A}_k^n e^{2\pi i j k/J}. \tag{6.2.6}$$

Substituting (6.2.6) into (6.2.5)

$$\sum_{|s|\le S} \sum_{k=0}^{J-1} [\mathbf{D}_s \mathbf{A}_k^{n+1} e^{2\pi i (j+s)k/J} - \mathbf{C}_s \mathbf{A}_k^n e^{2\pi i (j+s)k/J}] = \mathbf{0}.$$

Rearranging the sums

$$\sum_{k=0}^{J-1} \sum_{|s|\leq S} [\mathbf{D}_s \mathbf{A}_k^{n+1} e^{2\pi iks/J} - \mathbf{C}_s e^{2\pi iks/J} \mathbf{A}_k^n] e^{2\pi ijk/J} = \mathbf{0}.$$

Using the orthogonality condition (3.2.2), we find

$$\sum_{|s|\leq S} \mathbf{D}_s e^{2\pi iks/J} \mathbf{A}_k^{n+1} = \sum_{|s|\leq S} \mathbf{C}_s e^{2\pi iks/J} \mathbf{A}_k^n. \tag{6.2.7}$$

Solving for $\mathbf{A}_k^n$

$$\mathbf{A}_k^n = (\mathbf{M}_k)^n \mathbf{A}_k^0, \tag{6.2.8a}$$

where

$$\mathbf{M}_k = \left(\sum_{|s|\leq S} \mathbf{D}_s e^{2\pi iks/J} \right)^{-1} \left(\sum_{|s|\leq S} \mathbf{C}_s e^{2\pi iks/J} \right). \tag{6.2.8b}$$

The amplification factor $\mathbf{M}_k$ is an $m \times m$ matrix. In order to demonstrate the stability of the difference scheme (6.2.5), we must find a constant C such that

$$\|(\mathbf{M}_k)^n\| \leq C, \qquad \Delta t \to 0, \qquad n\Delta t \leq T. \tag{6.2.9}$$

The von Neumann necessary condition for stability implies that there exists a constant c such that

$$\rho(\mathbf{M}_k) \equiv \max_{1\leq i\leq m} |\lambda_i(\mathbf{M}_k)| \leq 1 + c\Delta t, \qquad \Delta t \to 0. \tag{6.2.10a}$$

For absolute stability, we require

$$\rho(\mathbf{M}_k) \leq 1. \tag{6.2.10b}$$

Conditions (6.2.9) and (6.2.10) may be quite difficult to verify in practice. We may try to bound the eigenvalues, but such bounds must be accurate in order to draw conclusions about stability.

6.3 Dissipation and Dispersion

As a prelude to understanding those difference schemes that will be effective for solving nonlinear problems with discontinuous solutions, let us return to the linear scalar initial value problem

$$u_t + au_x = 0, \qquad t > 0, \qquad u(x,0) = e^{2\pi ikx}, \qquad -\infty < x < \infty, \tag{6.3.1}$$

Many of the results developed in this section are qualitatively the same for nonlinear problems and vector systems, but, naturally, more difficult to analyze.

Consider an explicit finite difference approximation of (6.3.1) having the form

$$U_j^{n+1} = U_j^n - \frac{\alpha}{2}(U_{j+1}^n - U_{j-1}^n) + \frac{\beta}{2}(U_{j+1}^n - 2U_j^n + U_{j-1}^n), \tag{6.3.2}$$

where

$$\alpha = a\frac{\Delta t}{\Delta x} \tag{6.3.3}$$

is the Courant number. Several popular difference schemes have this form for different values of β and some are listed in Table 6.3. The centered scheme is not absolutely stable for any choice of α, whereas the other schemes are absolutely stable for some range of α (Chapter 3) and we seek to understand the reasons why this is so. For completeness, we'll also include the leap frog scheme

$$U_j^{n+1} = U_j^{n-1} - \alpha(U_{j+1}^n - U_{j-1}^n) \tag{6.3.4}$$

in our study.

Method	β	σ
Centered	0	0
Lax-Wendroff	α^2	$a^2\Delta t/2$
Upwind	$\lvert\alpha\rvert$	$\lvert a\rvert\Delta x/2$
Lax-Friedrichs	1	$\Delta x^2/2\Delta t$

Table 6.3.1: Values of the parameters β and σ for several difference schemes having the form of (6.3.2).

The scheme (6.3.2) may also be regarded as an approximation of the convection-diffusion problem

$$u_t + au_x = \sigma u_{xx}, \qquad t > 0, \qquad u(x,0) = e^{2\pi ikx}, \qquad -\infty < x < \infty, \tag{6.3.5a}$$

with forward time and centered space differencing. The exact solution of this problem is

$$u(x,t) = e^{-4\pi^2\sigma k^2 t} e^{2\pi ik(x-at)}. \tag{6.3.5b}$$

We have already seen that:

- The term σu_{xx} in (6.3.5a) is dissipative or diffusive. It decreases the $\mathcal{L}^2$ energy of the system (Section 1.2). When $\sigma = 0$ the $\mathcal{L}^2$ energy of the system is conserved.
- The dissipation is frequency dependent with the higher frequency terms having the greatest decrease in amplitude.
- When $4\pi^2\sigma k^2 t \ll 1$, the solution of (6.3.5a) is a good approximation of the solution of (6.3.1).

We have also seen that instabilities in finite difference schemes result in a rapid growth in the amplitudes of the high-frequency modes of the discrete solution. This suggests the addition of a small diffusion term (*e.g.*, a discrete approximation of σu_{xx}) to a finite difference scheme to kill the growth of the high-frequency oscillations without greatly affecting accuracy of the solution.

The methods of Table 6.3 are meant to approximate (6.3.1); however, with the exception of the centered scheme, they can also be regarded as discrete approximations of (6.3.5a) with various diffusivities. For example:

- The Lax-Wendroff scheme is an approximation of (6.3.5a) with forward time differences, centered space differences and $\sigma = \beta\Delta x^2/2\Delta t = a^2\Delta t/2$. Since the diffusivity $\sigma = O(\Delta t)$, the Lax-Wendroff scheme is consistent with (6.3.1) as $\Delta t \to 0$. The diffusivity is *artificial* since it is proportional to a numerical parameter (Δt).
- The upwind scheme is an approximation of (6.3.5a) using forward time and centered space differences with $\sigma = \beta\Delta x^2/2\Delta t = |a|\Delta x/2$.

- Similarly, the Lax-Friedrichs scheme approximates (6.3.5a) with the same differencing and $\sigma = \beta\Delta x^2/2\Delta t = \Delta x^2/2\Delta t$.

The solution of (6.3.2) for the special initial conditions (6.3.1) is (*cf.* Problem 3.2.3)

$$U_j^n = (M_k)^n e^{2\pi i k j/J}, \tag{6.3.6a}$$

where

$$M_k = 1 - 2\beta \sin^2\theta - i\alpha \sin 2\theta, \tag{6.3.6b}$$

and

$$\theta = k\pi/J. \tag{6.3.6c}$$

The magnitude of the amplification factor is

$$|M_k|^2 = 1 + (\alpha^2 - \beta^2)\sin^2 2\theta - 4\beta(1-\beta)\sin^2\theta. \tag{6.3.6d}$$

The magnitude of the amplification factor when $\alpha = 0.75$ is shown as a function of θ in Figure 6.3.1 for each of the schemes given in Table 6.3. The centered scheme amplifies the initial data for all values of θ while the upwind, Lax-Friedrichs, and Lax-Wendroff schemes dissipate the initial amplitude for virtually all frequencies.

Definition 6.3.1. *A finite difference scheme is dissipative of order 2r if there exists a constant $c > 0$, independent of Δt and Δx, such that*

$$|M_k(\theta)| \le 1 - c|\theta|^{2r}, \qquad 0 \le \theta \le \pi/2. \tag{6.3.7a}$$

There are other equivalent definitions. First, observe that it suffices to find a constant $\hat{c} > 0$ satisfying

$$|M_k(\theta)|^2 \le 1 - \hat{c}|\theta|^{2r}, \qquad 0 \le \theta \le \pi/2. \tag{6.3.7b}$$

Strikwerda [21], Section 5.1, observes that many common schemes have an amplification factor satisfying

$$|M_k(\theta)|^2 \le 1 - \bar{c}\sin^{2r}\theta, \qquad 0 \le \theta \le \pi/2, \tag{6.3.7c}$$

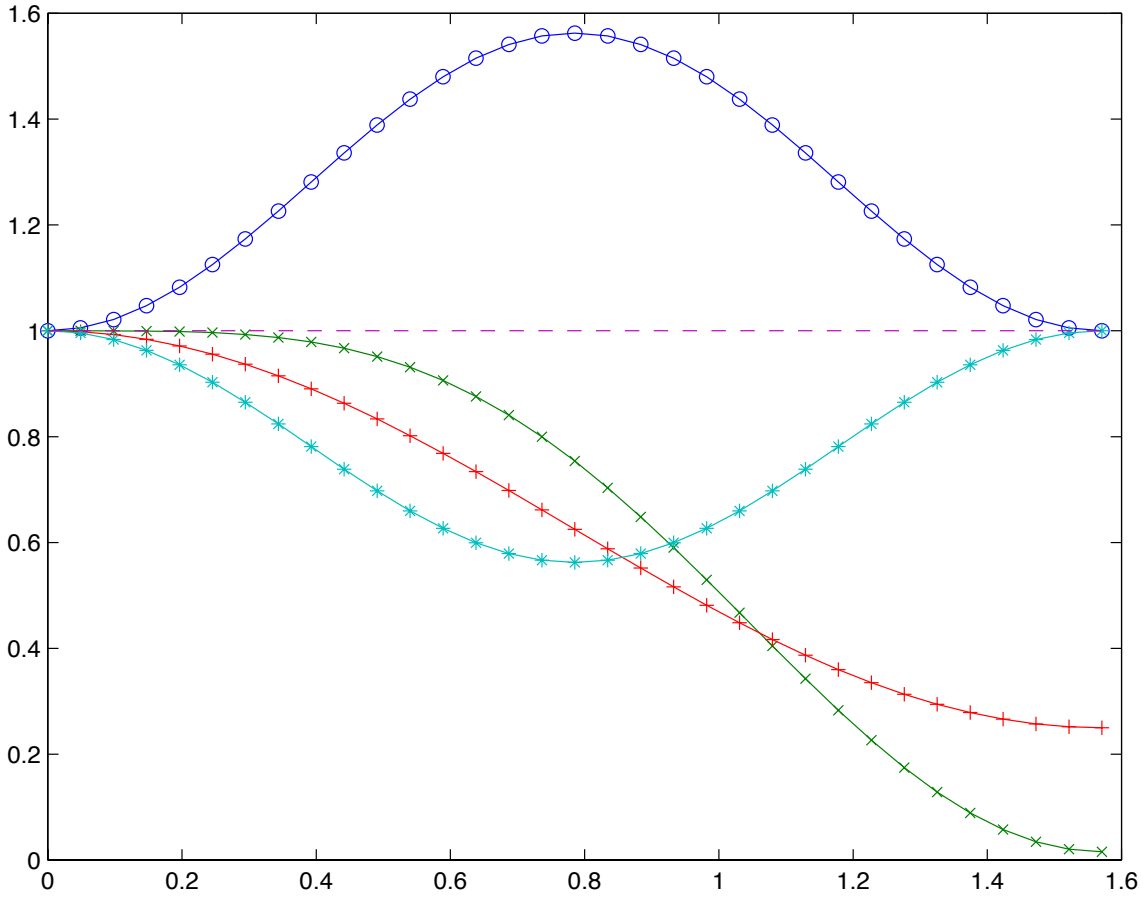

Figure 6.3.1: Magnitudes of the amplification factors when $\alpha = 0.75$ as functions of $\theta = k\pi/J$ for the centered (o), Lax-Wendroff (x), upwind (+), Lax-Friedrichs (*), and leap frog ($--$) schemes.

with $\bar{c} > 0$. Since $\theta \geq \sin\theta$, $0 \leq \theta \leq \pi/2$, (6.3.7a) and (6.3.7c) are equivalent.

Example 6.3.1. Using Definition 6.3.1 and (6.3.6d), let's examine the order of dissipation of the Lax-Wendroff, upwind, Lax-Friedrichs, and centered schemes. Beginning with the Lax-Wendroff scheme, we set $\beta = \alpha^2$ in (6.3.6d) to obtain

$$|M_k|^2 = 1 + \alpha^2(1-\alpha^2)\sin^2 2\theta - 4\alpha^2(1-\alpha^2)\sin^2\theta.$$

Using the double angle formula

$$|M_k|^2 = 1 - 4\alpha^2(1-\alpha^2)\sin^4\theta.$$

Thus, the Lax-Wendroff scheme is dissipative of order four.

For upwind differencing, we set $\beta = |\alpha|$ in (6.3.6d) and find

$$|M_k|^2 = 1 - 4|\alpha|(1 - |\alpha|)\sin^2\theta.$$

Thus, upwind differencing is dissipative of order two. The Lax-Wendroff scheme with fourth-order dissipation gives a more accurate representation of the lower-frequency modes than upwind differencing and greater dissipation of the high-frequency modes.

For the Lax-Friedrichs scheme, we set $\beta = 1$ in (6.3.6d) and obtain

$$|M_k|^2 = 1 - (1 - \alpha^2)\sin^2 2\theta.$$

For small θ, we can find a constant c ($\approx 4(1-\alpha^2)$) so that (6.3.7a) is satisfied. However, this bound will not hold when $\theta \approx \pi/2$. Thus, the Lax-Friedrichs scheme is not strictly dissipative. It will reduce the magnitude of the amplification factor for most frequencies, but not those of the highest-frequency modes (Figure 6.3.1).

Finally, the centered space scheme has $|M_k(\theta)| > 1$, $0 \le \theta \le \pi/2$, and is not dissipative.

Using the von Neumann method on the leap frog scheme (6.3.4), we find the two amplification factors as (*cf.* Problem 1 at the end of this Section)

$$M_k^{\pm} = i\alpha \sin 2\theta \pm \sqrt{1 - \alpha^2 \sin^2 2\theta}. \tag{6.3.8a}$$

There are two amplification factors since the leap frog method is a two-level scheme and, as described in Section 4.3, the discrete solution involves a linear combination of the two. Assuming that $|\alpha| \le 1$, we see that both factors have unit magnitude, *i.e.*,

$$|M_k^{\pm}|^2 = 1. \tag{6.3.8b}$$

Thus, the leap frog scheme neither dissipates nor amplifies the magnitudes of the initial data for any frequency (Figure 6.3.1).

With the prescribed initial data, the exact solution of (6.3.1) has unit magnitude. Thus, the leap frog scheme has no amplitude error. The question, of course, is if the leap frog scheme has no amplitude error, what error does it have? In order to answer this query, write the amplification factor (either (6.3.6b) or (6.3.8a) in the polar form

$$M_k = |M_k| e^{-i\phi_k} \tag{6.3.9a}$$

where

$$\tan\phi_k = \frac{\alpha \sin 2\theta}{1 - 2\beta \sin^2\theta} \tag{6.3.9b}$$

for (6.3.6b) and

$$\tan\phi_k^{\pm} = \pm\frac{\alpha \sin 2\theta}{\sqrt{1 - \alpha^2 \sin^2 2\theta}} \tag{6.3.9c}$$

for (6.3.8a). Using (6.3.6a), the solution of the difference equation (6.3.2) may be written in the form

$$U_j^n = |M_k|^n e^{-in\phi_k} e^{2\pi ikj/J} = |M_k|^n e^{2\pi ik(x_j - \gamma_k t_n)} \tag{6.3.10a}$$

where

$$x_j = j\Delta x = \frac{j}{J}, \qquad t_n = n\Delta t, \qquad \gamma_k = \frac{\phi_k}{2\pi k \Delta t}. \tag{6.3.10b}$$

We will also regard this as being one component of the solution of the leap frog scheme (6.3.4) when (6.3.9c) is used for ϕ.

The *phase speed* of a wave of the form

$$e^{i(\kappa x - \omega t)}$$

is the quantity ω/κ. It is the speed at which waves of frequency ω propagate. If the phase speed depends on the wave number κ then the wave is said to be *dispersive*, otherwise it is non-dispersive. The exact solution of the model problem (6.3.1) is

$$u(x,t) = e^{2\pi ik(x-at)}.$$

Hence, $\omega = 2\pi ka$ and $\kappa = 2\pi k$. The phase speed is a and the wave is non-dispersive. However, the phase speed γ_k of either finite difference scheme (6.3.2) or (6.3.4) depends on k and, hence, these waves are dispersive. Normalized phase speeds (γ_k/a) are shown as functions of θ for the schemes of Table 6.3 and the leap frog scheme (6.3.4). Jumps in the phase speed correspond to branch cuts in the arctangent function. While phase speeds appear reasonable when θ is small, all have large errors when θ is near $\pi/2$. With $\theta = \pi/2$, for example, both (6.3.9b) and (6.3.9c) give $\tan\phi_k = 0$ and, consequently,

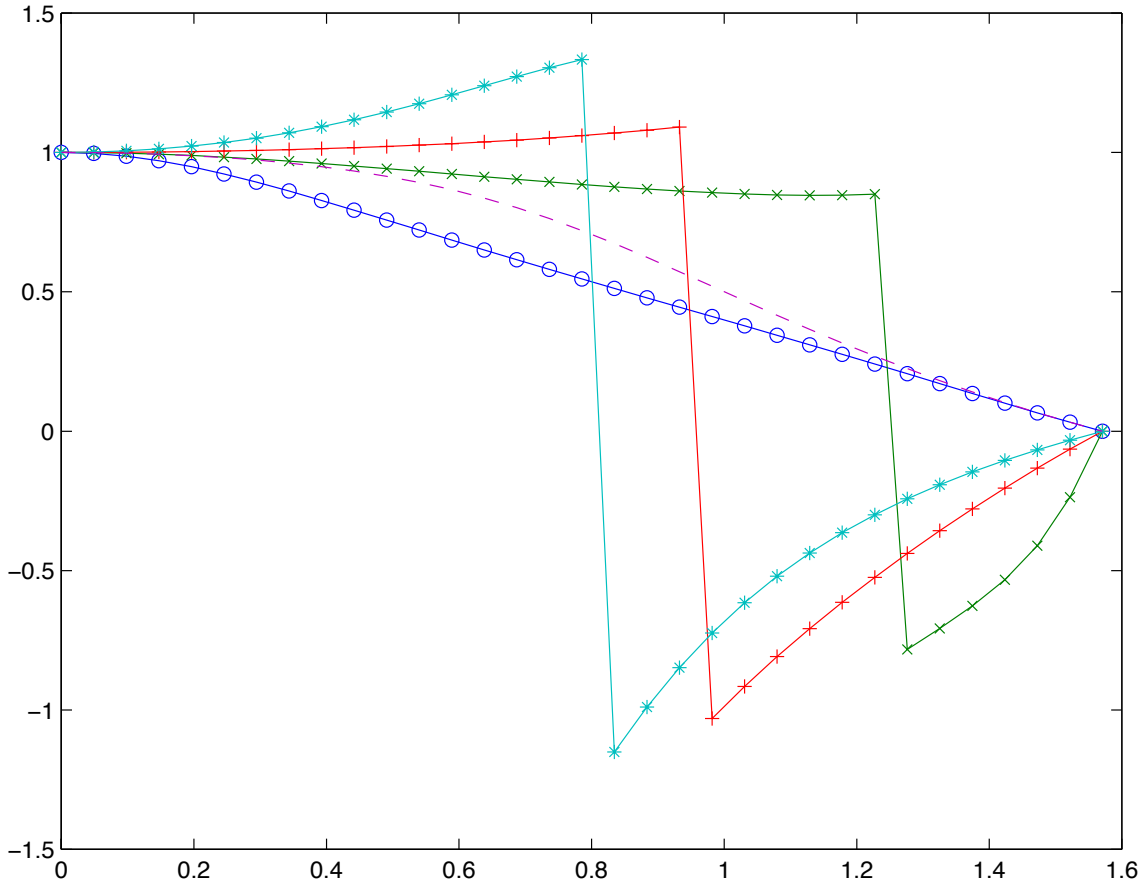

Figure 6.3.2: Normalized phase speeds γ_k/a when $\alpha = 0.75$ as functions of $\theta = k\pi/J$ for the centered (o), Lax-Wendroff (x), upwind (+), Lax-Friedrichs (*), and leap frog (−−) schemes.

$\gamma_k = 0$ (Figure 6.3.9b). Thus, instead of having speed a, waves of both schemes (6.3.2) and (6.3.4) are practically stationary when $\theta \approx \pi/2$.

In order to get additional qualitative information about the degradation of the phase speed with increasing θ, let's expand (6.3.9b) in a series about $\theta = 0$. Thus,

$$\tan\phi_k \approx 2\theta\alpha(1 - 2\theta^2/3)(1 + 2\beta\theta^2) \approx 2\theta\alpha[1 + 2\theta^2(\beta - 1/3)].$$

For small z, $\tan^{-1} z \approx z - z^3/3$; thus,

$$\phi_k \approx 2\theta\alpha[1 - \frac{2\theta^2}{3}(1 + 2\alpha^2 - 3\beta)].$$

Using (6.3.10b)

$$\gamma_k = \frac{\phi_k}{2\pi k\Delta t} \approx \frac{2(k\pi/J)(a\Delta t/\Delta x)}{2\pi k\Delta t}[1 - \frac{2\theta^2}{3}(1 + 2\alpha^2 - 3\beta)]$$

Method	β	Approximate γ_k
Centered	0	$a[1-2\theta^2(1+2\alpha^2)/3]$
Lax-Wendroff	α^2	$a[1-2\theta^2(1-\alpha^2)/3]$
Upwind	$\|\alpha\|$	$a[1-2\theta^2(1-\alpha)(1-2\alpha)/3]$
Lax-Friedrichs	1	$a[1+4\theta^2(1-\alpha^2)/3]$
Leap frog		$\pm a[1-2\theta^2(1-\alpha^2)/3]$

Table 6.3.2: Approximate phase speeds for the difference schemes (6.3.2) and (6.3.4) when $\theta << 1$.

or

$$\gamma_k \approx a[1-\frac{2\theta^2}{3}(1+2\alpha^2-3\beta)]. \tag{6.3.11a}$$

The results are summarized in Table 6.3 for the four schemes of Table 6.3. A similar analysis of the leap frog scheme gives (*cf.* Problem 1 at the end of this Section)

$$\gamma_k \approx \pm a[1-\frac{2\theta^2}{3}(1-\alpha^2)]. \tag{6.3.11b}$$

Clearly, the positive sign corresponds to the correct solution and the negative sign to a parasitic solution introduced by the multi-level scheme.

Comparing the data in Figure 6.3.2 and Table 6.3, all schemes considered have approximately the correct phase speed for low-frequency waves ($\theta \ll 1$), but have large errors when $\theta \approx \pi/2$. Assuming that $0 < \alpha \leq 1$, the data of Table 6.3 indicates that low-frequency waves travel slower than the correct speed for the centered scheme, the Lax-Wendroff scheme, the leap frog scheme, and the upwind scheme when $\alpha \leq 1/2$. Under the same assumption, low-frequency waves travel faster than the correct speed for the Lax-Friedrichs scheme and the upwind scheme when $\alpha > 1/2$. These conclusions are verified for $\alpha = 0.75$ in Figure 6.3.2.

Since the high-frequency waves have large phase-speed errors, it seems appropriate to dissipate them. The upwind and Lax-Wendroff schemes dissipate the high-frequency waves the most (Figure 6.3.1); thus, they would be good schemes to use when high-frequency information is present. The leap frog scheme, on the other hand, does not dissipate waves of any frequency and would not be a good one to use in connection with high-frequency data. While the Lax-Friedrichs scheme is dissipative for most frequencies,

it is less so for the highest-frequency modes. Finally, the centered scheme is not dissipative and should not be used.

Problems

1. Use the Von Neumann method to show that the amplification factors of the Leap frog difference scheme (6.3.4) for the initial value problem (6.3.1) are given by (6.3.8a). Show that the amplitudes and phase speeds of the amplification factors are given by (6.3.8b) and (6.3.9c), respectively. Obtain approximations of the phase speeds for small and large values of θ. Compare the approximate and true phase speeds.

6.4 Nonlinear Problems

The time step restriction implied by the Courant, Friedrichs, Lewy Theorem is not restrictive for many hyperbolic problems; thus, explicit finite difference schemes will often be acceptable. Implicit schemes are useful when (*i*) a problem involves widely disparate characteristic speeds, (*ii*) a steady-state (time independent) solution is sought, or (*iii*) reactions are involved. The treatment of implicit schemes for hyperbolic systems is fundamentally the same as for parabolic systems, so we'll concentrate on the simpler explicit schemes for solving a system of one-dimensional conservation laws having the form (6.1.1a).

We have just learned that the Lax-Wendroff scheme has good dissipative properties, particularly regarding high-frequency modes. It is, however, difficult to use because it generally requires knowledge of the derivatives of the Jacobian $\mathbf{A}$ (*cf.* (6.1.1c and Section 3.3). An analogy with Runge-Kutta methods for ordinary differential equations would suggest that derivatives of $\mathbf{A}$ could be avoided at the expense of additional evaluations of the flux $\mathbf{f}$. This typically leads to predictor-corrector methods and the first one that we will study is due to Richtmyer ([17], Section 12.7). Using this method, a predicted solution of (6.1.1a) is calculated by using the Lax-Friedrichs scheme for one-half time step, *i.e.*,

$$\mathbf{U}^{n+1/2}_{j+1/2} = \mu_x \mathbf{U}^n_{j+1/2} - \frac{\Delta t}{2\Delta x}\delta_x \mathbf{f}^n_{j+1/2} = \frac{\mathbf{U}^n_{j+1} + \mathbf{U}^n_j}{2} - \frac{\Delta t}{2\Delta x}(\mathbf{f}^n_{j+1} - \mathbf{f}^n_j). \qquad (6.4.1a)$$

This solution is corrected using the leap frog scheme

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}\delta_x \mathbf{f}_j^{n+1/2} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}(\mathbf{f}_{j+1/2}^{n+1/2} - \mathbf{f}_{j-1/2}^{n+1/2}). \tag{6.4.1b}$$

The computational stencils of the predictor and corrector steps of this scheme, which we'll call the *Richtmyer two-step method*, are shown in Figure 6.4.1. The local discretization error can be shown to be $O(\Delta t^2) + O(\Delta x^2)$. The $O(\Delta t^2)$ error is due to the centering of the leap frog corrector about $t_{n+1/2}$.

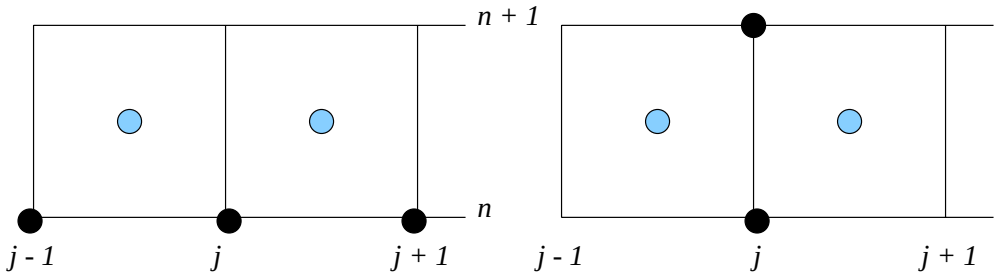

Figure 6.4.1: Computational stencils of the predictor (left) and corrector (right) stages of the Richtmyer two-step method (6.4.1). Predicted solutions are shown in light blue.

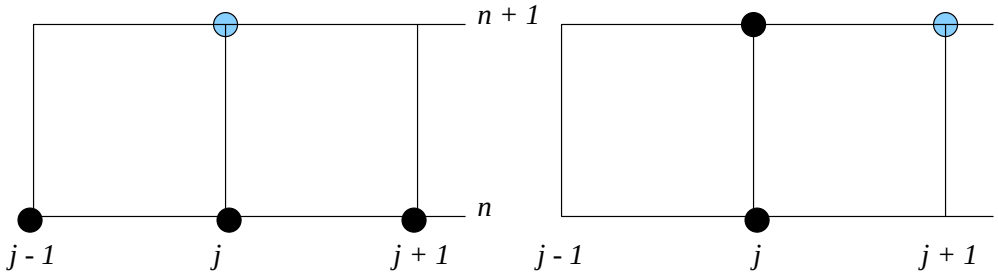

Figure 6.4.2: Computational stencils of the predictor (left) and corrector (right) stages of MacCormack's method (6.4.2). Predicted solutions are shown in light blue.

The Richtmyer two-step and Lax-Wendroff schemes are identical when $\mathbf{f}(\mathbf{u}) = \mathbf{A}\mathbf{u}$ and $\mathbf{A}$ is a constant. However, the Richtmyer two-step scheme is much simpler to implement than the Lax-Wendroff scheme for nonlinear systems. The equivalence of the two methods implies that the Richtmyer two-step scheme has similar properties to the Lax-Wendroff scheme. Indeed, both are stable for linear problems when the Courant, Friedrichs, Lewy condition $\max_{1\le i\le m} |\lambda_i|\Delta t/\Delta x \le 1$, is satisfied. Here λ_i, $1 \le i \le m$, is an eigenvalue of $\mathbf{A}$.

MacCormack's scheme (*cf.* [16], Section 4.5) uses a forward time-backward space predictor and a backward time-forward space corrector for one-half time step, *i.e.*,

$$\hat{\mathbf{U}}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}\nabla_x \mathbf{f}_j^n = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}(\mathbf{f}_j^n - \mathbf{f}_{j-1}^n), \tag{6.4.2a}$$

$$\mathbf{U}_j^{n+1} = \frac{\mathbf{U}_j^n + \hat{\mathbf{U}}_j^{n+1}}{2} - \frac{\Delta t}{2\Delta x}\Delta_x \hat{\mathbf{f}}_j^{n+1} = \frac{\mathbf{U}_j^n + \hat{\mathbf{U}}_j^{n+1}}{2} - \frac{\Delta t}{2\Delta x}(\hat{\mathbf{f}}_{j+1}^{n+1} - \hat{\mathbf{f}}_j^{n+1}) \tag{6.4.2b}$$

where $\hat{\mathbf{f}}_j^{n+1} \equiv \mathbf{f}(\hat{\mathbf{U}}_j^{n+1})$.

The computational stencils of the predictor and corrector stages of MacCormack's scheme are shown in Figure 6.4.2. MacCormack's scheme can be shown to possess the properties noted for the Richtmyer two-step scheme, including identity with the Lax-Wendroff scheme for linear constant-coefficient problems. Finally, MacCormack's scheme can also be used in reverse order as

$$\hat{\mathbf{U}}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}\Delta_x \mathbf{f}_j^n = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}(\mathbf{f}_{j+1}^n - \mathbf{f}_j^n), \tag{6.4.3a}$$

$$\mathbf{U}_j^{n+1} = \frac{\mathbf{U}_j^n + \hat{\mathbf{U}}_j^{n+1}}{2} - \frac{\Delta t}{2\Delta x}\nabla_x \hat{\mathbf{f}}_j^{n+1} = \frac{\mathbf{U}_j^n + \hat{\mathbf{U}}_j^{n+1}}{2} - \frac{\Delta t}{2\Delta x}(\hat{\mathbf{f}}_j^{n+1} - \hat{\mathbf{f}}_{j-1}^{n+1}). \tag{6.4.3b}$$

6.4.1 Artificial Viscosity

We have seen that nonlinear conservation laws can give rise to discontinuous solutions such as shock waves and contact surfaces. The principal computational approaches to treat problems with discontinuities involve (*i*) explicit tracking and (*ii*) implicit "capturing." Tracking is the more accurate of the two, but it is difficult to implement for problems in two and three dimensions. We'll concentrate on shock-capturing schemes that isolate discontinuities without knowing their locations. Let us begin with a simple scalar problem.

Example 6.4.1. Consider the Riemann problem for the inviscid Burgers' equation (*cf.* Problem 1.3.1)

$$u_t + \frac{(u^2)_x}{2} = 0, \qquad -\infty < x < \infty, \qquad t > 0, \tag{6.4.4a}$$

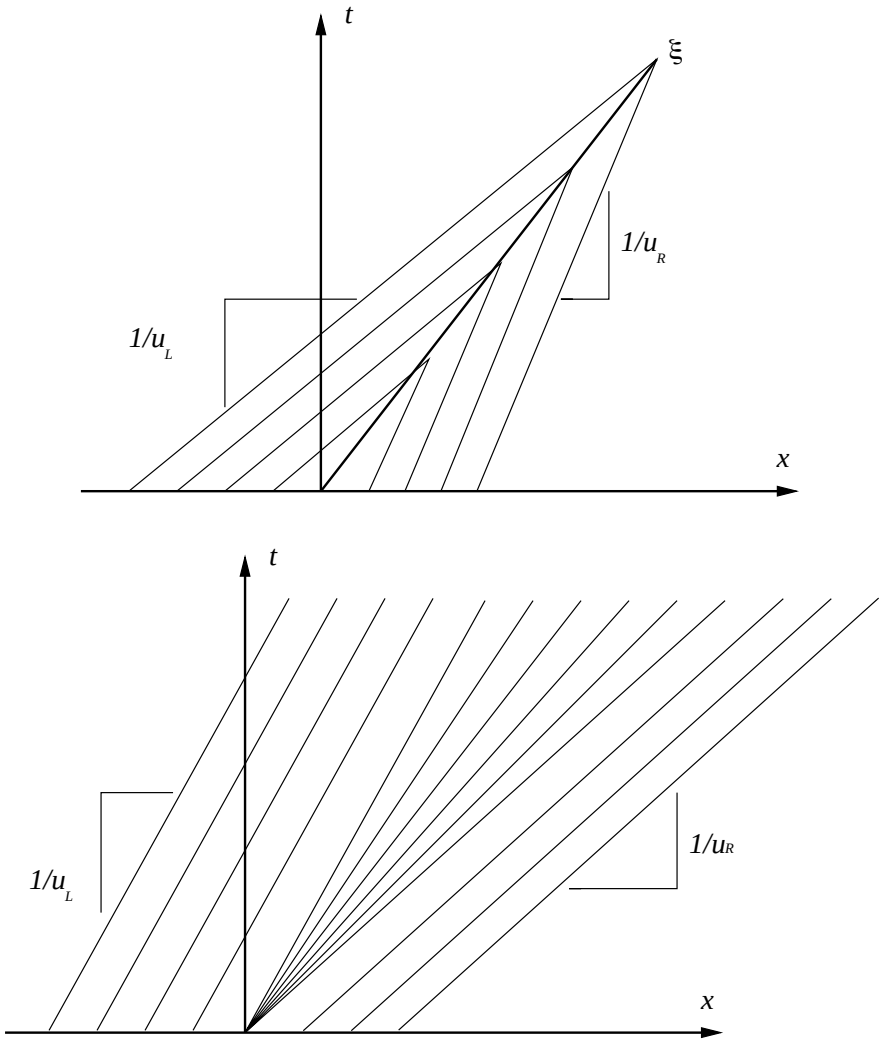

Figure 6.4.3: Shock (top) and expansion (bottom) wave characteristics of the Riemann problem (6.4.4).

$$u(x,0) = \begin{cases} u_L, & \text{if } x < 0 \\ u_R, & \text{if } x \geq 0 \end{cases} . \tag{6.4.4b}$$

Recall that a Riemann problem is a Cauchy (initial value) problem with piecewise constant initial data. If $u_L > u_R$ the solution of (6.4.4) features the shock wave

$$u(x,t) = \begin{cases} u_L, & \text{if } x/t < \dot{\xi} \\ u_R, & \text{if } x/t \geq \dot{\xi} \end{cases} , \qquad \dot{\xi} = \frac{u_L + u_R}{2}. \tag{6.4.5a}$$

If $u_L < u_R$ the solution of (6.4.4) has the expansion

$$u(x,t) = \begin{cases} u_L, & \text{if } x/t < u_L \\ x/t, & \text{if } u_L \leq x/t < u_R \\ u_R, & \text{if } x/t \geq u_R \end{cases} . \tag{6.4.5b}$$

The characteristics of the shock and expansion solutions are shown in Figure 6.4.3. The characteristics are straight lines with slopes of $1/u_L$ for $x/t < u_L$ and $1/u_R$ for $x/t \geq u_R$. No characteristics are generated within the expansion wave. The entropy condition that shocks can only exist when characteristics intersect can be used to infer that the two characteristic families are not separated by a shock. Viscosity arguments can be used to construct the expansion fan [14].

Consider solving a shock problem for (6.4.4) with $u_L = 1$ and $u_R = 0$ by the Lax-Friedrichs scheme

$$U_j^{n+1} = \frac{U_{j+1}^n + U_{j-1}^n}{2} - \frac{\bar{\alpha}}{4}[(U_{j+1}^n)^2 - (U_{j-1}^n)^2]$$

and the Richtmyer two-step (6.4.1) scheme

$$U_{j+1/2}^{n+1/2} = \frac{U_{j+1}^n + U_j^n}{2} - \frac{\bar{\alpha}}{4}[(U_{j+1}^n)^2 - (U_j^n)^2],$$

$$U_j^{n+1} = U_j^n - \frac{\bar{\alpha}}{2}[(U_{j+1/2}^{n+1/2})^2 - (U_{j-1/2}^{n+1/2})^2],$$

where

$$\bar{\alpha} = \frac{\Delta t}{\Delta x}. \tag{6.4.6}$$

Selecting $\Delta x = 0.1$ and $\bar{\alpha} = 0.75$, we show solutions obtained by the two methods for five time steps in Figure 6.4.4. The two numerical solutions are compared with the exact solution at $t = 3/40$ in Figure 6.4.5. The Lax-Friedrichs solution is a monotone decreasing function of x through the shock while the Richtmyer two step solution has an overshoot.

This situation is the same as the one that we observed for convection-diffusion systems (Section 4.4.2):

- first-order finite difference methods preserve monotonicity of the solution near a discontinuity but have excess dissipation and
- higher-order methods introduce oscillation near discontinuities.

Let us formally explore these notions, at least for scalar problems.

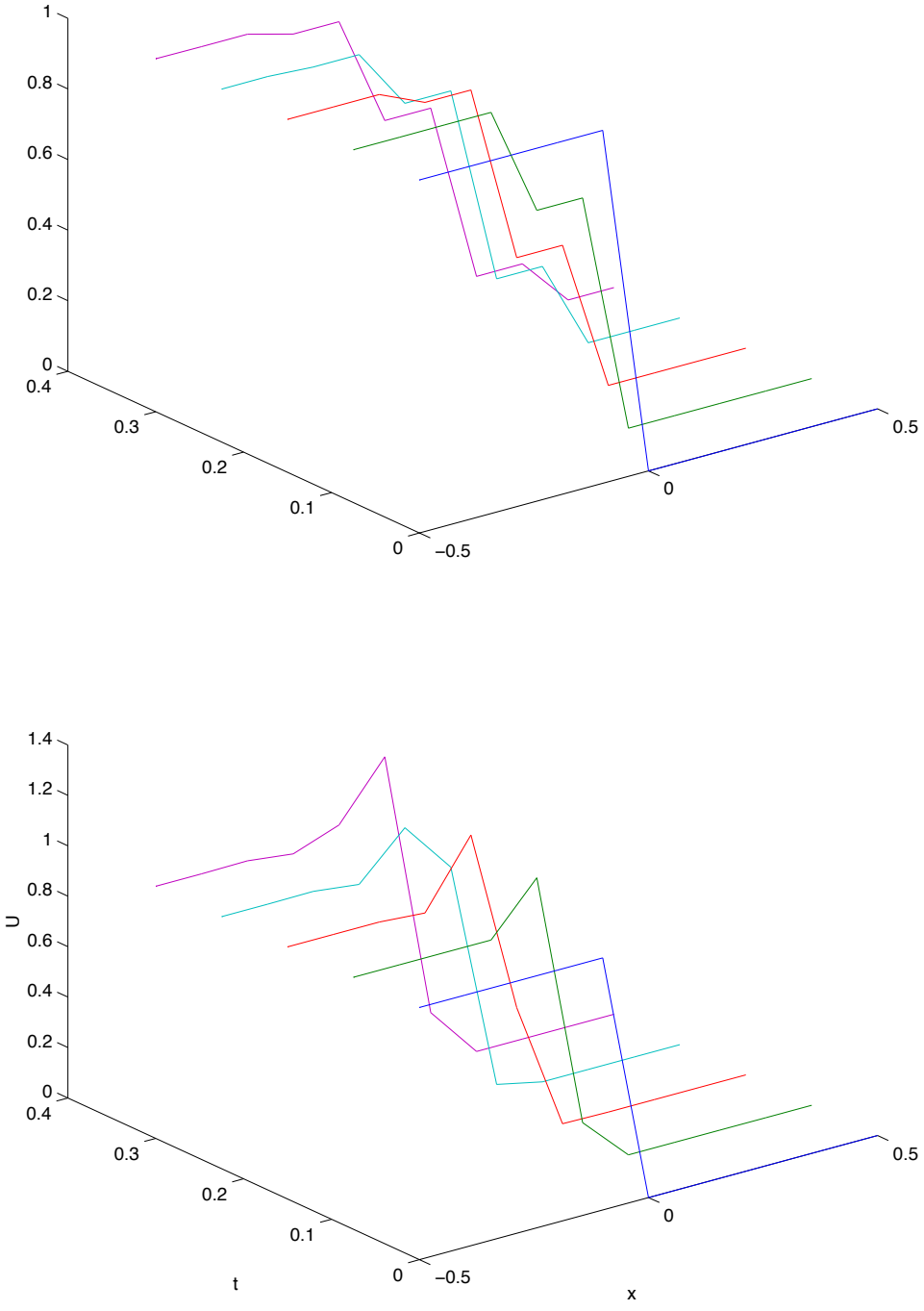

Figure 6.4.4: Shock wave solution of the Riemann problem (6.4.4) using the Lax-Friedrichs (top) and Richtmyer two-step (botom) schemes for five time steps.

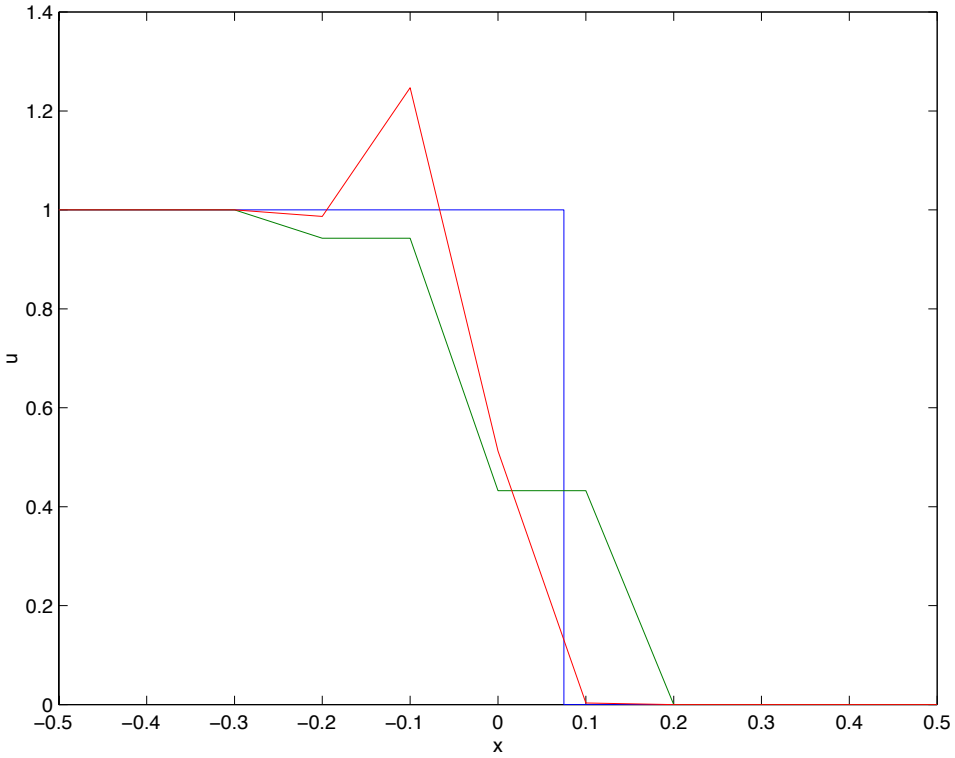

Figure 6.4.5: Comparison of the exact, Lax-Friedrichs, and Richtmyer two-step solutions of (6.4.4) at $t = 2\Delta t$.

Definition 6.4.1. *A finite-difference approximation of (6.1.1a) is in conservation form if it can be written as*

$$\frac{\mathbf{U}_j^{n+1} - \mathbf{U}_j^n}{\Delta t} + \frac{\mathbf{F}_{j+1/2}^n - \mathbf{F}_{j-1/2}^n}{\Delta x} = 0 \tag{6.4.7a}$$

where

$$\mathbf{F}_{j+1/2}^n = \mathbf{F}(\mathbf{U}_{j-S+1}^n, \mathbf{U}_{j-S+2}^n, \ldots, \mathbf{U}_{j+S}^n) \tag{6.4.7b}$$

and

$$\mathbf{F}_{j-1/2}^n = \mathbf{F}(\mathbf{U}_{j-S}^n, \mathbf{U}_{j-S+1}^n, \ldots, \mathbf{U}_{j+S-1}^n) \tag{6.4.7c}$$

for some integer $S > 0$. The function $\mathbf{F}$ is called the numerical flux and it is consistent with the physical flux $\mathbf{f}$ if

$$\mathbf{F}(\mathbf{v}, \mathbf{v}, \ldots, \mathbf{v}) = \mathbf{f}(\mathbf{v}) \tag{6.4.7d}$$

for any smooth function v.

Remark 1. Consistency, in this situation, implies that the numerical and exact fluxes agree for uniform data. You might check whether or not this is equivalent to our usual notion of consistency.

Definition 6.4.2. *A scalar finite difference scheme having the form*

$$U_j^{n+1} = G(U_{j-S}^n, U_{j-S+1}^n, \ldots, U_{j+S}^n) \tag{6.4.8a}$$

is monotone if it is a monotone non-decreasing function of its arguments, i.e., if

$$\frac{\partial G}{\partial U_{j+k}^n} \geq 0, \qquad |k| \leq S. \tag{6.4.8b}$$

For conservative schemes (6.4.7),

$$G(U_{j-S}^n, U_{j-S+1}^n, \ldots, U_{j+S}^n) = U_j^n - \bar{\alpha}(F_{j+1/2}^n - F_{j-1/2}^n). \tag{6.4.8c}$$

Remark 2. This definition applies to vector systems of conservation laws; however, we present it for scalar equations for simplicity.

Example 6.4.2. Consider the Lax-Friedrichs scheme for a scalar conservation law having the form (6.1.1a)

$$U_j^{n+1} = G(U_{j-1}^n, U_j^n, U_{j+1}^n) = \frac{U_{j+1}^n + U_{j-1}^n}{2} - \frac{\bar{\alpha}}{2}[f(U_{j+1}^n) - f(U_{j-1}^n)].$$

Differentiating

$$\frac{\partial G}{\partial U_{j-1}^n} = \frac{1 + \alpha(U_{j-1}^n)}{2}, \qquad \frac{\partial G}{\partial U_j^n} = 0, \qquad \frac{\partial G}{\partial U_{j+1}^n} = \frac{1 - \alpha(U_{j+1}^n)}{2},$$

where

$$\alpha(v) = \frac{a(v)\Delta t}{\Delta x}, \qquad a(v) = f'(v).$$

Thus,

$$\frac{\partial G}{\partial U_k^n} \geq 0, \qquad k = j, j \pm 1,$$

and the scheme is monotone when $|\alpha| \leq 1$.

Define

$$F_{j+1/2}^n = F(U_j^n, U_{j+1}^n) = -\frac{1}{2\bar{\alpha}}(U_{j+1}^n - U_j^n) + \frac{f_{j+1}^n + f_j^n}{2}$$

and

$$F^n_{j-1/2} = F(U^n_{j-1}, U^n_j) = -\frac{1}{2\bar{\alpha}}(U^n_j - U^n_{j-1}) + \frac{f^n_j + f^n_{j-1}}{2}$$

Thus, the Lax-Friedrichs scheme can be written in conservation form (6.4.7). In a similar fashion, we can show that the Richtmyer two-step scheme (6.4.1) can be written in conservation form but it is not monotone (*cf.* Problem 1 at the end of this Section).

Monotone difference schemes produce smooth transitions near discontinuities; however, as indicated by the following theorem, their existence is limited.

Theorem 6.4.1. *A monotone finite difference scheme in the conservation form (6.1.1a) must be first-order accurate.*

Proof. cf. Sod [19] or Harten, *et al.* [9]. □

The results of Section 6.3 (Figure 6.3.1) indicate that the Lax-Wendroff (hence, the Richtmyer two-step) scheme has less dissipation for moderately high frequencies than either the Lax-Friedrichs or upwind schemes. Since discontinuous solutions have all frequencies present, perhaps the oscillations could be reduced by adding even more viscosity.

Consider any second-order difference scheme and add a diffusive term to its right-hand side having the form

$$\frac{\Delta x}{2}\frac{\delta_x(\mathbf{Q}_j\delta_x\mathbf{U}^n_j)}{\Delta x^2} = \frac{1}{2\Delta x}[\mathbf{Q}_{j+1/2}(\mathbf{U}^n_{j+1} - \mathbf{U}^n_j) - \mathbf{Q}_{j-1/2}(\mathbf{U}^n_j - \mathbf{U}^n_{j-1})], \tag{6.4.9a}$$

where $\mathbf{Q}_{j+1/2} = \mathbf{Q}_{j+1/2}(\mathbf{U}^n_j, \mathbf{U}^n_{j+1})$ is an $m \times m$ matrix that vanishes when $\mathbf{U}^n_{j+1} - \mathbf{U}^n_j = 0$. Artificial viscosity terms of this form were originally studied by Lax and Wendroff [15] and are discussed in Richtmyer and Morton [17], Section 12.14, and Sod [19], Section 4.2. For scalar problems, Lax and Wendroff [15] chose

$$Q_{j+1/2}(U^n_j, U^n_{j+1}) = \frac{\epsilon}{2}|a^n_{j+1} - a^n_j|, \tag{6.4.9b}$$

where ϵ is a parameter of order unity. Thus, for example, the Lax-Wendroff scheme with artificial viscosity would become (*cf.* Section 3.3)

$$U^{n+1}_j = U^n_j - \alpha^n_j\mu_x\delta_x U^n_j + (\alpha^n_j)^2\delta^2_x U^n_j + \frac{\Delta t}{2\Delta x}\delta_x(Q_j\delta_x\mathbf{U}^n_j), \tag{6.4.9c}$$

where α^n_j is the Courant number.

Extending the artificial viscosity model to vector systems of conservation laws is more difficult. Lax and Wendroff [15] define the matrix $\mathbf{Q}_{j+1/2}$ in (6.4.9a) to have eigenvalues $\epsilon_i|\lambda_{i,j+1} - \lambda_{i,j}|$, $i = 1, 2, \ldots, m$, where $\lambda_{i,j}$, $i = 1, 2, \ldots, m$, are the eigenvalues of the Jacobian $\mathbf{A}(\mathbf{U}_j^n) = \mathbf{f}_\mathbf{u}(\mathbf{U}_j^n)$ and ϵ_i, $i = 1, 2, \ldots, m$, are approximately one. This eigenvalue condition is not sufficient to uniquely determine $\mathbf{Q}_{j\pm1/2}$ and Lax and Wendroff [15] required $\mathbf{Q}_{j\pm1/2}$ to commute with $\mathbf{A}$. They did this by prescribing $\mathbf{Q}_{j\pm1/2}$ as $m-1$ *th* degree polynomials in $\mathbf{A}$.

Let us illustrate the performance of the artificial viscosity technique using a simple example.

Example 6.4.3. Consider an initial value problem for the inviscid Burgers' equation (6.4.4a) with the data

$$u(x,0) = \phi(x) = \begin{cases} 1, & \text{if } x < 0 \\ 1-x, & \text{if } 0 \le x < 1 \\ 0, & \text{if } 1 \le x \end{cases} .$$

Recall (Section 1.3) that the exact solution is

$$u(x,t) = \begin{cases} 1, & \text{if } x < t \\ \frac{x-1}{t-1}, & \text{if } t \le x < 1 \\ 0, & \text{if } 1 \le x \end{cases} , \qquad t < 1,$$

$$u(x,t) = \begin{cases} 1, & \text{if } x < (t+1)/2 \\ 0, & \text{if } x \ge (t+1)/2 \end{cases} , \qquad t \ge 1.$$

The initial ramp steepens into a shock wave at $t = 1$. For $t > 1$, the shock travels in the positive x direction with speed $1/2$.

We solved this problem using upwind differencing, the Lax-Friedrichs method, and the Richtmyer two-step method with and without the artificial viscosity model (6.4.9). Sod [19] solved the same problem using other methods as well as these. In each case, we used $\Delta x = 0.02$, $\Delta t = 0.01$, and the artificial viscosity parameter $\epsilon = 0.9$. Solutions are shown in Figures 6.4.6 - 6.4.9 as functions of x for $t = 0.5$, 1.0, and 2.0.

The solution shown in Figure 6.4.7 using the Lax-Friedrichs scheme has the greatest dissipation. The compression wave and shock have been spread over several computational cells. The upwind scheme, shown in Figure 6.4.6, has much less apparent dissipation. As expected, the Richtmyer two-step scheme has oscillations trailing the shock;

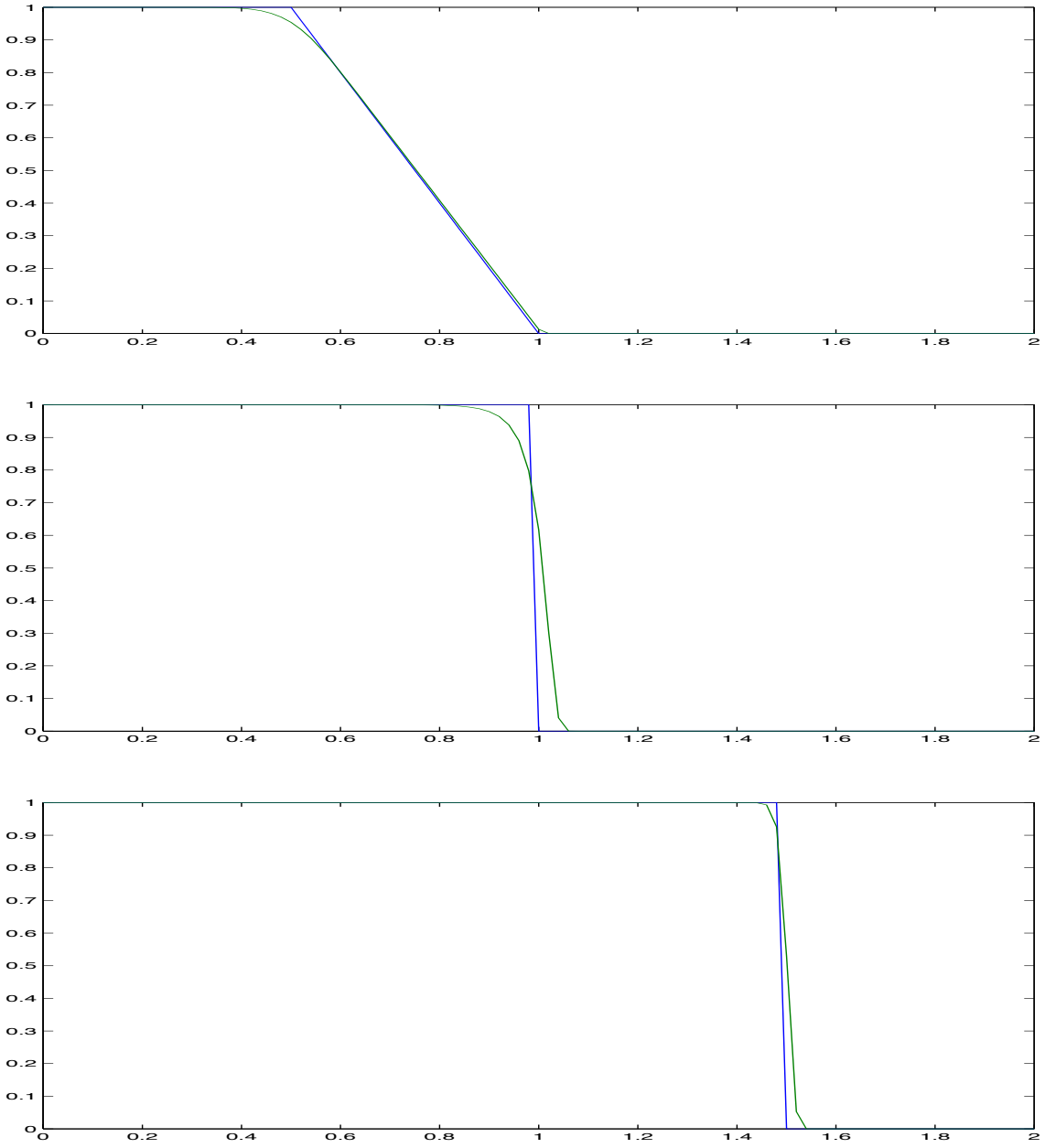

Figure 6.4.6: Comparison of exact and upwind-difference solutions of Example 6.4.3 at $t = 0.5, 1.0, 2.0$ (top to bottom).

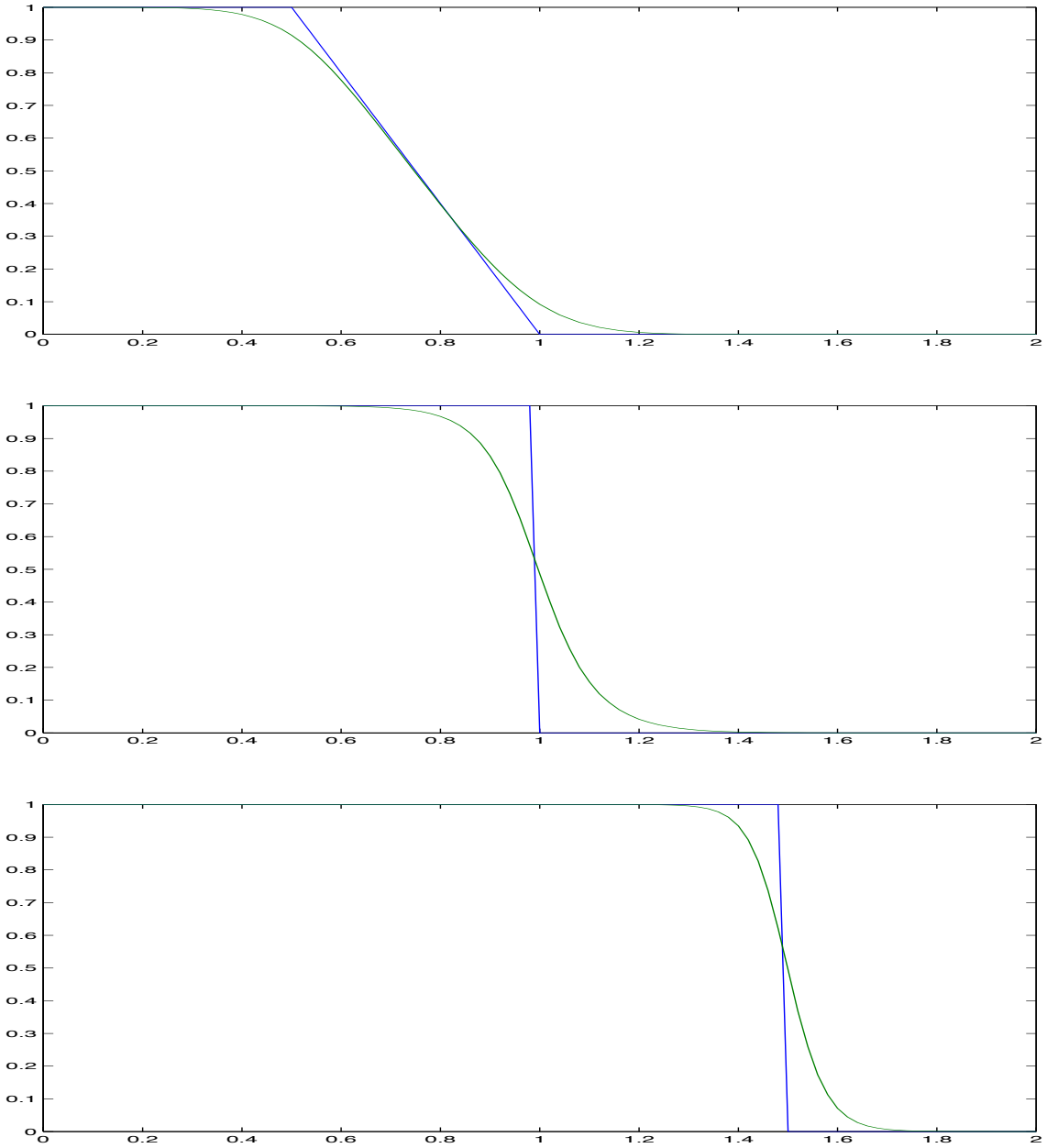

Figure 6.4.7: Comparison of exact and Lax-Friedrichs solutions of Example 6.4.3 at $t = 0.5, 1.0, 2.0$ (top to bottom).

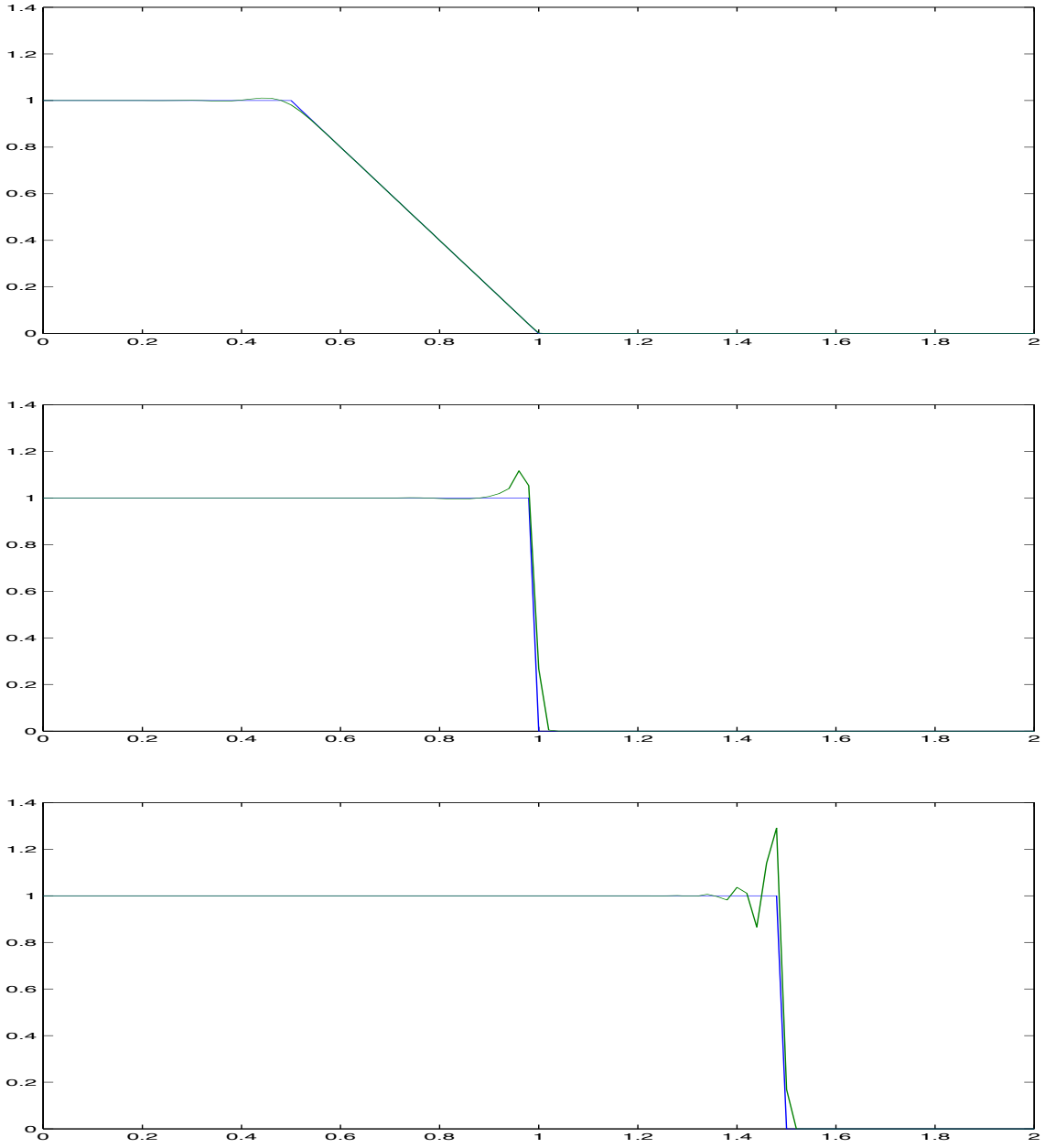

Figure 6.4.8: Comparison of exact and Richtmyer two-step solutions of Example 6.4.3 at $t = 0.5, 1.0, 2.0$ (top to bottom).

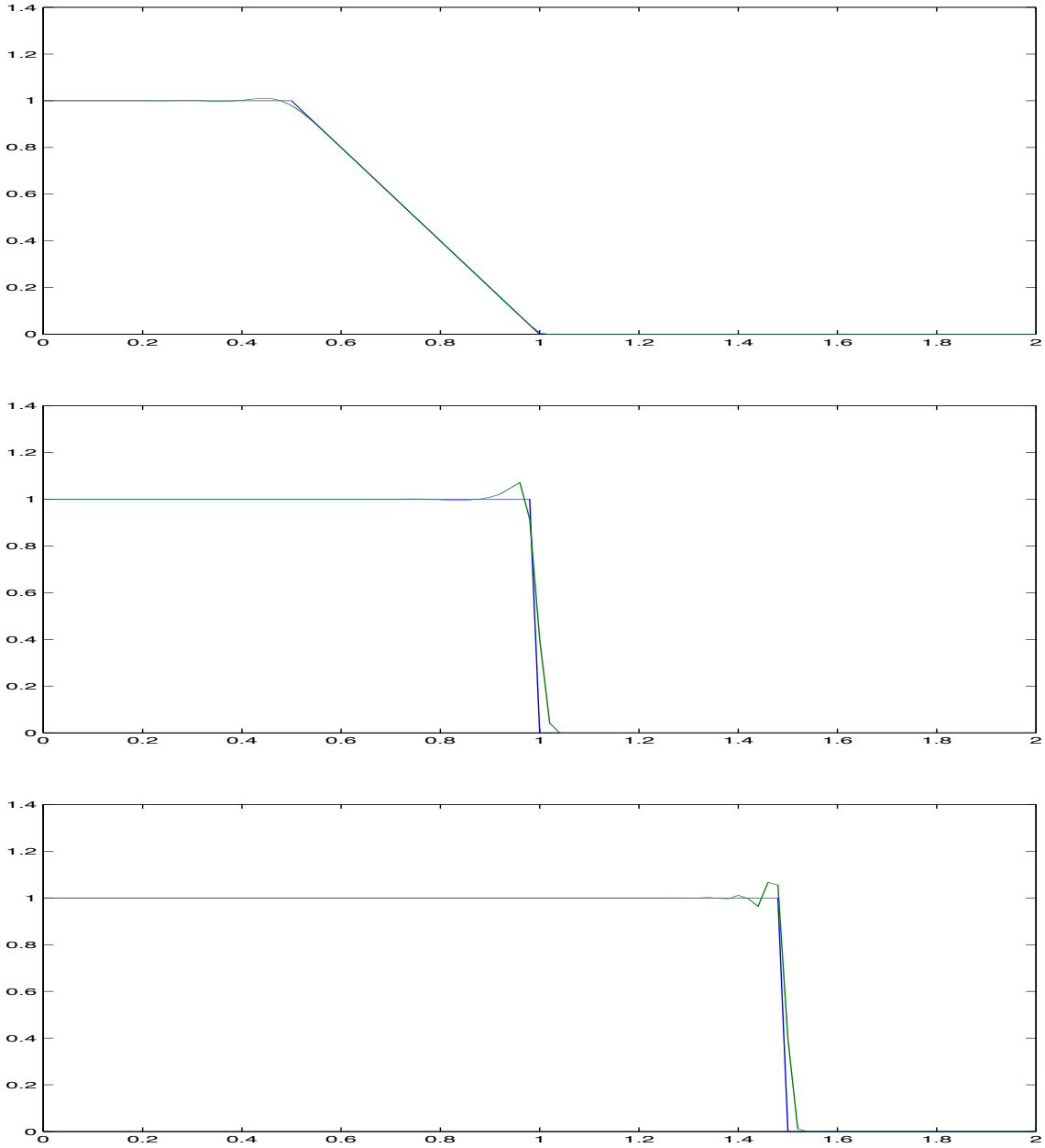

Figure 6.4.9: Comparison of exact and Richtmyer two-step with added viscosity solutions of Example 6.4.3 at $t = 0.5, 1.0, 2.0$ (top to bottom).

however, accuracy when the solution is continuous is quite good. The oscillations of the Richtmyer two-step scheme are reduced when artificial viscosity is added (Figure 6.4.9). For shock waves, the excess dissipation does not spread as time increases. This is not the case with the linear waves studied in Section 6.3. The intersecting characteristics tend to confine the excess dissipation to the vicinity of the shock. This does not happen for linear problems where discontinuities propagate along characteristics.

Problems

1. Show that the Richtmyer two-step scheme (6.4.1) can be written in conservation form but is not monotone. What is the numerical flux function?

2. Consider the initial value problem

$$\mathbf{u}_t + \mathbf{f}(\mathbf{u})_x = \mathbf{g}(\mathbf{u}), \qquad \mathbf{u}(x,0) = \boldsymbol{\phi},$$

where $\mathbf{u}$, $\mathbf{f}$, $\mathbf{g}$, and $\boldsymbol{\phi}$ are m-vectors that are periodic in x with period 1.

2.1. Write a computer program to solve the above initial value problem on $0 < x < 1$, $0 < t \leq T$ using the Richtmyer two-step procedure (6.4.1).

2.2. Let $m = 2$, $u_1 = u$, $u_2 = v$, and consider the linear wave equation

$$u_t - v_x = 0, \qquad v_t - u_x = 0,$$

with the initial conditions on $0 \leq x < 1$ given as

$$u(x,0) = 0, \qquad v(x,0) = \begin{cases} 1, & \text{if } 0 \leq x < 1/2 \\ -1, & \text{if } 1/2 \leq x < 1 \end{cases}.$$

Solve the problem on $0 < t \leq T = 1$ for all combinations of $J, N = 10, 20, 30$ such that $\bar{\alpha} = \Delta t/\Delta x \leq 1$. Calculate the exact solution and plot the exact and numerical values of u_1, u_2 as functions of x for $t = 0$, 0.25, 0.5, 0.75, 1. Calculate the errors in the maximum norm at $t = 1$ for each component of the solution and each mesh. Comment on the accuracy and convergence rate.

2.3. Let $m = 1$, $u = u_1$, and consider the inviscid Burgers' equation (6.4.4a) with the initial conditions on $0 \leq x < 1$ given as

$$u(x,0) = \begin{cases} 1, & \text{if } 0 \leq x < 0.25 \\ 0, & \text{if } 0.25 \leq x < 0.75 \\ 1, & \text{if } 0.75 \leq x < 1 \end{cases}.$$

Solve this problem on $0 < t \leq T = 2$ for the same combinations of J and N given in Part 2. Calculate the exact solution and compare the exact and numerical solutions at by plotting them at $t = 0, 0.5, 1, 1.5, 2$. Calculate errors in the maximum norm at $t = 2$. Comment on the accuracy and estimate the convergence rate.

6.5 Godunov Schemes

The shock capturing finite difference schemes for nonlinear hyperbolic problems have not been very satisfying. To reiterate, the first-order methods can preserve solution monotonicity near discontinuities by introducing excess dissipation and the higher-order methods oscillate near discontinuities. In this section, we will study some modern approaches that try to alleviate these difficulties. The problem is not solved and the subject matter is still an active area of research. Thus, our results will be somewhat tentative.

Once again, let us focus on a one-dimensional system of conservation laws having the form (6.1.1). When discontinuities are present, solutions are expected to satisfy the integrated form of (6.1.1a)

$$\frac{d}{dt}\int_a^b \mathbf{u}dx = -\mathbf{f}(\mathbf{u})\Big|_a^b \tag{6.5.1}$$

on any interval $a < x < b$ (*cf.* Section 1.3). Discontinuities move according to the Rankine-Hugoniot conditions

$$\dot{\xi}(\mathbf{u}_R - \mathbf{u}_L) = \mathbf{f}(\mathbf{u}_R) - \mathbf{f}(\mathbf{u}_L), \tag{6.5.2}$$

where $\mathbf{u}_R$ and $\mathbf{u}_L$ are solutions values immediately to the right and left, respectively, of the discontinuity. The Rankine-Hugoniot conditions do not necessarily determine a unique shock speed for vector systems (6.1.1). A way of introducing discontinuities directly into the solution is to construct a "weak form" of the problem using Galerkin's method. Weak solutions automatically satisfy the partial differential system (6.1.1a) in regions where $\mathbf{u}$ is smooth and the Rankine-Hugoniot conditions (6.5.2) across jumps. Galerkin's method is at the core of the finite element method. While we won't discuss this approach here, we will note that problems still exist. Weak solutions are not uniquely determined from

the initial data. The expansion wave of Example 6.4.1 illustrates this. One way of eliminating the non-uniqueness is to view solutions of (6.1.1) as limiting solutions of a viscous problem, *e.g.*,

$$\mathbf{u}_t + \mathbf{f}(\mathbf{u})_x = \epsilon \mathbf{u}_{xx}, \tag{6.5.3}$$

as the viscosity $\epsilon \to 0$. The proper form of the viscous term may not be known and different limiting solutions may result from different viscous models.

An alternate and preferable means of eliminating the non-uniqueness of solutions is through the use of an entropy function. Again, we will not develop this topic further but note that additional material may be found in, *e.g.*, Harten and Lax [10] and Jeffrey and Taniuti [13].

Having studied Riemann problems in Section 6.4, let us attempt to use their solutions to construct numerical schemes. Thus, let $\mathbf{v}(x/t, \mathbf{u}_L, \mathbf{u}_R)$ be the solution of a Riemann problem for (6.1.1a) with the piecewise-constant initial data

$$\mathbf{u}(x,0) = \begin{cases} \mathbf{u}_L, & \text{if } x < 0 \\ \mathbf{u}_R, & \text{if } x \geq 0 \end{cases} . \tag{6.5.4}$$

Example 6.5.1. The solution of the Riemann problem for the inviscid Burgers' equation (6.4.4a) is

$$v(x/t, u_L, u_R) = \begin{cases} u_L, & \text{if } x/t < (u_L + u_R)/2 \\ u_R, & \text{if } x/t \geq (u_L + u_R)/2 \end{cases} , \qquad \text{if } u_L > u_R,$$

and

$$v(x/t, u_L, u_R) = \begin{cases} u_L, & \text{if } x/t < u_L \\ x/t, & \text{if } u_L \leq x/t < u_R \\ u_R, & \text{if } x/t \geq u_R \end{cases} , \qquad \text{if } u_L \leq u_R.$$

Godunov [7] was the first to use solutions of Riemann problems to construct finite-difference procedures. Glimm [6] introduced a random component into the solution of Riemann problems like (6.1.1a, 6.5.4) as an analytical tool to study existence and uniqueness of solutions of initial value problems for (6.1.1). Chorin [2] showed that Glimm's study could be used to create a numerical technique called the "random choice method." There are several variants of Chorin's approach and Sod [19], Section 3.12, discusses some of them. The difference between the Glimm-Chorin random choice procedure and

Godunov's finite-difference scheme and its successors is that the former renders shocks as exact discontinuities. Shock speeds will generally only be correct on average, but the discontinuities will not have any dissipation associated with them. Godunov schemes are shock capturing schemes that have dissipation associated with them and, thus, spread discontinuities over a few computational cells.

Shocks do not have to be perfectly sharp for most applications. Some dissipation is generally acceptable. Additionally, exact solution of the associated Riemann problem may neither be possible nor practical. Let us, thus, consider the construction of finite-difference schemes that are based on the approximate solution of Riemann problems and we'll begin with a difference approximation of (6.1.1) in conservative form (6.4.7). Lax and Wendroff [15] and Sod [19], Section 4.2, show that finite difference schemes in conservation form converge as Δx and Δt approach zero to functions $\mathbf{w}(x,t)$ that are weak solutions of (6.1.1a).

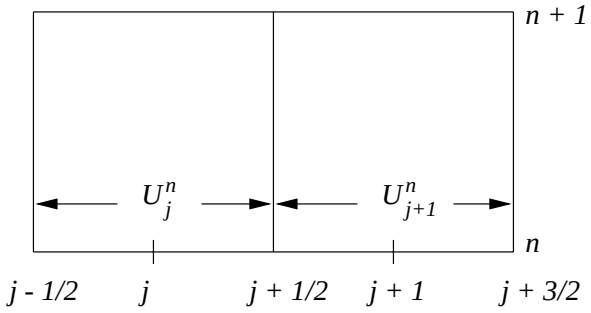

Figure 6.5.1: Grid structure for Godunov's method.

Let us introduce a uniform grid of spacing $\Delta x \times \Delta t$ and regard the solution at time level n as being a piecewise constant function of x given by

$$\mathbf{U}^n(x) = \mathbf{U}_j^n, \qquad (j-1/2)\Delta x < x \le (j+1/2)\Delta x, \qquad -\infty < j < \infty, \tag{6.5.5}$$

(Figure 6.5.1). The solution of a Riemann problem "breaking" at $(x_{j+1/2}, t_n)$ with this piecewise constant data is $\mathbf{v}((x-x_{j+1/2})/(t-t_n), \mathbf{U}_j^n, \mathbf{U}_{j+1}^n)$, $x \in (x_j, x_{j+1}]$, $t \in (t_n, t_{n+1}]$, provided that solutions of Riemann problems on neighboring subintervals do not interact. This will be the case if time steps are selected to satisfy the Courant, Friedrichs, Lewy

condition

$$\max_j \max_{1\le i\le m} |\lambda_i(\mathbf{A}_j^n)|\frac{\Delta t}{\Delta x} \le \frac{1}{2}, \tag{6.5.6a}$$

where λ_i, $i = 1, 2, \ldots, m$, are the eigenvalue of $\mathbf{A}_j^n$. Assuming (6.5.6a) is satisfied, the solution of all of the local Riemann problems can be written in the compact form

$$\mathbf{V}(x,t) = \mathbf{v}((x - x_{j+1/2})/(t - t_n), \mathbf{U}_j^n, \mathbf{U}_{j+1}^n), \qquad x \in (x_j, x_{j+1}], \qquad t \in (t_n, t_{n+1}]. \tag{6.5.6b}$$

With Godunov's [7] method, the solution of an initial value problem at time t_{n+1} is obtained by projecting the solution of the Riemann problems onto piecewise constant functions, *i.e.*,

$$\mathbf{U}_j^{n+1} = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{V}(x, t_{n+1})dx. \tag{6.5.7}$$

Since $\mathbf{V}(x,t)$ is an exact solution of the the partial differential system (6.1.1a),

$$\mathbf{0} = \int_{x_{j-1/2}}^{x_{j+1/2}}\int_{t_n}^{t_{n+1}} [\mathbf{V}_t + \mathbf{f}_x(\mathbf{V})]dxdt = \int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{V}(x,t)|_{t_n}^{t_{n+1}}dx + \int_{t_n}^{t_{n+1}} \mathbf{f}(\mathbf{V})|_{x_{j-1/2}}^{x_{j+1/2}}dt. \tag{6.5.8}$$

However, $\mathbf{V}(x_{j-1/2}, t) = \mathbf{v}(0, \mathbf{U}_{j-1}^n, \mathbf{U}_j^n)$ and $\mathbf{V}(x_{j+1/2}, t) = \mathbf{v}(0, \mathbf{U}_j^n, \mathbf{U}_{j+1}^n)$ are independent of t; thus,

$$\int_{t_n}^{t_{n+1}} \mathbf{f}(\mathbf{V})|_{x_{j-1/2}}^{x_{j+1/2}}dt = [\mathbf{f}(\mathbf{v}(0, \mathbf{U}_j^n, \mathbf{U}_{j+1}^n)) - \mathbf{f}(\mathbf{v}(0, \mathbf{U}_{j-1}^n, \mathbf{U}_j^n))]\Delta t.$$

Using this and (6.5.7) in (6.5.8) yields

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \bar{\alpha}[\mathbf{f}(\mathbf{v}(0, \mathbf{U}_j^n, \mathbf{U}_{j+1}^n)) - \mathbf{f}(\mathbf{v}(0, \mathbf{U}_{j-1}^n, \mathbf{U}_j^n))] \tag{6.5.9a}$$

where $\bar{\alpha} = \Delta t/\Delta x$. This is Godunov's [7] method. It clearly has the conservation form (6.4.7) with the numerical flux

$$\mathbf{F}_G(\mathbf{U}_j^n, \mathbf{U}_{j+1}^n) = \mathbf{f}(\mathbf{v}(0, \mathbf{U}_j^n, \mathbf{U}_{j+1}^n)). \tag{6.5.9b}$$

The subscript G is used merely to identify the numerical flux associated with Godunov's method.

Example 6.5.2. Using the solution of the Riemann problem for the inviscid Burgers' equation of Example 6.5.1, we have

$$v(0,u_L,u_R)=\begin{cases} u_L, & \text{if } u_L,u_R>0\\ u_R, & \text{if } u_L,u_R<0\\ 0, & \text{if } u_L<0,\ u_R>0\\ u_L, & \text{if } u_L>0,\ u_R<0,\ (u_L+u_R)/2>0\\ u_R, & \text{if } u_L>0,\ u_R<0,\ (u_L+u_R)/2<0 \end{cases}.$$

The characteristics for these solutions are illustrated in Figure 6.5.2. The numerical flux is

$$F_G(u_L,u_R)=\begin{cases} u_L^2/2, & \text{if } u_L,u_R>0\\ u_R^2/2, & \text{if } u_L,u_R<0\\ 0, & \text{if } u_L<0,\ u_R>0\\ u_L^2/2, & \text{if } u_L>0,\ u_R<0,\ (u_L+u_R)/2>0\\ u_R^2/2, & \text{if } u_L>0,\ u_R<0,\ (u_L+u_R)/2<0 \end{cases}.$$

This flux can be written more compactly by letting

$$u^+=\max(u,0),\qquad u^-=\min(u,0). \tag{6.5.10a}$$

Then

$$F_G(u_L,u_R)=\max[(u_L^+)^2/2,(u_R^-)^2/2]. \tag{6.5.10b}$$

Several schemes deriving from Godunov's [7] early work are now available. We'll describe techniques developed by Roe [18], van Leer [22], and Cockburn and Shu [3].

6.5.1 Roe's Method

Roe [18] developed a scheme having the conservation law form (6.4.7) with the numerical flux

$$\mathbf{F}_R(\mathbf{U}_j^n,\mathbf{U}_{j+1}^n)=\mathbf{f}(\mathbf{w}(0,\mathbf{U}_j^n,\mathbf{U}_{j+1}^n)), \tag{6.5.11a}$$

where $\mathbf{w}(x/t,\mathbf{u}_L,\mathbf{u}_R)$ is the exact solution of the linearized Riemann problem

$$\mathbf{w}_t+\hat{\mathbf{A}}(\mathbf{u}_L,\mathbf{u}_R)\mathbf{w}_x=0. \tag{6.5.11b}$$

The $m\times m$ matrix $\hat{\mathbf{A}}$ is an approximation of the Jacobian $\mathbf{f}_\mathbf{u}$ satisfying

$$\mathbf{f}(\mathbf{u}_R)-\mathbf{f}(\mathbf{u}_L)=\hat{\mathbf{A}}(\mathbf{u}_L,\mathbf{u}_R)(\mathbf{u}_R-\mathbf{u}_L). \tag{6.5.11c}$$

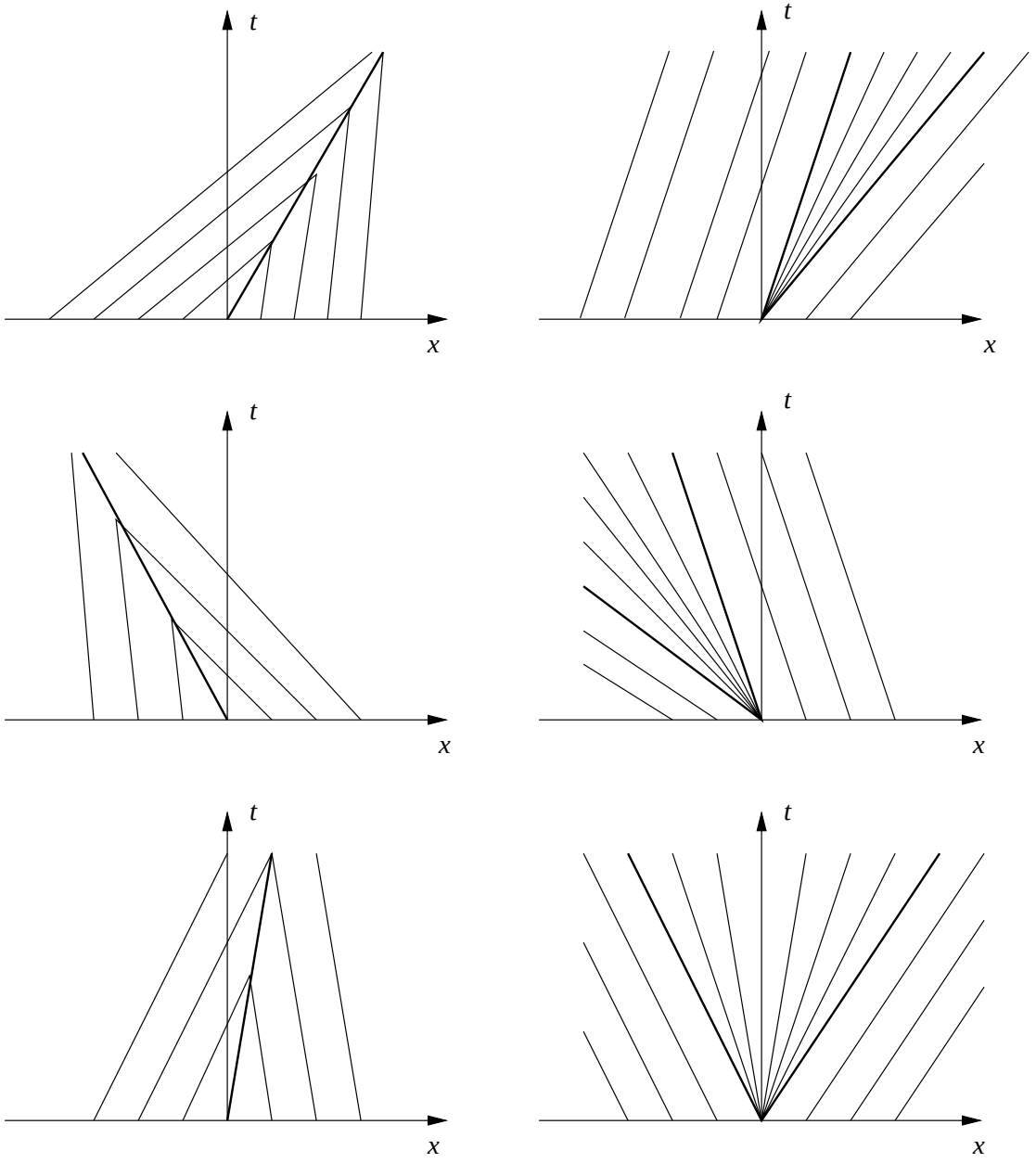

Figure 6.5.2: Characteristics of Riemann problems for Burgers' equation when $u_L, u_R > 0$ (top); $u_L, u_R < 0$ (center); $u_L > 0$, $u_R < 0$, $(u_L + u_R)/2 > 0$ (bottom left); and $u_L < 0, u_R > 0$ (bottom right).

Example 6.5.3. The flux for the inviscid Burgers' equation is $f(u) = u^2/2$ and (6.5.11c) becomes

$$(u_R^2 - u_L^2)/2 = \hat{A}(u_L, u_R)(u_R - u_L).$$

Thus, for this scalar problem

$$\hat{A}(u_L, u_R) = (u_R + u_L)/2 = \dot{\xi},$$

where $\dot{\xi}$ is the shock speed. Thus, Roe's numerical flux for the inviscid Burgers' equation is

$$F_R(u_L, u_R) = \begin{cases} u_L^2/2, & \text{if } \dot{\xi} \geq 0 \\ u_R^2/2, & \text{if } \dot{\xi} < 0 \end{cases} .$$

The approximate Riemann solver is exact when $u_L > u_R$ and a shock is produced; however, shocks are also produced when $u_L < u_R$ and an expansion wave should occur. Thus, Roe's scheme violates the entropy condition. Harten and Hyman [8] introduced a modification that eliminates the entropy violation. We'll depart from their development and present an approach due to Harten, Lax, and van Leer [11]. Both approaches appear in Sod [19], Section 4.4.

Let ν_{min} ν_{max}, be the minimum and maximum signal speeds associated with a Riemann problem breaking at $(x_{j+1/2}, t_n)$ (Figure 6.5.3). The solution for

$$\frac{x - x_{j+1/2}}{t - t_n} > \nu_{max}$$

is $\mathbf{u}_R$. Similarly, the solution for

$$\frac{x - x_{j+1/2}}{t - t_n} < \nu_{min}$$

is $\mathbf{u}_L$ (Figure 6.5.3). Several solution states may exist between these extremes. In order to simplify the resulting difference method, Harten, Lax, and van Leer [11] considered approximations with one and two states between the maximum and minimum signals. We'll illustrate their method when one state is present. In this case, their approximate Riemann solution is

$$\mathbf{w}(x/t, \mathbf{u}_L, \mathbf{u}_R) = \begin{cases} \mathbf{u}_L, & \text{if } x/t < \nu_{min} \\ \mathbf{u}_*, & \text{if } \nu_{min} \leq x/t < \nu_{max} \\ \mathbf{u}_R, & \text{if } \nu_{max} \leq x/t \end{cases} . \tag{6.5.12}$$

In order to determine the intermediate state $\mathbf{u}_*$, they integrate the conservation law (6.1.1a) over the cell $(x_j, x_{j+1}) \times (t_n, t_{n+1})$ (Figure 6.5.3) to obtain

$$0 = \int_{x_j}^{x_{j+1}} \int_{t_n}^{t_{n+1}} [\mathbf{u}_t + \mathbf{f}_x(\mathbf{u})] dx dt = \int_{x_j}^{x_{j+1}} \mathbf{u}(x,t)|_{t_n}^{t_{n+1}} dx + \int_{t_n}^{t_{n+1}} \mathbf{f}(\mathbf{u})|_{x_j}^{x_{j+1}} dt.$$

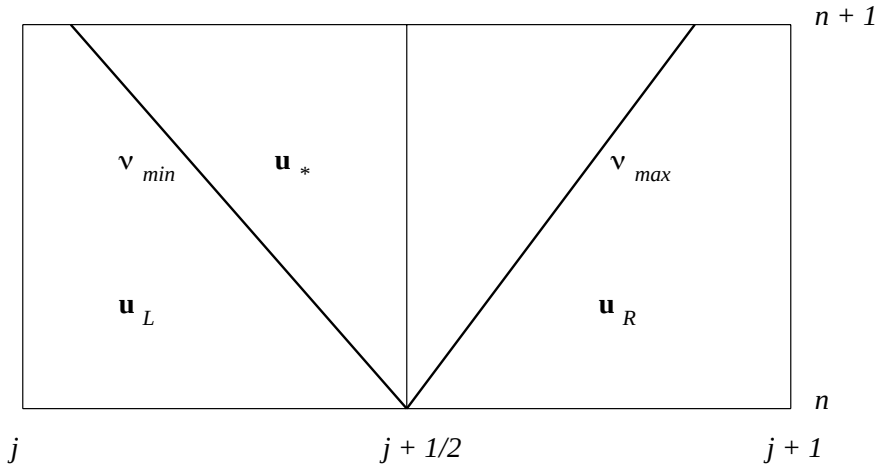

Figure 6.5.3: Computational cell for the Riemann problem for Roe's method. The minimum ν_{min} and maximum ν_{max} signal speeds are assumed to be separated by one constant state $\mathbf{u}_*$.

Since the solutions at $x = x_j$ and x_{j+1} are $\mathbf{u}_L$ and $\mathbf{u}_R$, respectively, we have

$$\int_{x_j}^{x_{j+1}} \mathbf{u}(x,t)|_{t_n}^{t_{n+1}} dx + \Delta t[\mathbf{f}(\mathbf{u}_R) - \mathbf{f}(\mathbf{u}_L)] = 0$$

From (6.5.12), we have

$$\mathbf{u}(x,t_n) = \begin{cases} \mathbf{u}_L, & \text{if } x_j \le x < x_{j+1/2} \\ \mathbf{u}_R, & \text{if } x_{j+1/2} \le x < x_{j+1} \end{cases}$$

and

$$\mathbf{u}(x,t_{n+1}) = \begin{cases} \mathbf{u}_L, & \text{if } x_{j-1} \le x < x_{j+1/2} + \nu_{min}\Delta t \\ \mathbf{u}_*, & \text{if } x_{j+1/2} + \nu_{min}\Delta t \le x < x_{j+1/2} + \nu_{max}\Delta t \\ \mathbf{u}_R, & \text{if } x_{j+1/2} + \nu_{max}\Delta t \le x < x_{j+1} \end{cases} .$$

(The speed ν_{min} has been assumed positive and not as shown in Figure 6.5.3.) Thus, the remaining integral may be evaluated to obtain

$$(\frac{\Delta x}{2} + \nu_{min}\Delta t)\mathbf{u}_L + (\nu_{max} - \nu_{min})\Delta t\mathbf{u}_* + (\frac{\Delta x}{2} - \nu_{max}\Delta t)\mathbf{u}_R - \frac{\Delta x}{2}(\mathbf{u}_L + \mathbf{u}_R)$$

$$+\Delta t[\mathbf{f}(\mathbf{u}_R) - \mathbf{f}(\mathbf{u}_L)] = 0$$

or

$$\mathbf{u}_* = \frac{\nu_{max}\mathbf{u}_R - \nu_{min}\mathbf{u}_L}{\nu_{max} - \nu_{min}} - \frac{\mathbf{f}(\mathbf{u}_R) - \mathbf{f}(\mathbf{u}_L)}{\nu_{max} - \nu_{min}}. \tag{6.5.13}$$

Harten, Lax, and van Leer [11] show that the modified Roe's scheme (6.4.7, 6.5.11a, 6.5.12, 6.5.13) no longer violates the entropy condition.

Harten and Hyman [8] described a procedure for calculating the signal speeds ν_{min} and ν_{max}, which we describe for scalar problems. They begin by selecting $\tilde{\nu}_{min}$ and $\tilde{\nu}_{max}$, respectively, as approximate speeds of the left and right ends of an expansion fan centered at $x = x_{j+1/2}$. These may be taken as

$$\tilde{\nu}_{min} = \min_{\theta\in[0,1]} \hat{A}(u_L, (1-\theta)u_L + \theta u_R), \tag{6.5.14a}$$

$$\tilde{\nu}_{max} = \max_{\theta\in[0,1]} \hat{A}((1-\theta)u_L + \theta u_R, u_R). \tag{6.5.14b}$$

Their search over values of $u \in [u_L, u_R]$ seeks to identify minimum and maximum signal speeds that are not at the extreme values of u_L and u_R.

They then select their signal speeds as

$$\nu_{min} = \hat{A}(u_L, u_R) - (\hat{A}(u_L, u_R) - \tilde{\nu}_{min})^+ \tag{6.5.14c}$$

and

$$\nu_{max} = \hat{A}(u_L, u_R) + (\tilde{\nu}_{max} - \hat{A}(u_L, u_R))^+. \tag{6.5.14d}$$

Using (6.5.11c) in (6.5.13), they compute u_* as

$$u_* = \frac{\hat{A}(u_L, u_R) - \nu_{min}}{\nu_{max} - \nu_{min}} u_L + \frac{\nu_{max} - \hat{A}(u_L, u_R)}{\nu_{max} - \nu_{min}} u_R \tag{6.5.14e}$$

The approximate solution of the Riemann problem is obtained from (6.5.12) translated to $(x_{j+1/2}, t_n)$ Equations (6.5.14) may also be applied to vector systems that have been put into canonical form by the transformation (6.1.3) ([19], Section 4.4).

Example 6.5.4. For the inviscid Burgers' equation, we have $\hat{A}(u_L, u_R) = (u_L + u_R)/2 = \dot{\xi}$ (Example 6.5.3). Using (6.5.14a)

$$\tilde{\nu}_{min} = \min_{\theta\in[0,1]} \frac{u_L + (1-\theta)u_L + \theta u_R}{2} = \min_{\theta\in[0,1]} \frac{(2-\theta)u_L + \theta u_R}{2}.$$

Since the argument of the minimization is a linear function of θ for Burgers' equation, extreme values occur at $\theta = 0, 1$.

$$\tilde{\nu}_{min} = \min(u_L, \frac{u_L + u_R}{2}) = \min(u_L, \dot{\xi}).$$

Similarly, using (6.5.14b)

$$\tilde{\nu}_{max} = \max_{\theta\in[0,1]} \frac{(1-\theta)u_L + (1+\theta)u_R}{2} = \max(\dot{\xi}, u_R).$$

In an expansion region, $u_L < \dot{\xi} < u_R$ and $\tilde{\nu}_{min} = u_L$ and $\tilde{\nu}_{max} = u_R$. Using these in (6.5.14c, 6.5.14d) yields $\nu_{min} = u_L$ and $\nu_{max} = u_R$, which are the exact speeds of the endpoints of the expansion fan. Using these in (6.5.14e) gives $u_* = \dot{\xi}$ and, from (6.5.12),

$$w(x/t, u_L, u_R) = \begin{cases} u_L, & \text{if } x/t < u_L \\ \dot{\xi}, & \text{if } u_L < x/t \le u_R \\ u_R, & \text{if } u_R \le x/t \end{cases} .$$

For a shock, $u_L > \dot{\xi} > u_R$, so $\tilde{\nu}_{min} = \dot{\xi}$ and $\tilde{\nu}_{max} = \dot{\xi}$. Substitution into (6.5.14c, 6.5.14d) yields $\nu_{min} = \nu_{max} = \dot{\xi}$. The intermediate solution disappears from (6.5.12) and the solution of the Riemann problem is

$$w(x/t, u_L, u_R) = \begin{cases} u_L, & \text{if } x/t < \dot{\xi} \\ u_R, & \text{if } \dot{\xi} \le x/t \end{cases} .$$

6.5.2 Van Leer's Method

Van Leer's [22] technique is simplest to present for the linear scalar problem

$$u_t + au_x = 0. \tag{6.5.15}$$

Suppose that we have an approximate solution $W^n(x)$ at time t_n, with $W^0(x)$ corresponding to the prescribed initial data. Using the mesh structure of Figure 6.5.1, construct a piecewise-constant approximation $U^n(x)$ of $W^n(x)$ as

$$U^n(x) = U_j^n = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} W^n(x)dx, \qquad x \in (x_{j-1/2}, x_{j+1/2}]. \tag{6.5.16a}$$

The solution $W^{n+1}(x)$ at t_{n+1} is obtained as the exact solution of (6.5.15); thus,

$$W^{n+1}(x) = U^n(x - a\Delta t). \tag{6.5.16b}$$

The process is repeated for the next time step. The steps in the procedure are illustrated in Figure 6.5.4. If the Courant, Friedrichs, Lewy condition $|a|\Delta t/\Delta x < 1$ is satisfied, then the translation of $U^n(x)$ by the amount $a\Delta t$ does not exceed Δx.

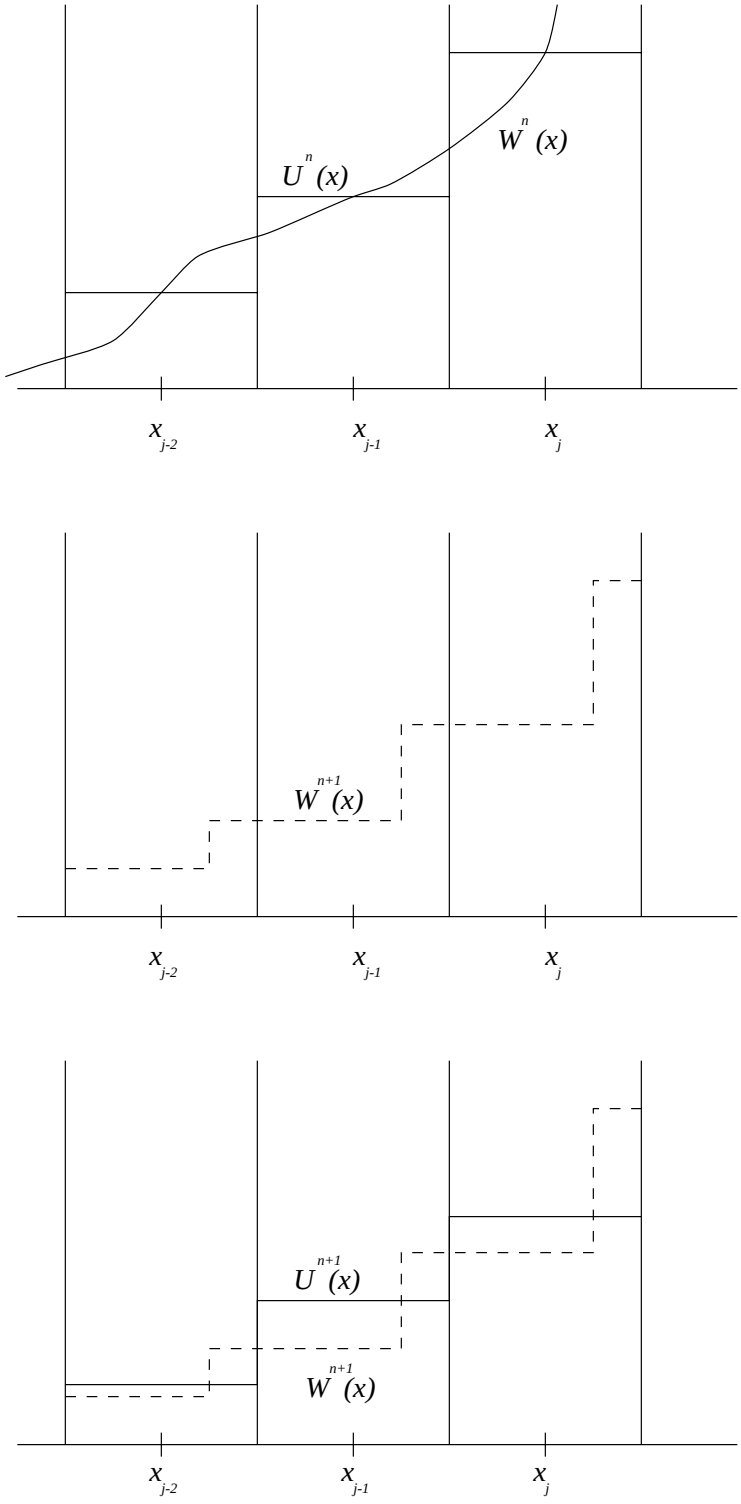

Figure 6.5.4: Projection of $W^n(x)$ onto a piecewise constant function $U^n(x)$ (top). Convection of the piecewise constant functions $U^n(x)$ (center) to obtain $W^{n+1}(x)$. Projection of $W^{n+1}(x)$ onto the piecewise constant function $U^{n+1}(x)$ (bottom).

If $a > 0$ then

$$W^{n+1}(x) = \begin{cases} U_{j-1}^n, & \text{if } x_{j-1/2} < x \le x_{j-1/2} + a\Delta t \\ U_j^n, & \text{if } x_{j-1/2} + a\Delta t < x \le x_{j+1/2} \end{cases}.$$

Substituting this into (6.5.16a) and performing the integration yields

$$U_j^{n+1} = a\bar{\alpha}U_{j-1}^n + (1 - a\bar{\alpha})U_j^n = U_j^n - a\bar{\alpha}(U_j^n - U_{j-1}^n), \tag{6.5.17}$$

which is recognized as the first-order upwind difference scheme.

Van Leer [22] extended this approach to higher orders of accuracy. For example, he obtained a second-order method by projecting $W^n(x)$ onto a space of piecewise linear polynomials. When this is done, one must figure a way of determining the slope of the solution on each subinterval. Van Leer [22] describes several possibilities and some of these also appear in Sod [19], Section 4.3. One possibility uses centered difference approximations from solution values on neighboring subintervals. The description of a second approach using moments follows.

6.5.3 Higher-Order Methods

Higher order methods can achieve excellent accuracy and performance in regions where the solution is smooth. We'll describe a method of constructing high-order methods that uses moments of the solution; however, instead of following the traditional approach [22], we'll describe an alternative treatment, called the *discontinuous Galerkin* method, as developed by Cockburn and Shu [3].

We'll use a method of lines formulation and let $\mathbf{W}(x,t)$ be a piecewise linear approximation of $\mathbf{u}(x,t)$ whose restriction to $(x_{j-1/2}, x_{j+1/2})$ is (Figure 6.5.5)

$$\mathbf{W}(x,t) = \mathbf{U}_j(t) + D\mathbf{U}_j(t)(x - x_j) \tag{6.5.18}$$

where $\mathbf{U}_j(t)$ and $D\mathbf{U}_j(t)$ are approximations of $u(x_j,t)$ and $u_x(x_j,t_n)$, respectively. Let us use the integral form of the conservation law (6.5.1) with (6.5.18) to obtain

$$\frac{d}{dt}\int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{W}(x,t)dx = -\mathbf{f}(\mathbf{W})\Big|_{x_{j-1/2}}^{x_{j+1/2}}.$$

Using the representation (6.5.18) and replacing $\mathbf{f}$ by a numerical flux yields

$$\Delta x\dot{\mathbf{U}}_j(t) = -\mathbf{F}_{j+1/2} + \mathbf{F}_{j-1/2} \tag{6.5.19a}$$

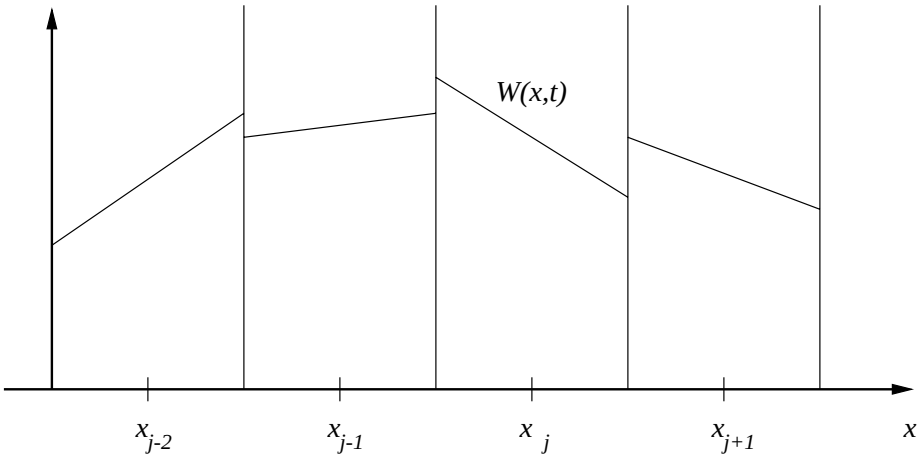

Figure 6.5.5: Piecewise linear function $W^n(x)$ used for the method of Cockburn and Shu [3].

where

$$\mathbf{F}_{j+1/2} = \mathbf{F}_{j+1/2}(\mathbf{U}_j(t) + \frac{\Delta x}{2} D\mathbf{U}_j(t), \mathbf{U}_{j+1}(t) - \frac{\Delta x}{2} D\mathbf{U}_{j+1}(t)). \tag{6.5.19b}$$

The gradient $D\mathbf{U}_j$ is determined by taking a moment of (6.1.1a); thus, we multiply (6.1.1a) by $x - x_j$, integrate over the subinterval $(x_{j-1/2}, x_{j+1/2})$, and replace the exact solution by the approximation (6.5.18) to obtain

$$\frac{d}{dt}\int_{x_{j-1/2}}^{x_{j+1/2}} (x - x_j)\mathbf{W}(x,t)dx + \int_{x_{j-1/2}}^{x_{j+1/2}} (x - x_j)\mathbf{f}_x(\mathbf{W})dx = 0$$

Integrating the flux term by parts

$$\frac{d}{dt}\int_{x_{j-1/2}}^{x_{j+1/2}} (x - x_j)\mathbf{W}(x,t)dx + (x - x_j)\mathbf{f}(\mathbf{W})|_{x_{j-1/2}}^{x_{j+1/2}} = \int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{f}(\mathbf{W})dx.$$

Using (6.5.18)

$$\frac{\Delta x^3}{12} D\dot{\mathbf{U}}_j + \frac{\Delta x}{2}[\mathbf{F}_{j+1/2} + \mathbf{F}_{j-1/2}] = \int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{f}(\mathbf{W})dx. \tag{6.5.19c}$$

Equations (6.5.19b) and (6.5.19c) are two vector ($2m$ scalar) equations for the unknowns $\mathbf{U}_j(t)$ and $D\mathbf{U}_j(t)$. To complete the specification of the scheme (6.5.18, 6.5.19), we must (i) choose a numerical flux, (ii) evaluate the integral on the right of (6.5.19c), (iii) choose a time integration strategy, and (iv) define initial conditions.

The flux can be evaluated by a Riemann or other numerical flux. Cockburn and Shu [3] describe several choices. Perhaps the simplest is the Lax-Friedrichs flux which, in this context, is

$$\mathbf{F}_{j+1/2}(\mathbf{W}_L, \mathbf{W}_R) = \frac{1}{2}[\mathbf{f}(\mathbf{W}_L) + \mathbf{f}(\mathbf{W}_R) - \max(|\lambda_{min}|, |\lambda_{max}|)(\mathbf{W}_R - \mathbf{W}_L)] \quad (6.5.20a)$$

where, using (6.5.18),

$$\mathbf{W}_L = \mathbf{U}_j(t) + \frac{\Delta x}{2} D\mathbf{U}_j(t), \qquad \mathbf{W}_R = \mathbf{U}_{j+1}(t) - \frac{\Delta x}{2} D\mathbf{U}_{j+1}(t) \quad (6.5.20b)$$

are the solutions to the immediate left and right, respectively, of $x_{j+1/2}$ and λ_{min} and λ_{max} are, respectively, the minimum and maximum eigenvalues of $\mathbf{f_u}(\mathbf{u})$, $\mathbf{W}_L \le \mathbf{u} \le \mathbf{W}_R$.

The integral on the right of (6.5.19c) is usually evaluated numerically. With the midpoint rule, we have

$$\int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{f}(\mathbf{W}) dx = \Delta x \mathbf{f}(\mathbf{U}_j) + O(\Delta x^3). \quad (6.5.20c)$$

Cockburn and Shu [3] recommend a total variation diminishing (TVD) Runge-Kutta scheme. Biswas *et al.* [1] found that classical Runge-Kutta formulas gave similar results. With forward Euler integration and (6.5.20c), the scheme (6.5.19) becomes

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}[\mathbf{F}_{j+1/2}^n - \mathbf{F}_{j-1/2}^n] \quad (6.5.21a)$$

$$D\mathbf{U}_j^{n+1} = D\mathbf{U}_j^n + \frac{6\Delta t}{\Delta x^2}[\mathbf{F}_{j+1/2}^n + \mathbf{F}_{j-1/2}^n - 2\mathbf{f}(\mathbf{U}_j^n)]. \quad (6.5.21b)$$

While the forward Euler method may be used for time integration, a second-order method would be more compatible with the higher spatial accuracy.

Finally, initial conditions may also be obtained by taking moments

$$\int_{xj-1/2}^{x_{j+1/2}} [\mathbf{W}(x,0) - \boldsymbol{\phi}(x)] dx = 0, \qquad \int_{xj-1/2}^{x_{j+1/2}} (x - x_j)[\mathbf{W}(x,0) - \boldsymbol{\phi}(x)] dx = 0. \quad (6.5.22)$$

The scheme may be recognized by some as a Galerkin method, which is the principal technique used with finite element methods. It uses a discontinuous solution representation; hence, the name discontinuous Galerkin method. It has several advantages:

- it readily extends to higher orders of accuracy by using higher-degree polynomials and taking higher-order moments of the conservation law (6.1.1a) [3, 1];
- it may be used in two and three dimensions in much the same way as in one dimension [1, 4];
- it extends to unstructured meshes of triangular or tetrahedral elements [4]; and
- it has a compact structure that is suitable for parallel computation.

Unfortunately, since it is second order, it will have oscillations near discontinuities. These can be reduced by *limiting* the computed solution. With limiting, the gradient $D\mathbf{U}_j(t)$ variable is reevaluated to eliminate overshoots in the solution. In particular $D\mathbf{U}_j$ is modified so that:

1. the linear function (6.5.18) does not take on values outside of the adjacent grid averages (Figure 6.5.6, upper left);
2. local extrema are set to zero (Figure 6.5.6, upper right); and
3. the gradient is replaced by zero if its sign is not consistent with its neighbors (Figure 6.5.6, lower center).

A formula for accomplishing these goals can be summarized concisely using the minimum modulus function as

$$D\mathbf{U}_{j,mod} = \text{minmod}(D\mathbf{U}_j, \nabla\mathbf{U}_j, \Delta\mathbf{U}_j) \tag{6.5.23a}$$

where

$$\text{minmod}(a,b,c) = \begin{cases} \text{sgn}(a)\min(|a|,|b|,|c|), & \text{if } \text{sgn}(a) = \text{sgn}(b) = \text{sgn}(c) \\ 0, & \text{otherwise} \end{cases} \tag{6.5.23b}$$

and ∇ and Δ are the backward and forward difference operators, respectively.

Example 6.5.5. Biswas *et al.* [1] solve the inviscid Burgers' equation (6.4.4a) with the initial data

$$u(x,0) = \frac{1+\sin x}{2}.$$

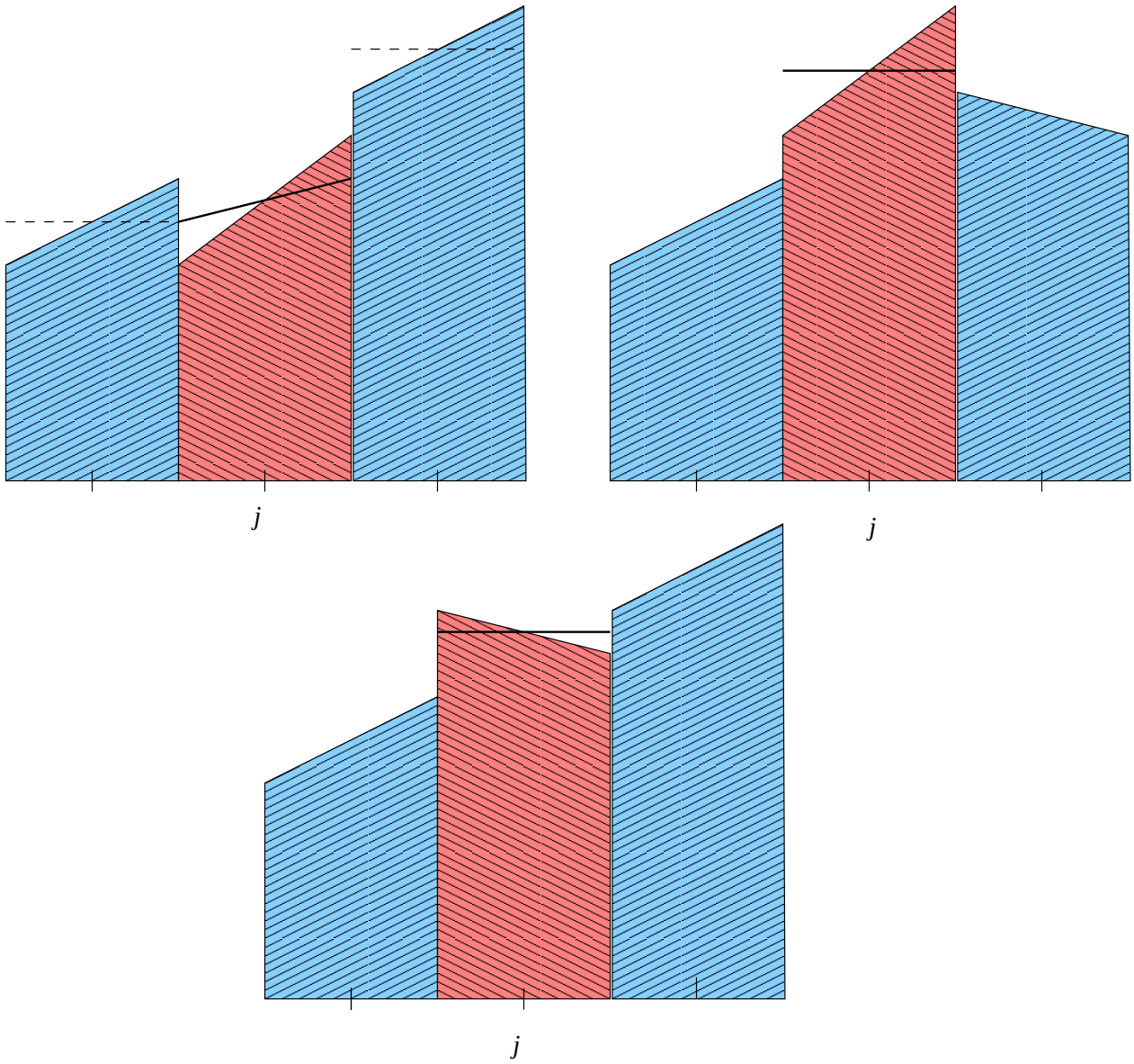

Figure 6.5.6: Solution limiting: reduce slopes to be within neighboring averages (upper left); set local extrema to zero (upper right); and set slopes to zero if they disagree with neighboring trends.

They use an upwind numerical flux

$$F_{j+1/2}(W_L, W_R) = \begin{cases} f(W_L), & \text{if } f'(U_j) \geq 0 \\ f(W_R), & \text{if } f'(U_j) < 0 \end{cases}$$

and solve problems with $\Delta x = 1/32$ using first-, second- and third-order methods. The first-order method is obtained from (6.5.21a) by setting $DU_j = 0$, $j = 0, 1, \ldots, J$. This is just upwind differencing. The second-order method is (6.5.18, 6.5.21–6.5.23) with the upwind numerical flux. The third-order method is derived in the same manner as the

second-order method using a piecewise quadratic function to represent the solution. Time integration was done using classical Runge-Kutta methods of orders 1-3, respectively, for the first-, second-, and third-order methods. The solutions are shown in Figure 6.5.7. The piecewise polynomial functions used to represent the solution are plotted at eleven points on each subinterval.

The first-order solution (upper left of Figure 6.5.7) is characteristically diffusive. The second-order solution (upper right of Figure 6.5.7) has greatly reduced the diffusion while not introducing any spurious oscillations. The minimum modulus limiter (6.5.23) has a tendency to flatten solutions near extrema. This can be seen in the third-order solution shown at the lower left of Figure 6.5.7. There is a loss of (local) monotonicity near the shocks. (Average solution values are monotone and this is all that the limiter (6.5.23) was designed to produce.) Biswas *et al.* [1] developed an adaptive limiter that reduces "flattening" and does a better job of preserving local monotonicity near discontinuities. The solution with this limiter is shown in the lower portion of Figure 6.5.7.

Evaluating numerical fluxes and using limiting for vector systems is more complicated than indicated by the previous scalar example. Cockburn and Shu [3] reported problems when applying limiting component-wise. At the price of additional computation, they applied limiting to the characteristic fields obtained by diagonalizing the Jacobian $\mathbf{f}_\mathbf{u}$. Biswas *et al.* [1] proceeded in a similar manner. "Flux-vector splitting" may provide a compromise between the two extremes. As an example, consider the solution and flux vectors for the one-dimensional Euler equations of compressible flow

$$\mathbf{u} = \begin{bmatrix} \rho \\ \rho u \\ e \end{bmatrix}, \qquad \mathbf{f}(\mathbf{u}) = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ u(e+p) \end{bmatrix} \tag{6.5.24a}$$

where ρ, u, e, and

$$p = (\gamma - 1)(e - \rho u^2)/2 \tag{6.5.24b}$$

are, respectively, the fluid's density, velocity, internal energy per unit volume, and pressure.

For this and related differential systems, the flux vector is a homogeneous function

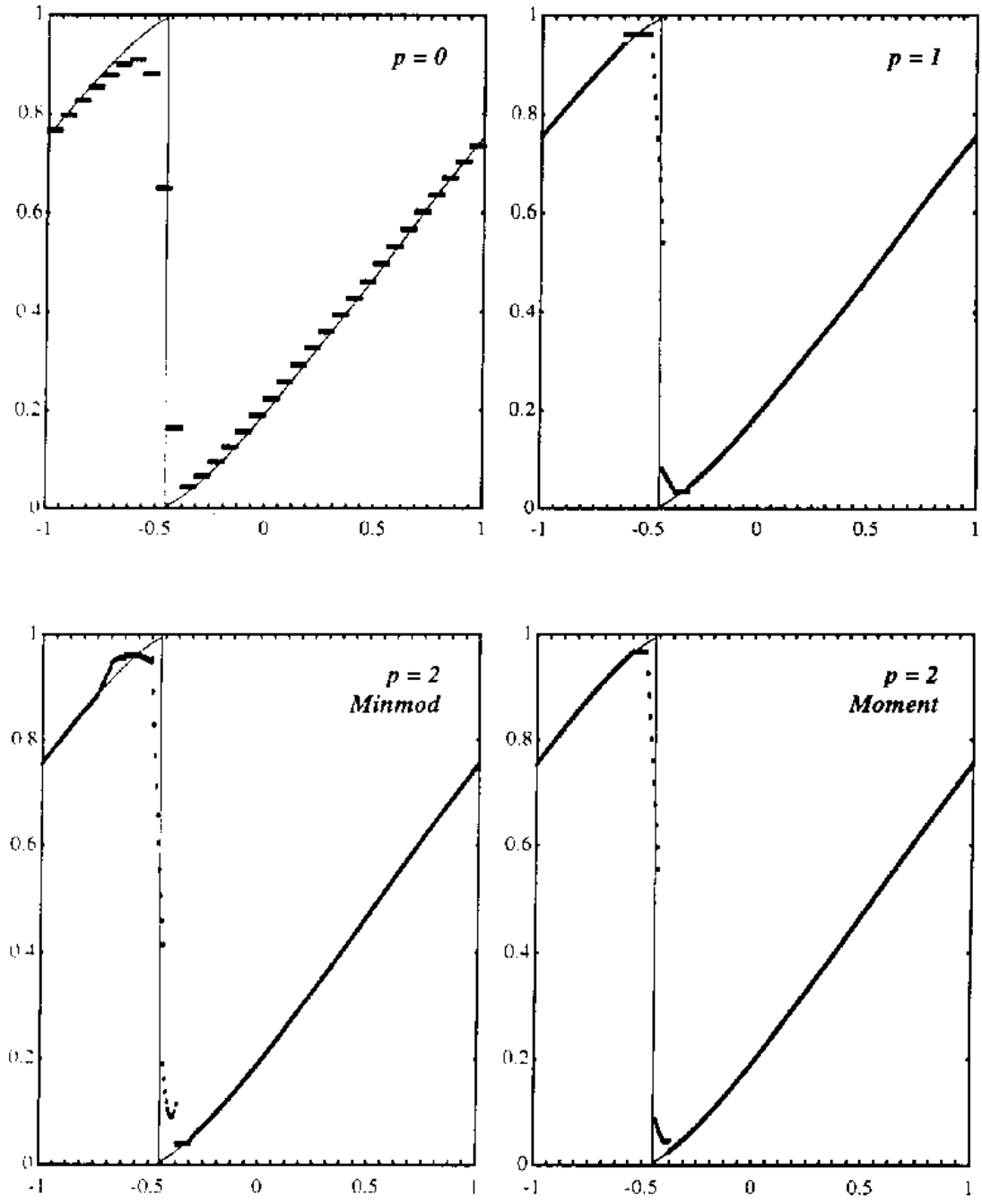

Figure 6.5.7: Exact (line) and numerical solutions of Example 5 for methods with piecewise polynomial approximations of degrees $p = 0$, 1, and 2. Solutions with a minmod limiter and an adaptive limiter of Biswas et al. [1] are shown for $p = 2$.

that may be expressed as

$$\mathbf{f}(\mathbf{u}) = \mathbf{A}\mathbf{u} = \mathbf{f}_{\mathbf{u}}(\mathbf{u})\mathbf{u}. \tag{6.5.25a}$$

Since the system is hyperbolic, the Jacobian $\mathbf{A}$ may be diagonalized as described in Section 6.1 to yield

$$\mathbf{f}(\mathbf{u}) = \mathbf{P}^{-1}\mathbf{\Lambda}\mathbf{P}\mathbf{u} \tag{6.5.25b}$$

where the diagonal matrix $\mathbf{\Lambda}$ contains the eigenvalues of $\mathbf{A}$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix} = \begin{bmatrix} u-c & & \\ & u & \\ & & u+c \end{bmatrix}. \tag{6.5.25c}$$

The variable $c = \sqrt{\partial p/\partial \rho}$ is the speed of sound in the fluid. The matrix $\mathbf{\Lambda}$ can be decomposed into components

$$\mathbf{\Lambda} = \mathbf{\Lambda}^+ + \mathbf{\Lambda}^- \tag{6.5.26a}$$

where $\mathbf{\Lambda}^+$ and $\mathbf{\Lambda}^-$ are, respectively, composed of the non-negative and non-positive components of $\mathbf{\Lambda}$

$$\lambda_i^{\pm} = \frac{\lambda_i \pm |\lambda_i|}{2}, \qquad i = 1, 2, \ldots, m. \tag{6.5.26b}$$

Writing the flux vector in similar fashion using (6.5.25)

$$\mathbf{f}(\mathbf{u}) = \mathbf{P}^{-1}(\mathbf{\Lambda}^+ + \mathbf{\Lambda}^-)\mathbf{P}\mathbf{u} = \mathbf{f}(\mathbf{u})^+ + \mathbf{f}(\mathbf{u})^-. \tag{6.5.27}$$

Split fluxes for the Euler equations were presented by Steger and Warming [20]. Van Leer [23] found an improvement that provided improvements near sonic and stagnation points of the flow. The split fluxes are evaluated by upwind techniques. Letting $\mathbf{W}_L$ and $\mathbf{W}_R$ be solution vectors on the left and right of an interface (*e.g.*, at $x_{j+1/2}$), then $\mathbf{f}^+$ is evaluated using $\mathbf{W}_L$ and $\mathbf{f}^-$ is evaluated using $\mathbf{W}_R$.

6.6 Nonlinear Instability and Boundary Conditions

Linear stability analyses (*e.g.*, using von Neumann's method) are necessary for stability, but, possibly, not sufficient because of nonlinearity or boundary conditions. In order to illustrate this effect, consider solving the inviscid Burgers' equation

$$u_t + uu_x = 0$$

by the leap frog scheme

$$U_j^{n+1} = U_j^{n-1} - \bar{\alpha} U_j^n (U_{j+1}^n - U_{j-1}^n), \tag{6.6.1}$$

where $\bar{\alpha} = \Delta t/\Delta x$. In order to analyze the sensitivity of (6.6.1) to small perturbations, let

$$U_j^n = \bar{U}_j^n + \delta U_j^n \tag{6.6.2}$$

where $\bar{U}_j^n$ is a solution of (6.6.1) and δU_j^n is a perturbation. Substituting (6.6.2) into (6.6.1)

$$\bar{U}_j^{n+1} + \delta U_{j+1}^n = U_j^{n-1} + \delta U_j^{n-1} - \bar{\alpha}(\bar{U}_j^n + \delta U_j^n)[(\bar{U}_{j+1}^n + \delta U_{j+1}^n) - (\bar{U}_{j-1}^n + \delta U_{j-1}^n)].$$

Expanding the above expression while using the fact that $\bar{U}_j^n$ satisfies (6.6.1) and neglecting squares and products of perturbations, we find

$$\delta U_j^{n+1} = \delta U_j^{n-1} - \bar{\alpha}\bar{U}_j^n(\delta U_{j+1}^n - \delta U_{j-1}^n) - \bar{\alpha}\delta U_j^n(\bar{U}_{j+1}^n - \bar{U}_{j-1}^n). \tag{6.6.3a}$$

The difference equation (6.6.3a) is linear, but has variable coefficients. In order to use a von Neumann stability analysis we would also have to assume that $\bar{U}_j^n = \bar{U}$, a constant. Were this done, (6.6.3a) would become

$$\delta U_j^{n+1} = \delta U_j^{n-1} - \bar{\alpha}\bar{U}(\delta U_{j+1}^n - \delta U_{j-1}^n). \tag{6.6.3b}$$

Using the von Neumann method (Problem 6.3.1), we would conclude that the leap frog scheme is absolutely stable when the Courant, Friedrichs, Lewy condition $|\bar{U}|\bar{\alpha} \leq 1$ is satisfied. We would like to use this condition as a local stability analysis (*cf.* Section 4.4) to infer that the leap frog scheme is absolutely stable for this nonlinear problem when

$$\bar{\alpha} \max_j |U_j^n| \leq 1. \tag{6.6.3c}$$

A local stability analysis works for many finite difference schemes and problems; however, if disturbances are sufficiently strong, they may cause nonlinear instabilities. Gary [5] introduced an example where the solution of (6.6.1) has the form

$$U_j^n = \bar{U} + c_n \cos \pi j/2 + s_n \sin \pi j/2 + v_n \cos \pi j. \tag{6.6.4}$$

The trigonometric terms may be regarded as high-frequency perturbations to a constant solution $\bar{U}$. By substituting (6.6.4) into (6.6.1) we may show that c_n, s_n, and v_n satisfy the difference equations

$$c_{n+1} - c_{n-1} = 2\bar{\alpha} s_n (v_n - \bar{U}), \tag{6.6.5a}$$

$$s_{n+1} - s_{n-1} = 2\bar{\alpha} c_n (v_n - \bar{U}), \tag{6.6.5b}$$

$$v_{n+1} = v_{n-1}. \tag{6.6.5c}$$

The solution of (6.6.5c) only takes on two values

$$v_n = \begin{cases} A, & \text{if } n \text{ is even} \\ B, & \text{if } n \text{ is odd} \end{cases}. \tag{6.6.5d}$$

We may use (6.6.5d) to eliminate v_n and s_n from (6.6.5a, 6.6.5b) to get

$$c_{n+2} - 2\mu c_n + c_{n-2} = 0, \tag{6.6.6a}$$

where

$$\mu = 1 + 2\bar{\alpha}^2 (B - \bar{U})(A + \bar{U}). \tag{6.6.6b}$$

The constant-coefficient difference equation (6.6.6a) may be solved to yield

$$c_n = C_+ r_+^n + C_- r_-^n \tag{6.6.7a}$$

where

$$r_\pm = -\mu \pm i\sqrt{1 - \mu^2}. \tag{6.6.7b}$$

The constants C_+ and C_- are determined from the initial and starting conditions.

The coefficients c_n, $n \geq 0$, will not grow as n increases as long as $|r_\pm| \leq 1$. This in turn requires that $|\mu| \leq 1$. Using (6.6.4–6.6.7), we may show that the local linear stability condition (6.6.3c) is satisfied when the following two inequalities are satisfied:

$$\bar{\alpha}[|\bar{U}| + \max(|A|, |B|)] \leq 1 \tag{6.6.8a}$$

and

$$-1 < \bar{\alpha}^2 (B - \bar{U})(A + \bar{U}). \tag{6.6.8b}$$

Under these conditions:

- If $|A| < \bar{U}$, $|B| \leq \bar{U}$, and the linear stability conditions (6.6.8a, 6.6.8b) are satisfied then $|\mu| < 1$.

- If $|A|$ and $|B|$ are not bounded by $\bar{U}$, then it is possible to satisfy (6.6.8a, 6.6.8b) with $|\mu| > 1$. For example, choose $A = 0$, $B = \bar{U} + \epsilon$, with $\bar{U}$ and ϵ positive. Then, (6.6.8b) is satisfied since

$$\bar{\alpha}^2 \epsilon \bar{U} > 0.$$

  Choosing $\bar{\alpha}$ such that

$$\bar{\alpha}(2\bar{U} + \epsilon) < 1$$

  ensures that (6.6.8a) is satisfied. However, using (6.6.6b),

$$\mu = 1 + 2\bar{\alpha}^2 \epsilon \bar{U} > 1.$$

  Thus, solutions will grow in magnitude. This example illustrates that high-frequency modes of a solution may interact with other modes to cause a nonlinear instability. This provides another reason to choose a scheme that dissipates high-frequency modes (Section 6.3).

6.6.1 Boundary Conditions

We need extra "artificial" boundary conditions with centered finite difference schemes like the Lax-Friedrichs and Lax-Wendroff methods. Let us illustrate this using a simple linear problem.

Example 6.6.1. Consider the initial-boundary value problem for the kinematic wave equation

$$u_t - u_x = 0, \qquad 0 < x < 1, \qquad t > 0, \tag{6.6.9a}$$

$$u(x,0) = \phi(x), \qquad u(1,t) = \psi(t), \tag{6.6.9b}$$

which has the exact solution

$$u(x,t) = \begin{cases} \phi(x+t), & \text{if } x + t \leq 1 \\ \psi(x+t-1), & \text{if } x + t > 1 \end{cases}.$$

Establishing a uniform grid on $[0,1] \times t > 0$ and using a centered spatial scheme with the four-point stencil shown in Figure 6.6.1, we see that additional information is needed to compute the solution along the line $x = 0$.

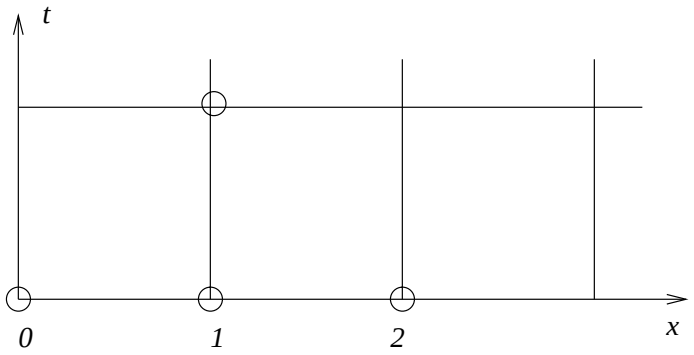

Figure 6.6.1: Uniform grid and four-point finite difference scheme illustrating the need for extra information at $x = 0$.

Consider, for example, a problem with $\phi(x) = x$. The exact solution for $x + t < 1$ is $u(x,t) = x + t$. Let's solve this problem using the leap frog scheme

$$U_j^{n+1} = U_j^{n-1} + \frac{\Delta t}{\Delta x}(U_{j+1}^n - U_{j-1}^n) \tag{6.6.10a}$$

and the Lax-Friedrichs scheme

$$U_j^{n+1} = \frac{U_{j+1}^n + U_{j-1}^n}{2} + \frac{\Delta t}{2\Delta x}(U_{j+1}^n - U_{j-1}^n) \tag{6.6.10b}$$

with $\Delta t = \Delta x = 0.1$ and simple constant extrapolation

$$U_0^{n+1} = U_1^{n+1}$$

as a boundary condition at $x = 0$. We further use the exact solution to start the multi-level leap frog scheme. The results, for a few time steps are shown in Table 6.6.1.

Both the leap frog and Lax-Friedrichs schemes produce the exact solution of this initial value problem since the Courant number is unity. Thus, the only errors arise due to the boundary approximation. From the table, we see that the Lax-Friedrichs scheme has confined the rather inaccurate treatment of the solution at $x = 0$ to the one cell adjacent to the boundary. With the leap frog scheme, however, the poor boundary approximation is polluting the accuracy of the solution in the interior of the domain. The choice of proper

n	Leap Frog				Lax-Friedrichs			
	j				j			
	0	1	2	3	0	1	2	3
0	0.0	0.1	0.2	0.3	0.0	0.1	0.2	0.3
1	0.1	0.2	0.3	0.4	0.2	0.2	0.3	0.4
2	0.3	0.3	0.4	0.5	0.3	0.3	0.4	0.5
3	0.3	0.3	0.5	0.6	0.4	0.4	0.5	0.6
4	0.5	0.5	0.7	0.7	0.5	0.5	0.6	0.7
5	0.5	0.5	0.7	0.7	0.6	0.6	0.7	0.8

Table 6.6.1: Solution of Example 6.6.1 using the leap frog and Lax-Friedrichs methods with constant extrapolation at the boundary $x = 0$.

artificial boundary conditions for hyperbolic problems has been thoroughly studied. A good account of the theory and practice appears in Strikwerda [21], Chapter 11. The two principal methods of analyzing stability involve Laplace transforms and matrix methods. The former approach is often called the Gustaffson, Kreiss, Sundström, and Osher theory ([21], Section 11.3. The latter is the technique that we studied in Section 3.3. We'll be much more brief and note that, as a rule of thumb, one should try to use dissipative difference schemes and dissipative boundary conditions. Thus, with the present example, we could consider using the Lax-Friedrichs scheme in the interior and the forward time-forward space scheme

$$U_0^{n+1} = U_0^n + \frac{\Delta t}{\Delta x}(U_1^n - U_0^n)$$

at the boundary. With unit Courant number, this combination gives the exact solution.

Prescribing boundary conditions for $m \times m$ linear vector systems (6.1.1b) requires the canonical form

$$\begin{bmatrix} \mathbf{w}^- \\ \mathbf{w}^+ \end{bmatrix}_t + \begin{bmatrix} \mathbf{\Lambda}^- & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}^+ \end{bmatrix} \begin{bmatrix} \mathbf{w}^- \\ \mathbf{w}^+ \end{bmatrix}_x = \mathbf{0}, \tag{6.6.11a}$$

where

$$\mathbf{AP} = \mathbf{P\Lambda}, \qquad \mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}^- & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}^+ \end{bmatrix}, \tag{6.6.11b}$$

$$\mathbf{\Lambda}^- = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix}, \qquad \mathbf{\Lambda}^+ = \begin{bmatrix} \lambda_{p+1} & & & \\ & \lambda_{p+2} & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}, \tag{6.6.11c}$$

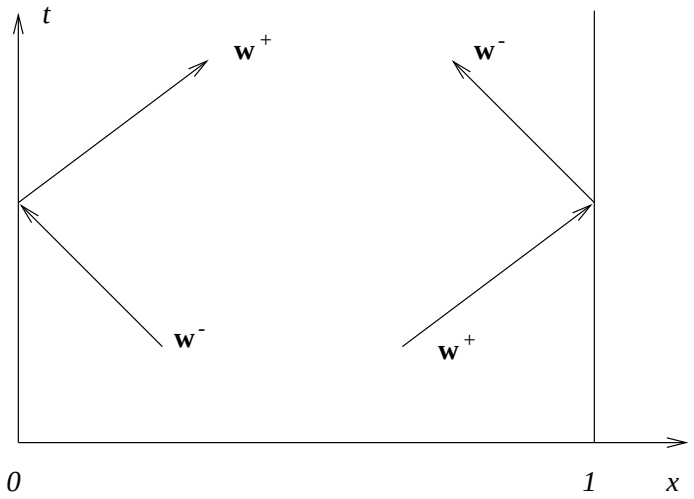

Figure 6.6.2: Boundary condition prescription for an initial-boundary value problem for (6.1.1b).

with $\lambda_1 < \lambda_2 < \ldots < \lambda_p < 0$ and $\lambda_{p+1} < \lambda_{p+2} < \ldots < \lambda_m > 0$. Boundary conditions are needed along those characteristics that enter the region. Thus, for an initial boundary value problem on $0 < x < 1$, boundary conditions should be prescribed for $\mathbf{w}^+$ at $x = 0$ and for $\mathbf{w}^-$ at $x = 1$ (Figure 6.6.2). Artificial boundary conditions would, therefore, be needed for $\mathbf{w}^-$ at $x = 0$ and for $\mathbf{w}^+$ at $x = 1$.

Example 6.6.2. Initial-boundary value problems on infinite domains must be approximated on finite domains. Artificial boundaries, placed far from the region of interest, can cause spurious wave reflections back into the domain. The goal is to construct boundary conditions on the artificial boundaries to eliminate or reduce these spurious reflections. Let us partition $\mathbf{P}$ as

$$\mathbf{A}\mathbf{P}^- = \mathbf{P}^-\mathbf{\Lambda}^-, \qquad \mathbf{A}\mathbf{P}^+ = \mathbf{P}^+\mathbf{\Lambda}^+. \tag{6.6.12a}$$

For simplicity, assume that $\mathbf{P}$ is orthogonal in the sense that

$$(\mathbf{P}^-)^T\mathbf{P}^+ = 0. \tag{6.6.12b}$$

Assuming that $\mathbf{P}$ is constant, the canonical transformation (6.1.3) may be differentiated to obtain

$$\dot{\mathbf{u}} = \mathbf{P}\dot{\mathbf{w}} = \mathbf{P}^-\dot{\mathbf{w}}^- + \mathbf{P}^+\dot{\mathbf{w}}^+$$

where $\dot{(\ )} = d(\ )/dt$.

Let the artificial boundary be $x = 0$ of Figure 6.6.2. No waves will enter the region if $\dot{\mathbf{w}}^+ = 0$; thus,

$$\dot{\mathbf{u}} = \mathbf{P}^- \dot{\mathbf{w}}^-$$

or, multiplying by $(\mathbf{P}^+)^T$ and using the orthogonality condition (6.6.12b),

$$(\mathbf{P}^+)^T \dot{\mathbf{u}} = 0. \tag{6.6.13a}$$

There are many possibilities for evaluating the directional derivative

$$\dot{\mathbf{u}} = \mathbf{u}_t + \dot{x}\mathbf{u}_x. \tag{6.6.13b}$$

Hedstrom [12] evaluates (6.6.13b) along the line $x = 0$. Another possibility is to assume that the initial data is trivial for $x \leq 0$; thus, in particular $\mathbf{u}_x = 0$. In either case, the nonreflecting boundary conditions (6.6.13) become

$$(\mathbf{P}^+)^T \mathbf{u}_t = 0. \tag{6.6.14}$$

Problems

1. For the solution (6.6.4) of the leap frog problem (6.6.1) verify that relations (6.6.5a–6.6.5c), (6.6.6a, 6.6.6b), (6.6.7a, 6.6.7b), and (6.6.8a, 6.6.8b) are satisfied.

2. ([21], Section 11.2). Consider the initial-boundary value problem (6.6.9) with the initial and boundary conditions prescribed so that its exact solution is

$$u(x,t) = \sin 2\pi(x+t).$$

   2.1. Solve this problem using the leap frog scheme (6.6.10a) with the following artificial boundary conditions applied at $x = 0$:

$U_0^{n+1} = U_1^{n+1}$	Constant extrapolation
$U_0^{n+1} = U_1^n$	Constant diagonal extrapolation
$U_0^{n+1} = U_0^n + \lambda(U_1^n - U_0^n)$	Upwind differencing
$U_0^{n+1} = U_0^{n-1} + 2\lambda(U_1^n - U_0^n)$	Pseudo leap frog
$U_0^{n+1} = 2U_1^{n+1} - U_2^{n+1}$	Linear extrapolation
$U_0^{n+1} = 2U_1^n - U_2^{n-1}$	Linear diagonal extrapolation

Use $\Delta x = 0.05$ and $\lambda = 0.95$. Display the solution at $t = 1.9$. Comment on the stability of each boundary condition based on the computational results. Stop any calculation prematurely if the computed solution exceeds two in magnitude. Display the solution at the last acceptable time in this case.

2.2. Repeat the computation for the Lax-Wendroff scheme.

Bibliography

[1] R. Biswas, K.D. Devine, and J.E. Flaherty. Parallel adaptive finite element methods for conservation laws. *Applied Numerical Mathematics*, 14:255–284, 1993.

[2] A.J. Chorin. Random choice solution of hyperbolic systems. *Journal of Computational Physics*, 25:517–533, 1976.

[3] B. Cockburn and C.-W. Shu. TVB Runge-Kutta local projection discontinuous finite element method for conservation laws II: General framework. *Mathematics of Computation*, 52:411–435, 1989.

[4] J.E. Flaherty, R.M. Loy, C. Özturan, M.S. Shephard, B.K. Szymanski, J.D. Teresco, and L.H. Ziantz. Parallel structures and dynamic load balancing for adaptive finite element computation. *Applied Numerical Mathematics*, 26:241–265, 1998.

[5] J. Gary. Lecture notes on the numerical solution of partial differential equations. Technical report, University of Colorado, Boulder, 1975.

[6] J. Glimm. Solutions in the large for nonlinear hyperbolic systems of equations. *Communications on Pure and Applied Mathematics*, 18:697–715, 1965.

[7] S.K. Godunov. A finite difference method for the numerical computation of discontinuous solutions of the equations of fluid dynamics. *Mat. Sbornik.*, 47:271–306, 1959.

[8] A. Harten and J.M. Hyman. Self-adjusting grid methods for one-dimensional hyperbolic conservation laws. Technical Report LASL Report LA-9105, Los Alamos Scientific Laboratory, Los Alamos, 1983.

[9] A. Harten, J.M. Hyman, and P.D. Lax. On finite-difference approximations and entropy conditions for shocks. *Communications on Pure and Applied Mathematics*, 29:297–322, 1976.

[10] A. Harten and P.D. Lax. A random choice finite difference scheme for hyperbolic conservation laws. *SIAM Journal on Numerical Analysis*, 18:289–315, 1981.

[11] A. Harten, P.D. Lax, and B. van Leer. On upstream difference and Godunov-type schemes for hyperbolic conservation laws. *SIAM Review*, 25:35–61, 1983.

[12] G. Hedstrom. Nonreflecting boundary conditions for nonlinear hyperbolic systems. *Journal of Computational Physics*, 30:222–237, 1979.

[13] A. Jeffrey and T. Taniuti. *Non-linear Wave Propagation.* Academic Press, New York, 1964.

[14] P.D. Lax. *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves.* Regional Conference Series in Applied Mathematics, No. 11. SIAM, Philadelphia, 1973.

[15] P.D. Lax and B. Wendroff. Systems of conservation laws. *Communications on Pure and Applied Mathematics*, 13:217–237, 1960.

[16] A.R. Mitchell and D.F. Griffiths. *The Finite Difference Method in Partial Differential Equations.* John Wiley and Sons, Chichester, 1980.

[17] R.D. Richtmyer and K.W. Morton. *Difference Methods for Initial Value Problems.* John Wiley and Sons, New York, second edition, 1967.

[18] P.L. Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *Journal of Computational Physics*, 43:357–372, 1981.

[19] G.A. Sod. *Numerical Methods in Fluid Dynamic.* Cambridge University Press, Cambridge, 1985.

[20] J.L Steger and R.F. Warming. Flux vector splitting of the inviscid gasdynamic equations with applications to finite difference methods. *Journal of Computational Physics*, 40:263–293, 1981.

[21] J.C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations.* Chapman and Hall, Pacific Grove, 1989.

[22] B. van Leer. Towards the ultimate conservation difference scheme IV. A new approach of numerical convection. *Journal of Computational Physics*, 23:276–299, 1977.

[23] B. van Leer. Flux-vector splitting gor the Euler equations. *Lecture Notes in Physics*, 170:507–512, 1982.